Université de Montréal

Transformation-Based Error-Driven learning

(Eric Brill)

par Pierre-Marc Jodoin

Université de Montréal
Montréal, Canada

31 mars 2003

Université de Montréal

Mise en contexte

Problématique:

- Simplifications markoviennes trop restrictives
- La plupart des modèles markoviens conditionnent les probabilités sur les n-mots précédents.

  ex: Taggers $\hat{t}_1^n = \arg\max_{t_1^n} p(w_1^n \mid t_1^n) \times p(t_1^n)$

  comment distinguer **Can** you help me?

  et **Can** of beer.
- Les modèles markoviens plus sophistiqués sont coûteux

Solution:

- Apprendre des règles de transformation explicites basées sur autre chose que des probabilités conditionnelles

Credit : Pierre-Marc Jodoin

Université de Montréal

Plan

- Introduction à la philosophie des TBEDLs.
- Applications et résultats
  - *Part of Speech tagging*
  - Induction automatique de grammaires
  - *Prepositional Phrase Disambiguation*
- Conclusion

Université de Montréal

Introduction aux TBEDLs

- **Observation** : des règles linguistiques complexes peuvent être observées par des «épiphénomènes » simples.

  When will the very first race /NOUN end?

  He will race /VERB the car

- Apprendre des transformations à partir d'un corpus donné et *sans probabilité conditionnelle*:

  Ex.: Si un pronom (HE) apparaît parmi les 2 mots précédents, change le tag /NOUN par /VERB

Credit : Pierre-Marc Jodoin

Université de Montréal

Introduction aux TBEDLs

- Qu'est-ce qu'une transformation?

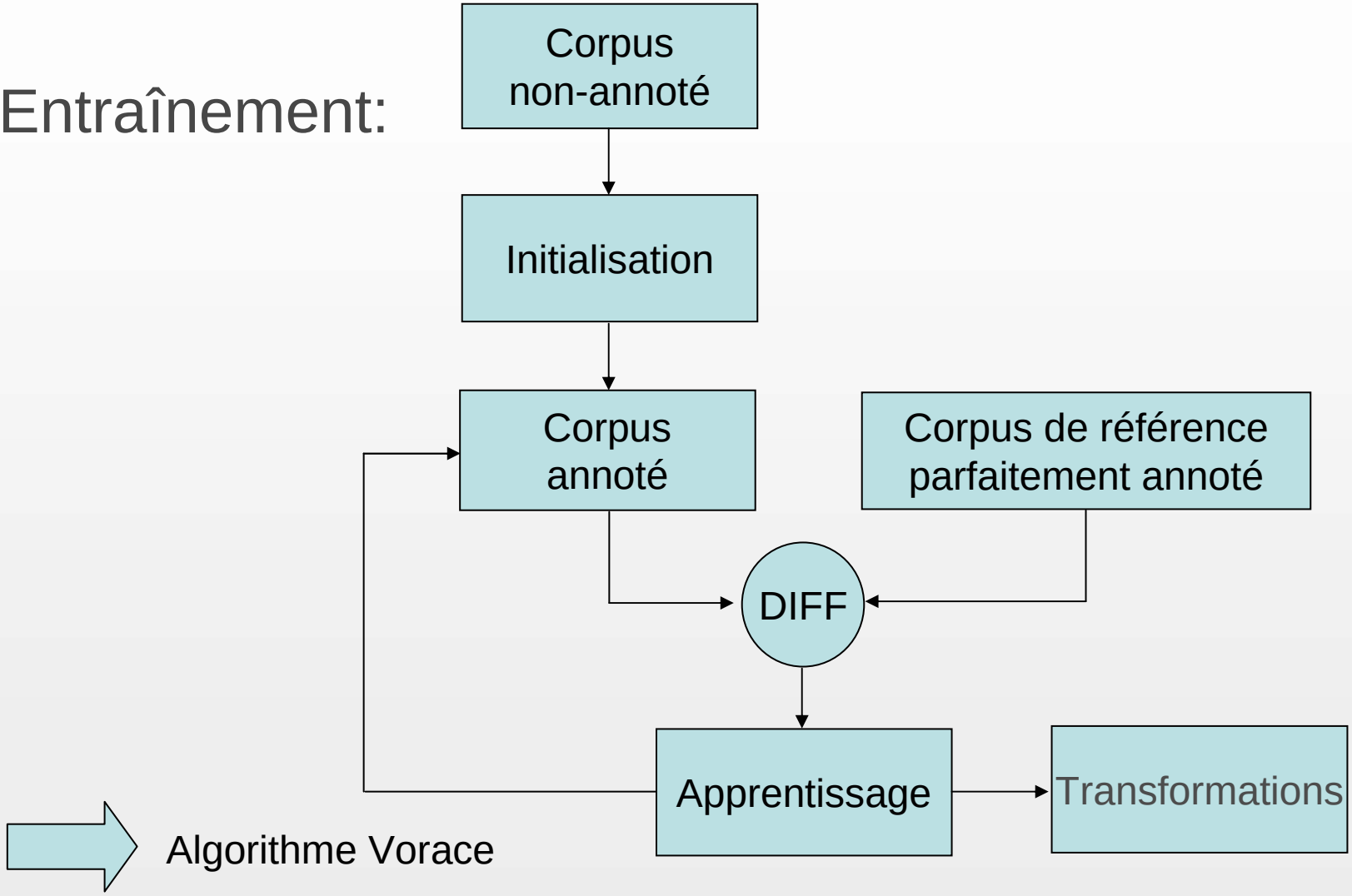
Transformation = Trigger + Règle de réécriture

Ex: Trigger : un pronom (HE) apparaît parmi les 2 mots précédents
Règle de réécriture : change le tag /NOUN par /VERB

Université de Montréal

Introduction aux TBEDLs

Entraînement:

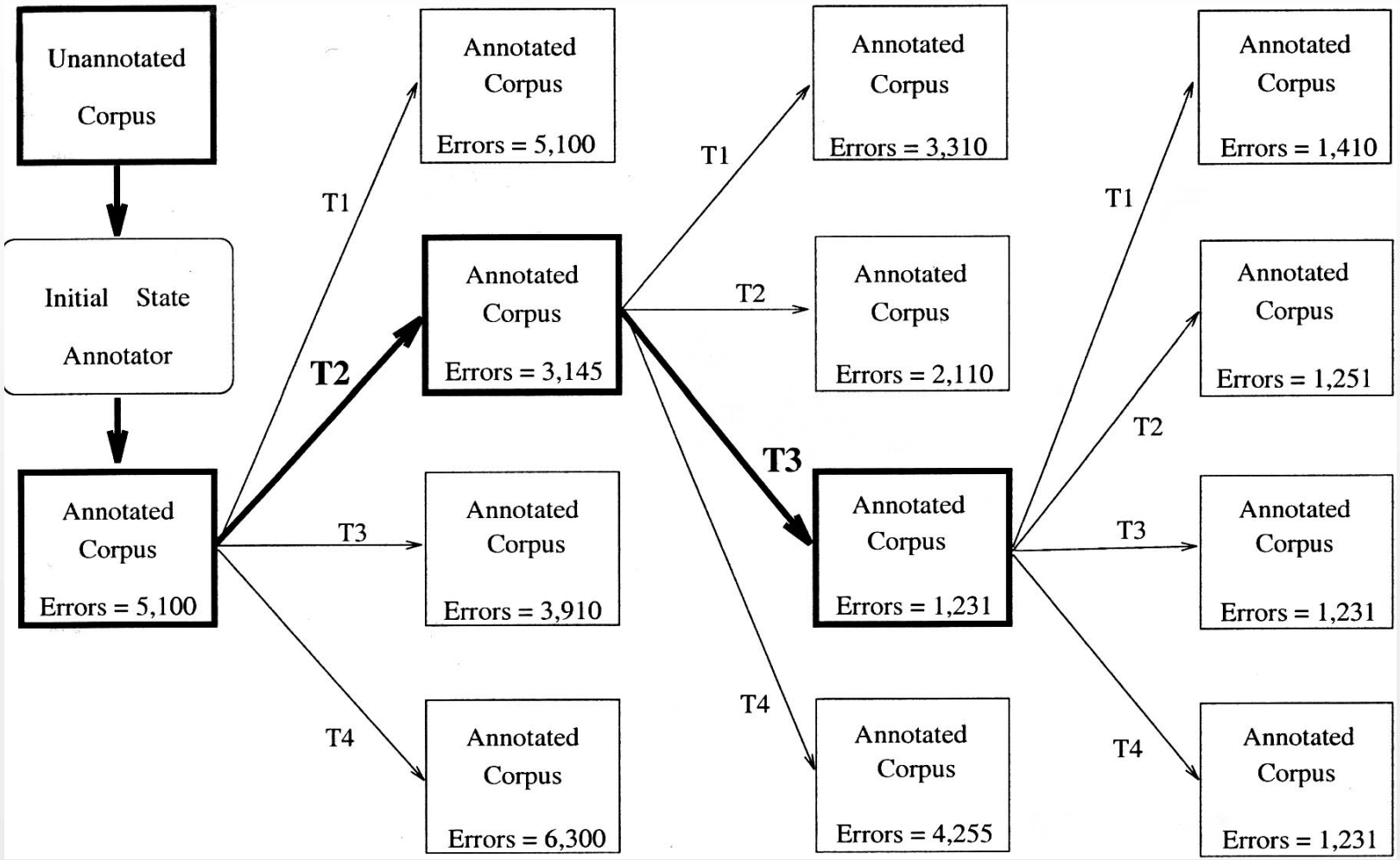


Credit : Pierre-Marc Jodoin

Université de Montréal

Introduction aux TBEDLs

Exemple de fonctionnement:



Credit : Pierre-Marc Jodoin

Université de Montréal

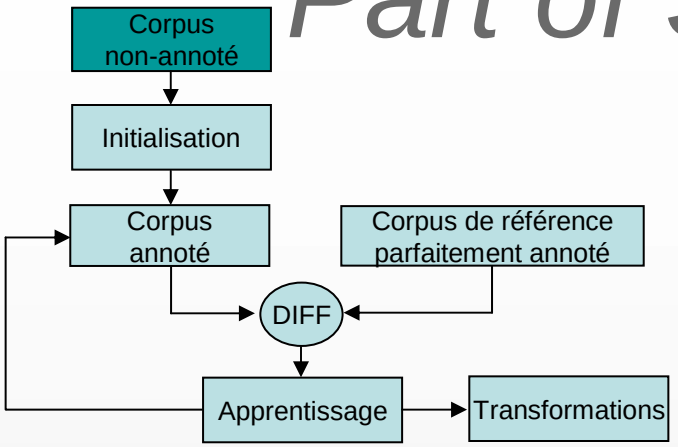
Plan

- Introduction à la philosophie des TBEDLs.
- Applications et résultats
  - *Part of Speech tagging*
  - Induction automatique de grammaires
  - *Prepositional Phrase Disambiguation*
- Conclusion

Credit : Pierre-Marc Jodoin

Université de Montréal

Part of Speech tagging



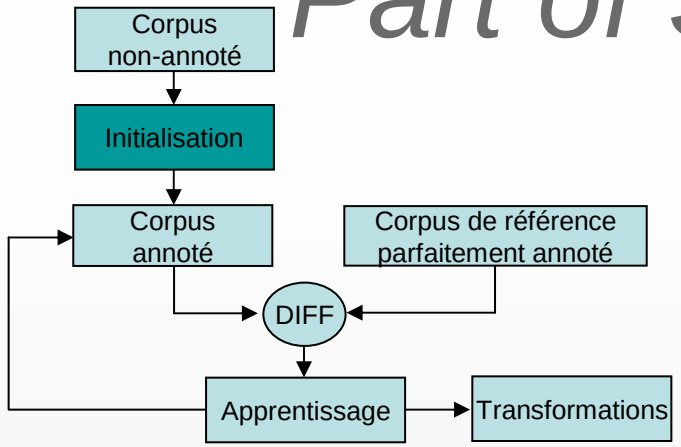
Corpus non taggué

W-3	W-2	W-2	W-1	W0	W1	W2	W3	W4

Credit : Pierre-Marc Jodoin

Université de Montréal

Part of Speech tagging



Corpus non taggué

W-3	W-2	W-2	W-1	W0	W1	W2	W3	W4

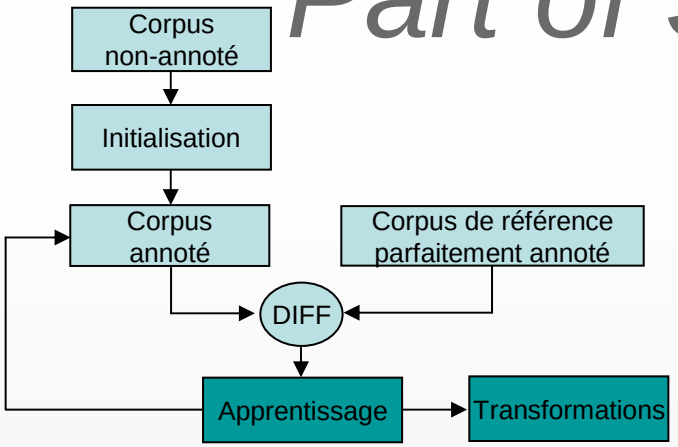
- Initialisation:

$$\hat{t}_i = \text{argmax}_{t_x} \, p(w_i | t_x)$$

W-3	W-2	W-2	W-1	W0	W1	W2	W3	W4
VB	WB	PR	AR	NN	RP	CC	TO	UH

Université de Montréal

Part of Speech tagging



Corpus non taggué

W-3	W-2	W-2	W-1	W0	W1	W2	W3	W4

- Initialisation:

$$\hat{t}_i = \text{argmax}_{t_x} \, p(w_i | t_x)$$

W-3	W-2	W-2	W-1	W0	W1	W2	W3	W4
VB	WB	PR	AR	NN	RP	CC	TO	UH

- Transformation pour W0:

Trigger : Si W-2 a le tag PR

Règle: Changer le tag NN en tag VB

W-3	W-2	W-2	W-1	W0	W1	W2	W3	W4
VB	WB	PR	AR	VB	RP	CC	TO	UH

Université de Montréal

Part of Speech tagging

Transformations Apprises

	Change Tag		
#	From	To	Condition
1	NN	VB	Previous tag is *TO*
2	VBP	VB	One of the previous three tags is *MD*
3	NN	VB	One of the previous two tags is *MD*
4	VB	NN	One of the previous two tags is *DT*
5	VBD	VBN	One of the previous three tags is *VBZ*
6	VBN	VBD	Previous tag is *PRP*
7	VBN	VBD	Previous tag is *NNP*
8	VBD	VBN	Previous tag is *VBD*
9	VBP	VB	Previous tag is *TO*
10	POS	VBZ	Previous tag is *PRP*
11	VB	VBP	Previous tag is *NNS*
12	VBD	VBN	One of previous three tags is *VBP*
13	IN	WDT	One of next two tags is *VB*
14	VBD	VBN	One of previous two tags is *VB*
15	VB	VBP	Previous tag is *PRP*
16	IN	WDT	Next tag is *VBZ*
17	IN	DT	Next tag is *NN*
18	JJ	NNP	Next tag is *NNP*
19	IN	WDT	Next tag is *VBD*
20	JJR	RBR	Next tag is *JJ*

Credit : Pierre-Marc Jodoin

Université de Montréal

Part of Speech tagging

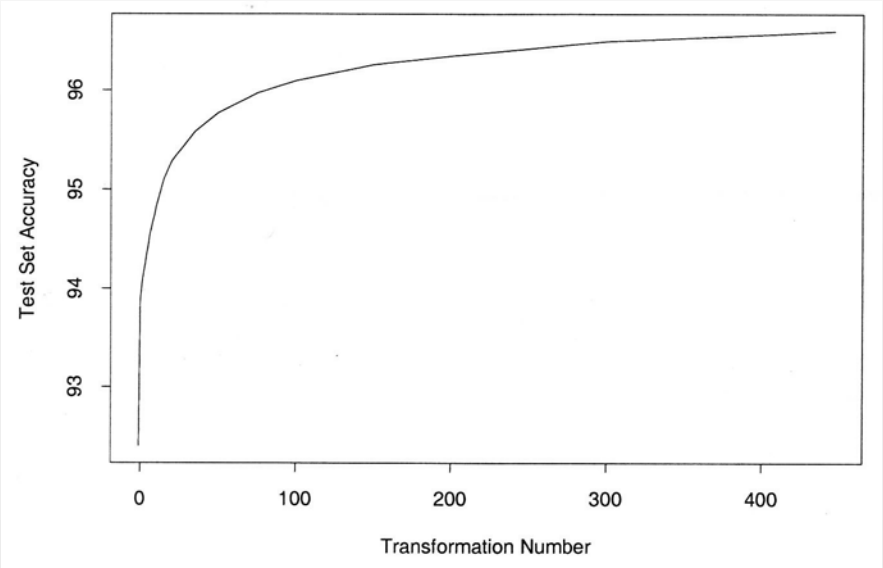
	Change Tag		
#	From	To	Condition
1	NN	NNS	Has suffix **-s**
2	NN	CD	Has character **.**
3	NN	JJ	Has character **-**
4	NN	VBN	Has suffix **-ed**
5	NN	VBG	Has suffix **-ing**
6	??	RB	Has suffix **-ly**
7	??	JJ	Adding suffix **-ly** results in a word.
8	NN	CD	The word **$** can appear to the left.
9	NN	JJ	Has suffix **-al**
10	NN	VB	The word **would** can appear to the left.
11	NN	CD	Has character **0**
12	NN	JJ	The word **be** can appear to the left.
13	NNS	JJ	Has suffix **-us**
14	NNS	VBZ	The word **it** can appear to the left.
15	NN	JJ	Has suffix **-ble**
16	NN	JJ	Has suffix **-ic**
17	NN	CD	Has character **1**
18	NNS	NN	Has suffix **-ss**
19	??	JJ	Deleting the prefix **un-** results in a word
20	NN	JJ	Has suffix **-ive**

Université de Montréal

Part of Speech tagging

Résultats:

Method	Training Corpus Size (Words)	# of Rules or Context. Probs.	Acc. (%)
Stochastic	64 K	6,170	96.3
Stochastic	1 Million	10,000	96.7
Rule-Based With Lex. Rules	64 K	215	96.7
Rule-Based With Lex. Rules	600 K	447	97.2
Rule-Based w/o Lex. Rules	600 K	378	97.0



Credit : Pierre-Marc Jodoin

Université de Montréal

Plan

- Introduction à la philosophie des TBEDLs.
- Applications et résultats
  - *Part of Speech tagging*
  - Induction automatique de grammaires
  - *Prepositional Phrase Disambiguation*
- Conclusion

Credit : Pierre-Marc Jodoin

Université de Montréal

Induction automatique de grammaires

(Sans TBEDL)

- Induction grammaticale : apprendre les paramètres d'une CFG ou d'une PCFG.

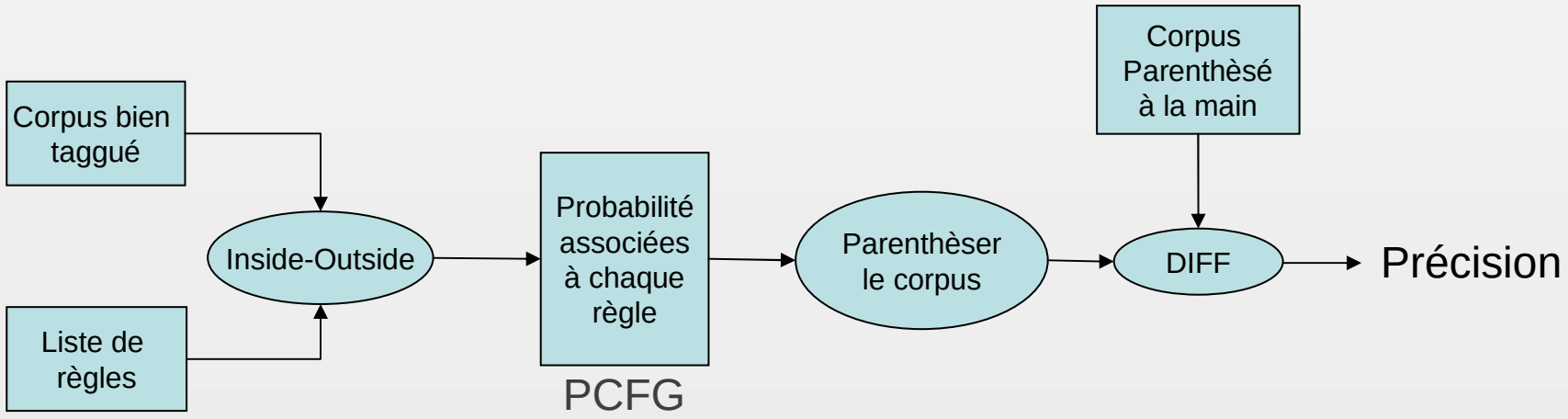
  Paramètres = règles (+ probabilités)

- De façon plus pragmatique:

  Il est plus facile d'apprendre des PCFGs;

  les règles sont prédéfinies;

  le meilleur algorithme est *Inside-Outside*.



Credit : Pierre-Marc Jodoin

Université de Montréal

Induction automatique de grammaires

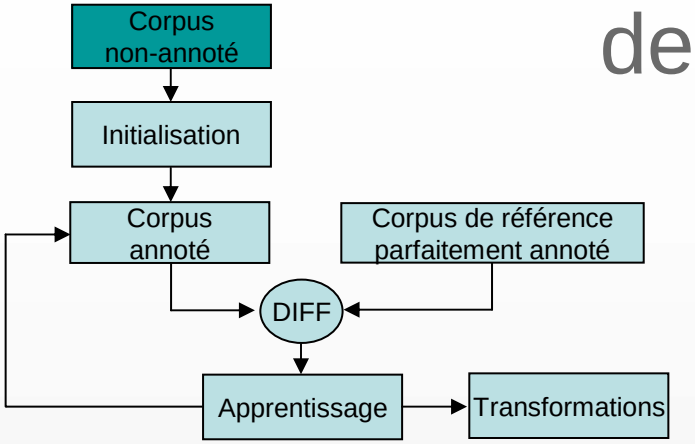
Pourquoi utiliser TBEDL?

1. TBEDL plus simple que Inside-Outside
2. nbTransformations << nbRègles
3. Précision plus élevée
4. Parenthèser un texte $\in \Theta(n)$
5. Plus robuste face à un texte bruité
6. Pas besoin d'exemples négatifs

Credit : Pierre-Marc Jodoin

Université de Montréal

Induction automatique de grammaires



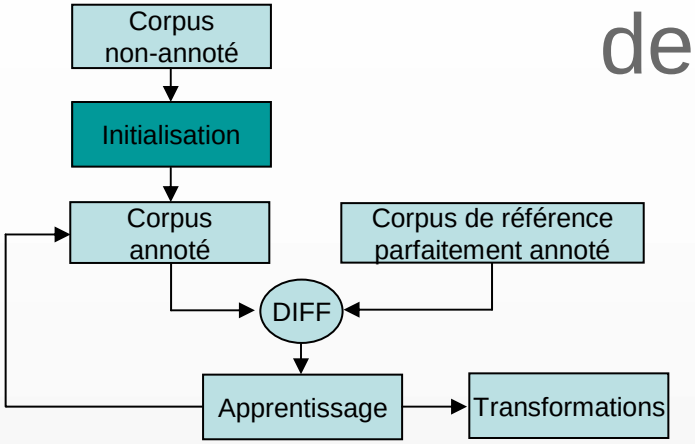
Corpus non parenthèsé mais taggué

The dog and old cat ate

Det NN CC JJ NN VB

Université de Montréal

Induction automatique de grammaires



Corpus non parenthèsé mais taggué

The dog and old cat ate
Det NN CC JJ NN VB

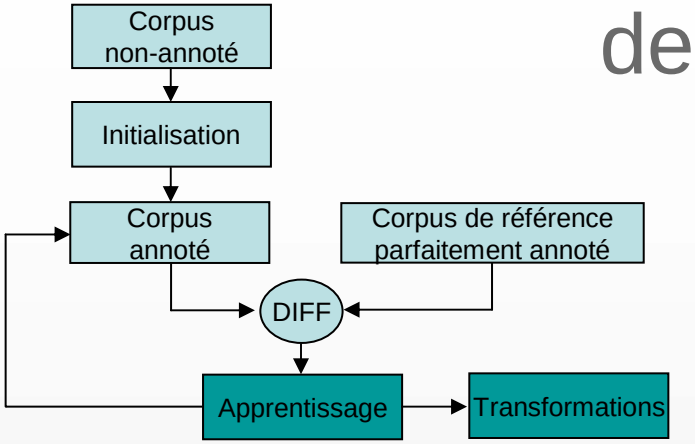
- Initialisation:

  Parenthèsage à gauche de TOUTES les phrases.

The dog and old cat ate
((The (dog (and (old (cat ate)))))).)

Université de Montréal

Induction automatique de grammaires



Corpus non parenthèsé mais taggué

The dog and old cat ate
Det NN CC JJ NN VB

- Initialisation:

  Parenthèsage à gauche de TOUTES les phrases.

The dog and old cat ate
((The (dog (and (old (cat ate)))))).)

- Transformations:

  Trigger : Si W_0 est a le tag NN

  Règle: Ajouter une parenthèse Droite à sa droite.

((The (dog (and (old (cat ate))))).)
Det NN CC JJ NN VB

(((The dog) and ((old cat) ate)))).)
Det NN CC JJ NN VB

Credit : Pierre-Marc Jodoin

Université de Montréal

Induction automatique de grammaires

Transformations Apprises

#	Add/ Delete	Left/ Right Paren	Environment
1	D	L	Left of NN
2	D	L	Left of NNS
3	A	R	Left of ,
4	D	L	Btwn NNP and NNP
5	D	L	Right of DT
6	A	R	Left of ,
7	D	R	Left of NNS
8	D	R	Btwn NN and NN
9	D	L	Btwn JJ and JJ
10	D	L	Right of $

Credit : Pierre-Marc Jodoin

Université de Montréal

Induction automatique de grammaires

- Résultats (Accuracy = précision)

Air Travel corpus (HGD90)

Method	# of Training Corpus Sentences	Accuracy
Inside-Outside	700	90.36%
Transformation Learner	150	91.12%

4096 règles

64 Transformations

Penn Treebank

TBEDL

Sent. Length	# Training Corpus Sents	# of Trans-formations	% Accuracy
2-15	250	83	88.1
2-15	500	163	89.3
2-15	1000	221	91.6
2-20	250	145	86.2
2-25	250	160	83.8

Inside-Outside

Sent. Length	# Training Corpus Sents	# of Rules	Précision
2-12	30000	10605	88.6
2-16	30000	10605	85
2-20	**30000**	**10605**	**83.5**
2-25	**30000**	**10605**	**80.6**
2-30	30000	10605	78.8

[Charniak]

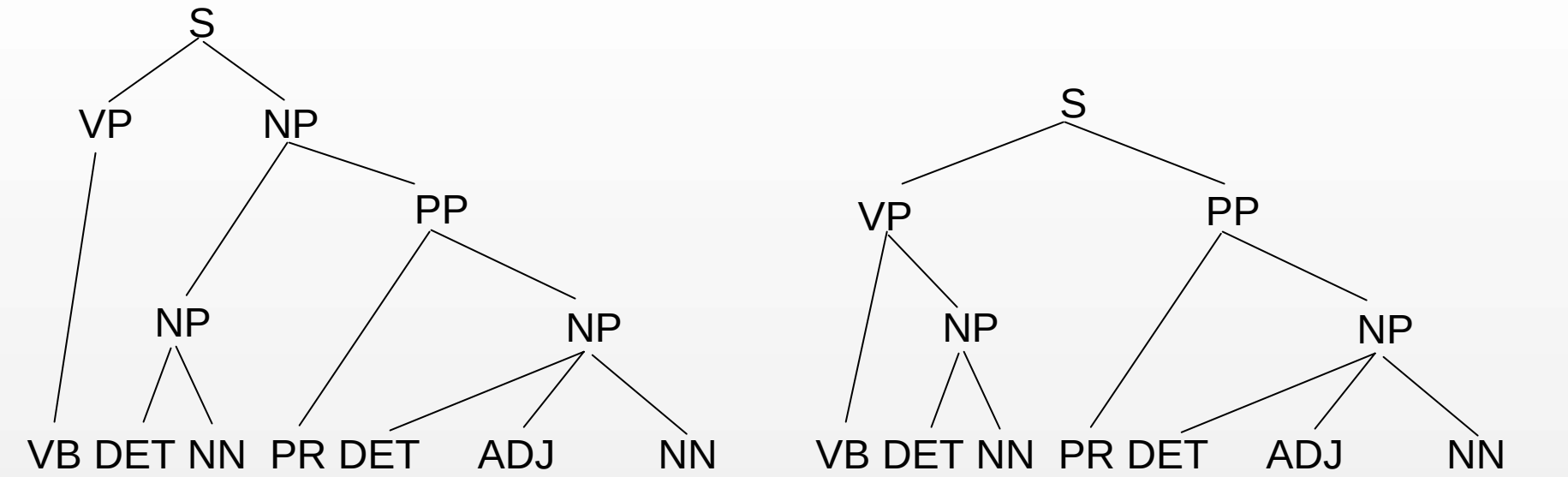
Université de Montréal

Plan

- Introduction à la philosophie des TBEDLs
- Applications et résultats
  - *Part of Speech tagging*
  - Induction automatique de grammaires
  - *Prepositional Phrase Disambiguation*
- Conclusion

Université de Montréal

Prepositional Phrase Disambiguation



Question: Est-ce que PP s'attache à NP ou à VP?

Université de Montréal

Prepositional Phrase Disambiguation

(sans TBEDL)

1. PCFG:

Trouver l'arbre de plus haute probabilité.

P(Arbre) = P(S → VP NP) x P(VP→ Buy) x P(NP → NP PP) x …

2. Simple methods [Hindle,Rooth] [Charniak]

Soit une phrase de la forme : « verbe np_1 prep np_2. »

P(A | prep, verb, np_1, np_2,*w*) = ?

Après simplification

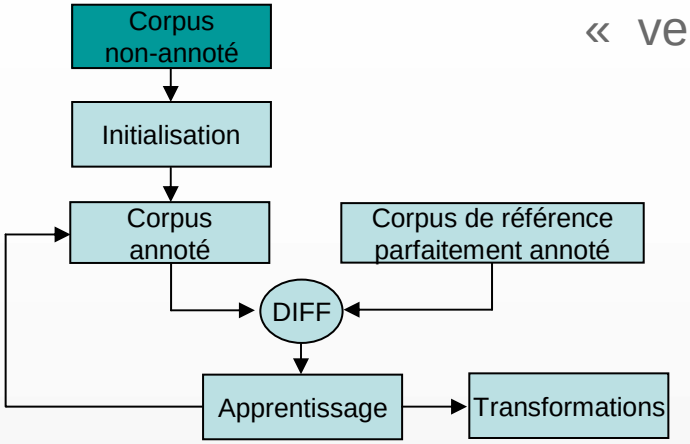
P(prep | verb) > P (prep | noun)

2.1 Information sémantique

Credit : Pierre-Marc Jodoin

Prepositional Phrase Disambiguation

« verbe np_1 prep np_2. »



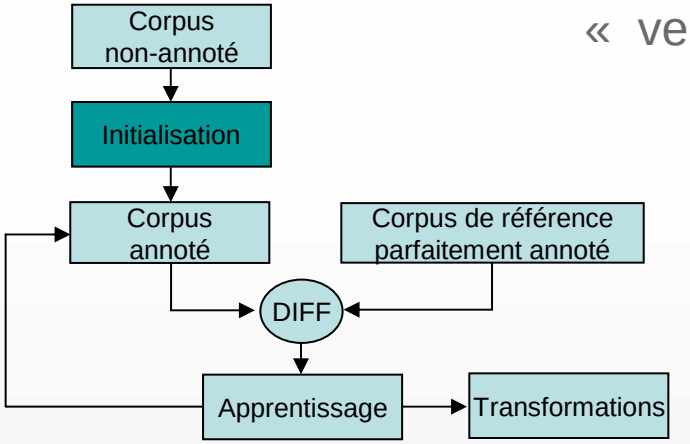
Corpus ambigu mais taggué

Buy	a	plant	with	Jane
VB	Det	NN	Prep	NN

Credit : Pierre-Marc Jodoin

Prepositional Phrase Disambiguation

« verbe np_1 prep np_2. »



Corpus ambigu mais taggué

Buy	a	plant	with	Jane
VB	Det	NN	Prep	NN

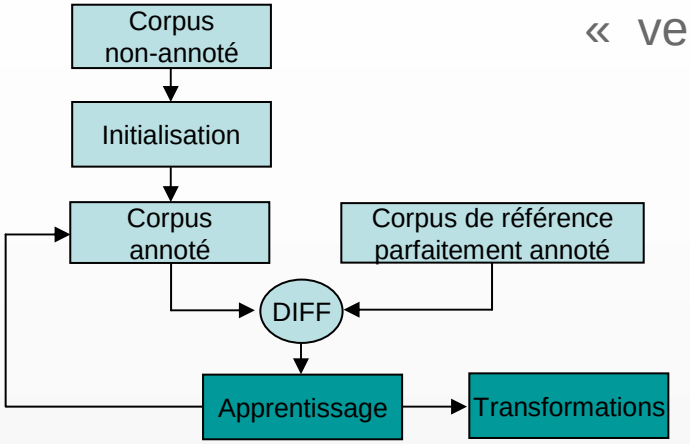
- Initialisation:

  Associer toutes les Prep. à NN .

Buy	a	plant	with	Jane
VB	Det	NN	Prep	NN

Prepositional Phrase Disambiguation

« verbe np_1 prep np_2. »



Corpus ambigu mais taggué

Buy	a	plant	with	Jane
VB	Det	NN	Prep	NN

- Initialisation:

  Associer toutes les Prep. à NN .

Buy	a	plant	with	Jane
VB	Det	NN	Prep	NN

- Transformations:

  Trigger : Si Prep est «with»

  Règle: changer l'attachement du nom au verbe

Buy	a	plant	with	Jane
VB	Det	NN	Prep	NN

Credit : Pierre-Marc Jodoin

Université de Montréal

Prepositional Phrase Disambiguation

Transformations Apprises

#	Change Attachment Location From	To	Condition
1	N1	V	P is *at*
2	N1	V	P is *as*
3	N1	V	P is *into*
4	N1	V	P is *from*
5	N1	V	P is *with*
6	N1	V	N2 is *year*
7	N1	V	P is *by*
8	N1	V	P is *in* and N1 is *amount*
9	N1	V	P is *through*
10	N1	V	P is *during*
11	N1	V	V is *put*
12	N1	V	N2 is *month*
13	N1	V	P is *under*
14	N1	V	P is *after*
15	N1	V	V is *have* and P is *in*
16	N1	V	P is *without*
17	V	N1	P is *of*
18	N1	V	V is *buy* and P is *for*
19	N1	V	P is *before*
20	N1	V	V is *have* and P is *on*

Credit : Pierre-Marc Jodoin

Université de Montréal

Prepositional Phrase Disambiguation

- Résultats:

[Hindle,Rooth]

Method	Accuracy	# of Transforms
t-Scores	70.4 - 75.8	
Transformations	80.8	471
Transformations (no N2)	79.2	418

Associated Press Newswire

Université de Montréal

Plan

- Introduction à la philosophie des TBEDLs.
- Applications et résultats
  - *Part of Speech tagging*
  - Induction automatique de grammaires
  - *Prepositional Phrase Disambiguation*
- Conclusion

Credit : Pierre-Marc Jodoin

Université de Montréal

Conclusion

- Approche **simple**, **vorace** et ***non probabilistes***
  - *Apprendre des transformations : trigger + règle de réécriture*
- Exploite des **régularités syntaxiques** et des **contextes** plus vastes que les modèles markoviens.
- Très **compact**
- Donne en moyenne de meilleurs **résultats**
- Requiert **peu** de transformations
- Bons résultats sur de **petits corpus** d'entraînement
- Résout plusieurs problèmes.

Université de Montréal

Bibliographie

- **Eric Brill. *Automatic grammar induction and parsing free text: A transformation-based approach*. In Proceedings of the 31 st Annual Meeting of the Association for Computational Linguistics, Columbus, Ohio, 1993.**
- **Eric Brill and Philip Resnik. *A rule-based approach to prepositional phrase attachment disambiguation*. Proceedings of the 15th International Conference on Computational Linguistics, pages 1198—1204,1994.**
- **Brill, Eric. *Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part of Speech Tagging*. Computational Linguistics,1995.**
- **Eugene Charniak. *Statistical Language Learning*. MIT Press, Cambridge MA, 1993.**
- **Philippe Langlais, *Notes du cours ift6010***