



Ressources textuelles incrémentales pour la modélisation des interactions linguistiques multiples

Maria ZIMINA

CLILLAC-ARP, Université Paris Diderot

maria.zimina@eila.univ-paris-diderot.fr

Serge FLEURY

CLESTHIA, Université de la Sorbonne nouvelle – Paris 3

serge.fleury@univ-paris3.fr



Plan

- Linguistique appliquée & approches quantitatives
- Modèle de données **Trame/Cadre**
- **Ressources textuelles incrémentales :**

Démonstrations sur le treebank **Rhapsodie**



sous **Le Trameur**

pour identifier des marqueurs formels de genres



Linguistique appliquée

Acquisition du langage, Analyse contrastive, Analyse et production du signal de parole, Enseignement de la langue maternelle, Enseignement de langues étrangères, Français langue étrangère, Langage des enfants, Langues de spécialité, Langues régionales et langue nationale, Lexicologie et Lexicographie, Nouvelles technologies éducatives, Politique linguistique, Terminologie, Troubles dans l'acquisition du langage, Troubles du langage chez l'adulte, Traduction, Langues en contact et variation linguistique en rapport avec la situation géographique, sociale et professionnelle, Rééducation linguistique en rapport avec l'audition ou la production, Traitement automatique des langues naturelles (analyseurs automatiques, traduction automatique, gestion des données documentaires)...

<http://www.afla-asso.org/>

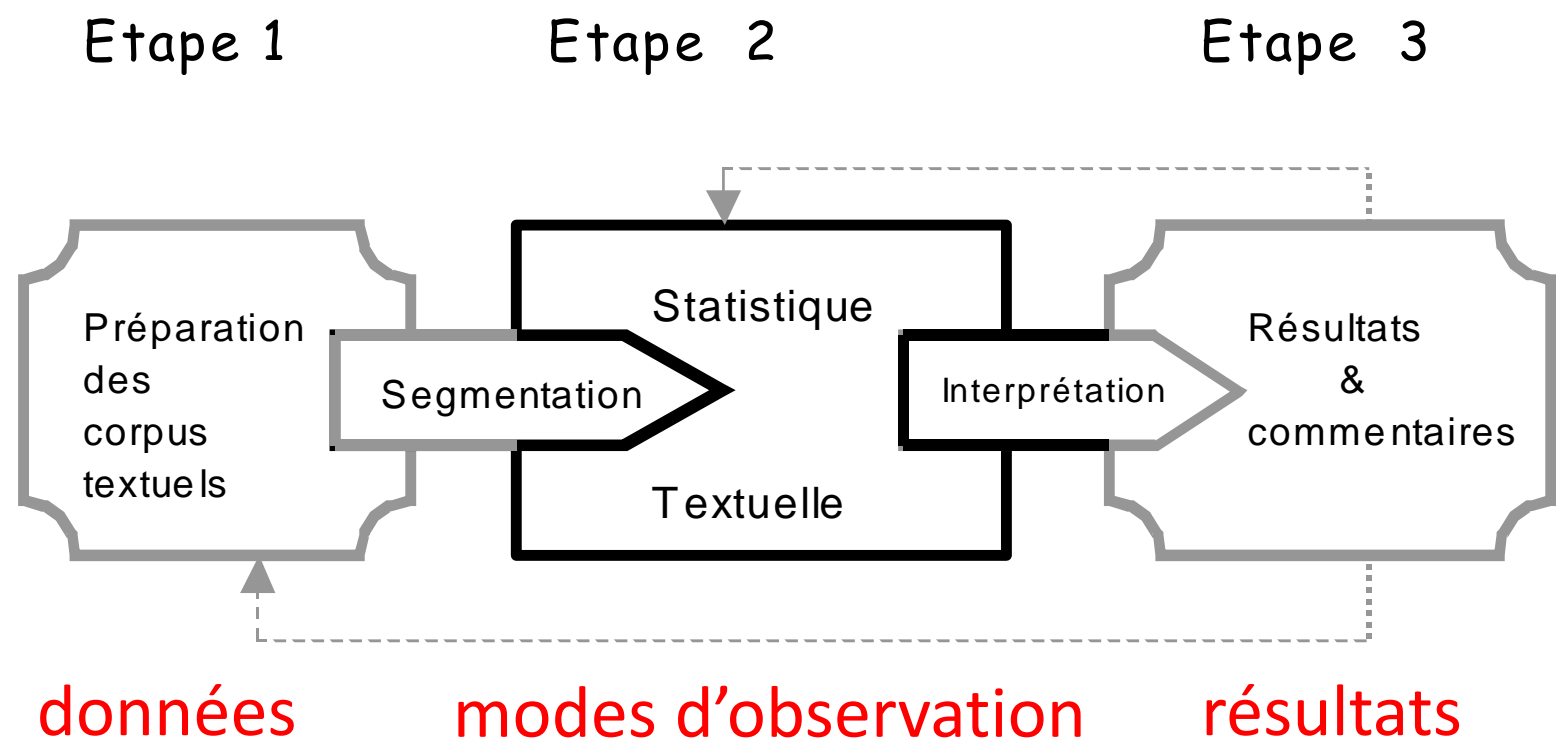
Objets d'études complexes...

Multiplés cadres théoriques

- les approches de la grammaire traditionnelle
- les procédures structuralistes
- les principes de la grammaire générative
- les approches de la linguistique cognitive
- La linguistique contrastive
- Les théories de la traduction
- L'analyse de discours
- ...

Annotations...

Approches quantitatives de corpus



Lamalle *et al.*, 2006

Architecture *Trame/Cadre*

Le texte dans le Trameur

1. Une segmentation
2. Systèmes de parties

on "découpe" des parts

Un texte est segmenté
en unités
(les mots en général)

Pour chaque mot

etc...

P.O.S

lemmes

Formes

On compte des « contenus » dans des « contenants »



Le Trameur

A Framework for Annotated Text Corpora Exploration

<http://www.tal.univ-paris3.fr/trameur/>

Texte (unités, délimiteurs) + parties

```

<part num=1>
x1 x2, x3 x2. x4
x5 . . . . xk §
<part num=2>
xk+1 x2, xk+2 x4.
x2 xk+3 . . . . xn
§
etc.

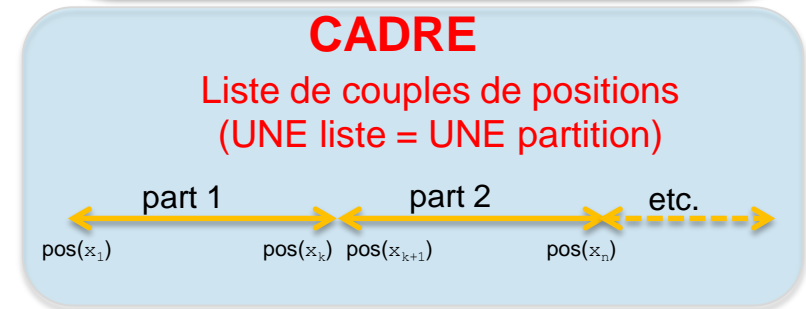
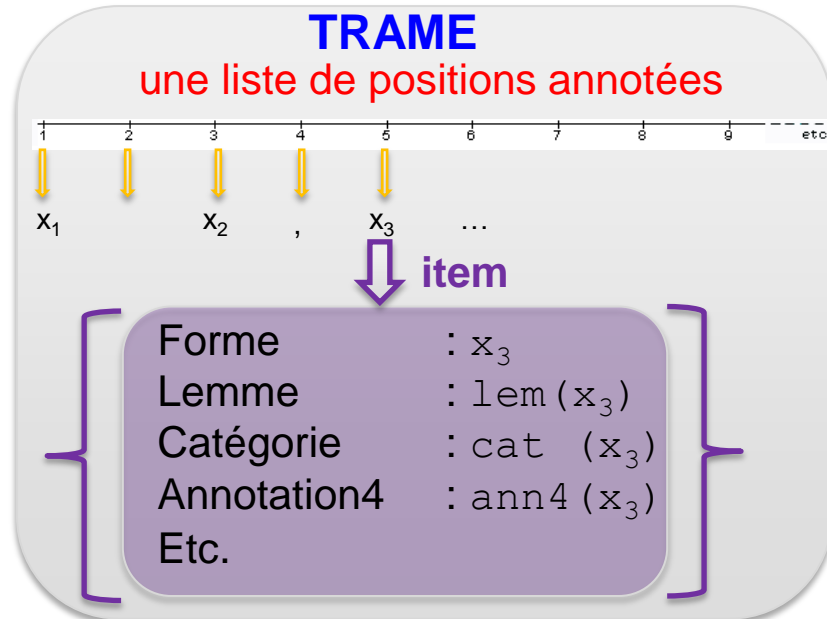
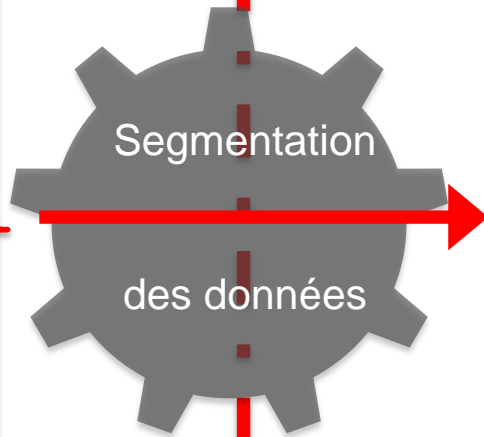
```

Délimiteurs

```

.,;! ? += ( ) § etc.

```





Le Treebank *Rhapsodie*

<http://www.projet-rhapsodie.fr/>

- Un corpus constitué de 57 échantillons de français parlé (5 minutes en moyenne), soit 3 heures de parole (33 000 mots, 89 locuteurs) munies d'une transcription orthographique et phonétique.
- Annotations **micro / macro syntaxique** et **prosodique**
- Développé avec comme objectif théorique la modélisation de l'interface prosodie, syntaxe, discours en français parlé.

[R^hapSodie]
corpus prosodique de référence
en français parlé

Rhapsodie sous Le Trameur

- Démonstration (v8 => 62 annotations)

Situation de communication => Marqueurs formels => Genres...

CADRE ACTIF : Clic sur partie : les formes | Clic-droit sur partie : les lemmes | Control+Clic sur partie : vocabulaire spécifique

ARBRE XML : Clic sur noeud : les formes

<input type="checkbox"/>	<SUBGENRE (PROCEDURAL [pos=557])>	
<input type="checkbox"/>	<INTERACTIVITY (NON-INTERACTIVE [pos=557])>	...
<input type="checkbox"/>	<SOCIAL_CONTEXT (PRIVATE [pos=557])>	...
<input type="checkbox"/>	<PLANNING_TYPE (SPONTANEOUS [pos=557])>	...
<input type="checkbox"/>	<TASK (TRAVEL-PLANNING [pos=557])>	...

Nb L. Partie : SUBGENRE=PROCEDURAL (557,1090) Annotation : 1 Aperçu : 50

7 tu **passes** un autre rond-point §

8 donc toujours tout droit # §

9 après tu **montes** une grande grande ligne droite # §

10 tu **passes** devant la piscine # §

11 euh il y a un stade aussi à côté §

12 et tu **arrives** à un rond-point donc là tu **rentres** dans le centre-ville # §

13 tu **arrives** à un rond-point §

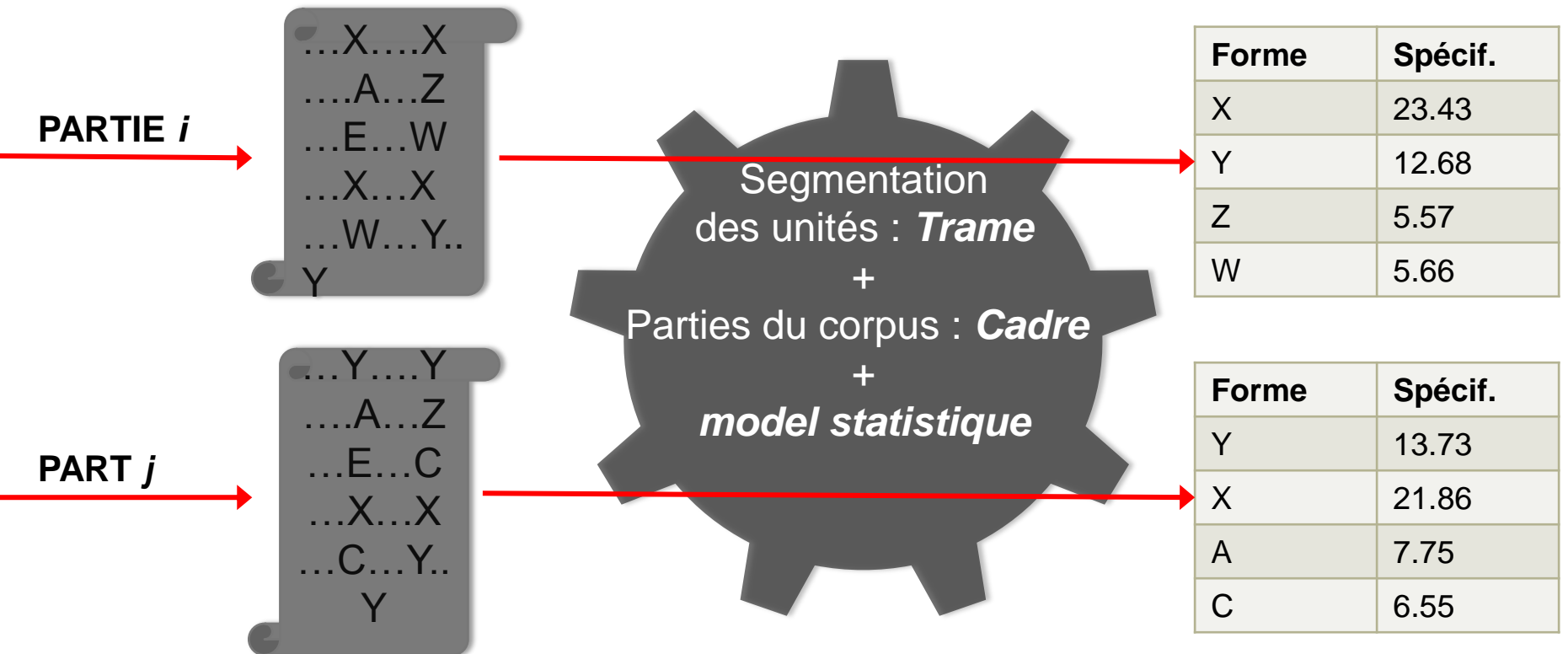
14 c'est donc à euh gauche # euh ensuite euh donc gauche §

15 là je **crois** que ça **sera** indiqué déjà gare # §

Principes d'analyse de données

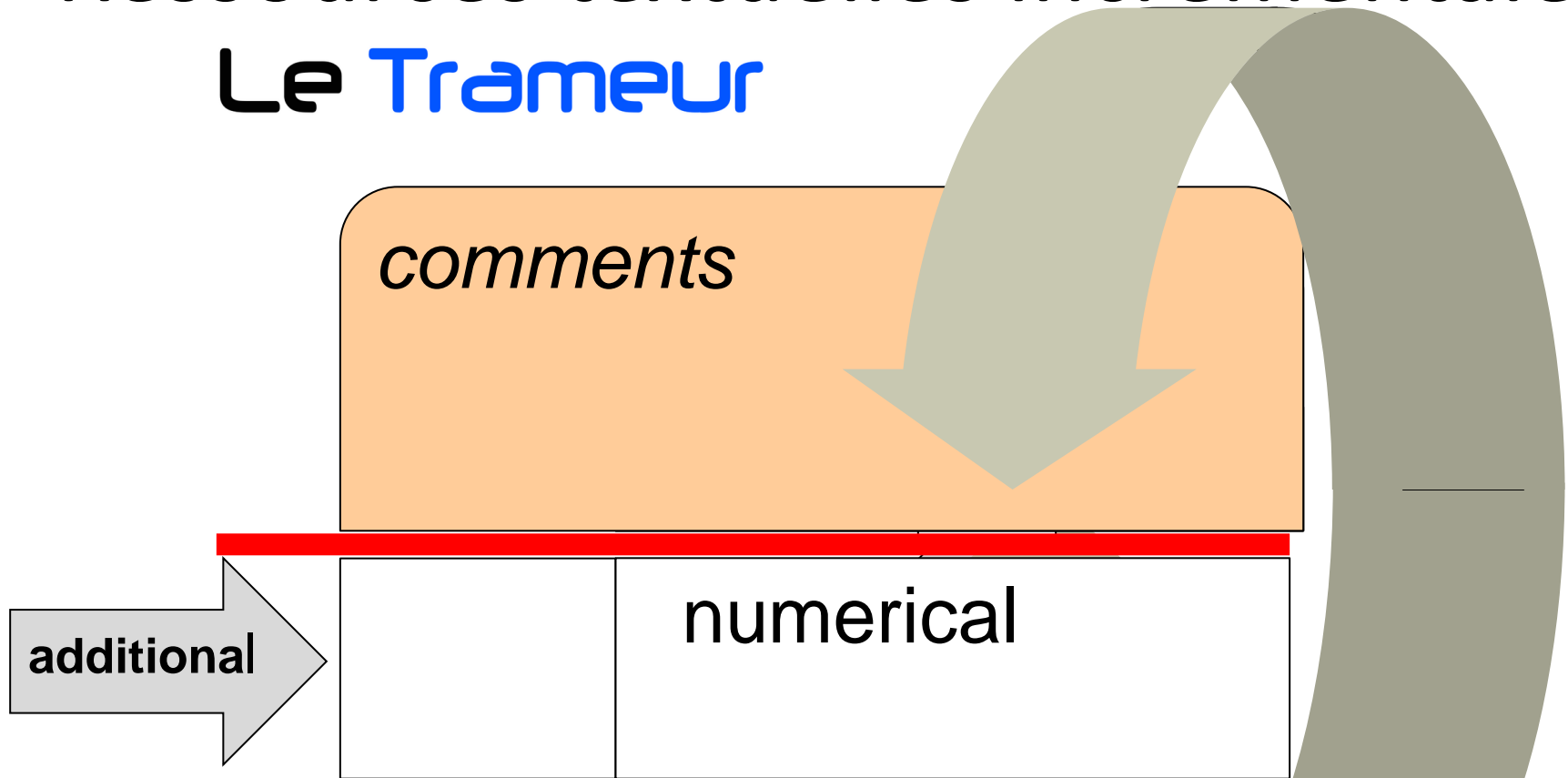
Moteur textométrique

Tableaux statistiques calculés dynamiquement



Ressources textuelles incrémentales

Le Trameur



André Salem 2011
(Présentation)



Conclusions

- Souplesse dans la **construction des parcours interprétatifs** dans les corpus avec des **annotations multiples** (treebanks).
- Flexibilité pour **l'analyse quantitative d'objets textuels complexes**
- Interface graphique, téléchargement libre :

Le Trameur



Références

- **S. Fleury 2013.** *Le Trameur. Propositions de description et d'implémentation des objets textométriques.* Sorbonne nouvelle – Paris 3.
En ligne : <http://www.tal.univ-paris3.fr/trameur/trameur-propositions-definitions-objets-textometriques.pdf>
- **S. Fleury and M. Zimina 2014.** «Trameur: A Framework for Annotated Text Corpora Exploration», *Proceedings of COLING 2014: System Demonstrations*, Dublin, 2014, pages 57-61.
- **C. Lamalle, S. Fleury, A. Salem 2006.** « Vers une description formelle des traitements textométriques ». *Journées Internationales d'Analyse Statistiques des Données Textuelles*, Besançon, 2006.
- **M. Zimina , S. Fleury 2014.** « Approche systémique de la résonance textuelle multilingue ». *Journées internationales d'Analyse statistique des Données Textuelles*, Paris, 2014.

<http://www.tal.univ-paris3.fr/trameur/>

<http://www.projet-rhapsodie.fr/>