

# Corpus parallèles et textométrie

---

Maria ZIMINA

EA 2290 SYLED

Université de la Sorbonne nouvelle – Paris 3

[zimina@msh-paris.fr](mailto:zimina@msh-paris.fr)

# Textométrie parallèle

- Les méthodes quantitatives apportent un éclairage précieux sur des régularités dans la structuration de discours parallèles.
- L'approche textométrique de corpus parallèles permet d'envisager la conception de nouvelles procédures informatisées pour le traitement automatique de l'information bi-textuelle.

# Approches quantitatives de corpus de traduction

## Types de tâches

Extraction de ressources traductionnelles

Alignement de corpus

Vérification de l'alignement

Synthèse de l'information bi-textuelle

Navigation dans le bi-texte

## Méthodes privilégiées

Segments répétés

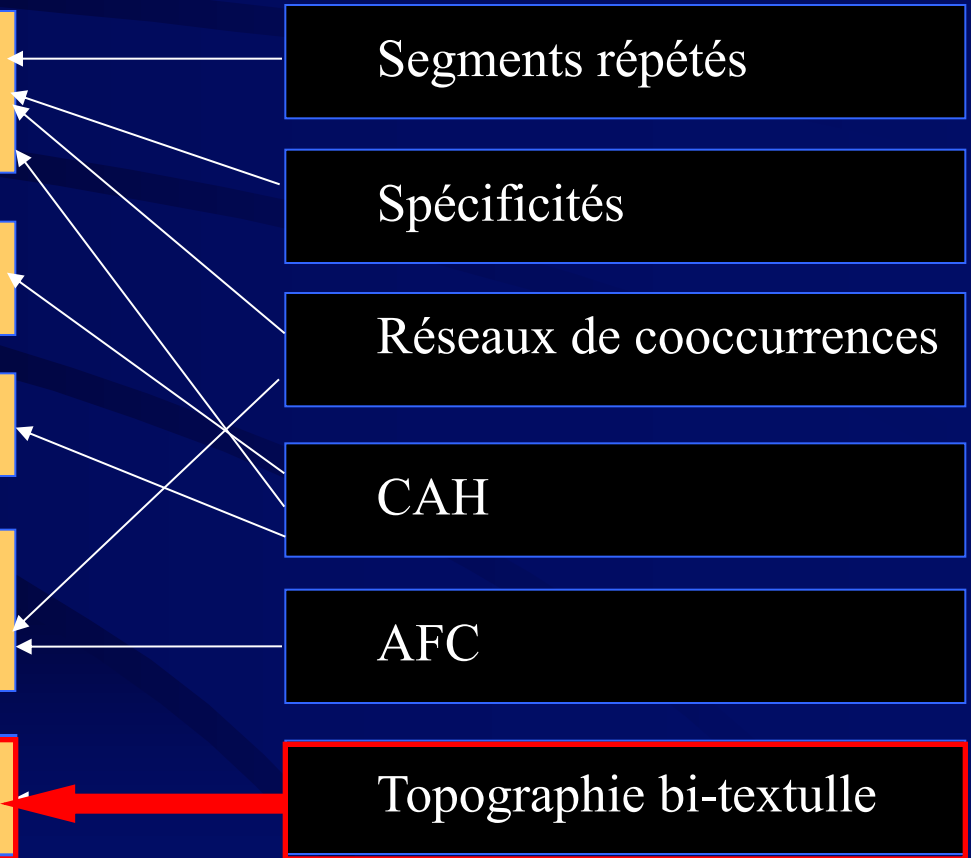
Spécificités

Réseaux de cooccurrences

CAH

AFC

Topographie bi-textuelle



# Exemple : la carte bi-textuelle (i)

volet français

volet anglais

eu égard à l'enjeu du litige pour l'intéressé, et même si les procédures devant la cour d'appel et la cour de cassation prises isolément ne paraissent pas excessivement longues, un délai global d'environ onze ans et huit mois ne saurait passer pour raisonnable.

Section : [icon] [icon]

Occurrence : [icon] [icon]

Section [icon]

regard being had to the importance of what was at stake for the applicant, and even though the proceedings in the court of appeal and the court of cassation, taken separately, do not appear excessively long, a total lapse of time of approximately eleven years and eight months cannot be regarded as reasonable.

Section : [icon] [icon]

Occurrence : [icon] [icon]

Section [icon]

[Grid of small squares representing text]

[Grid of small squares representing text]

[Grid of small squares representing text]

[Grid of small squares representing text]

[Grid of small squares representing text]

[Grid of small squares representing text]

47 sections

Spécificités  positives  négatives

Terme	Frq Tot.	Frq...	Sp...
fonctionnaires	49	49	109
les fonctionnaires	14	14	31
des fonctionnaires	14	14	31
de loyauté	36	14	22
loyauté	42	14	21
de loyauté politique	22	10	17
loyauté politique	24	10	16
de fonctionnaires	7	7	16
obligation de loyauté politique	15	8	14
obligation de loyauté	21	8	13

47 sections

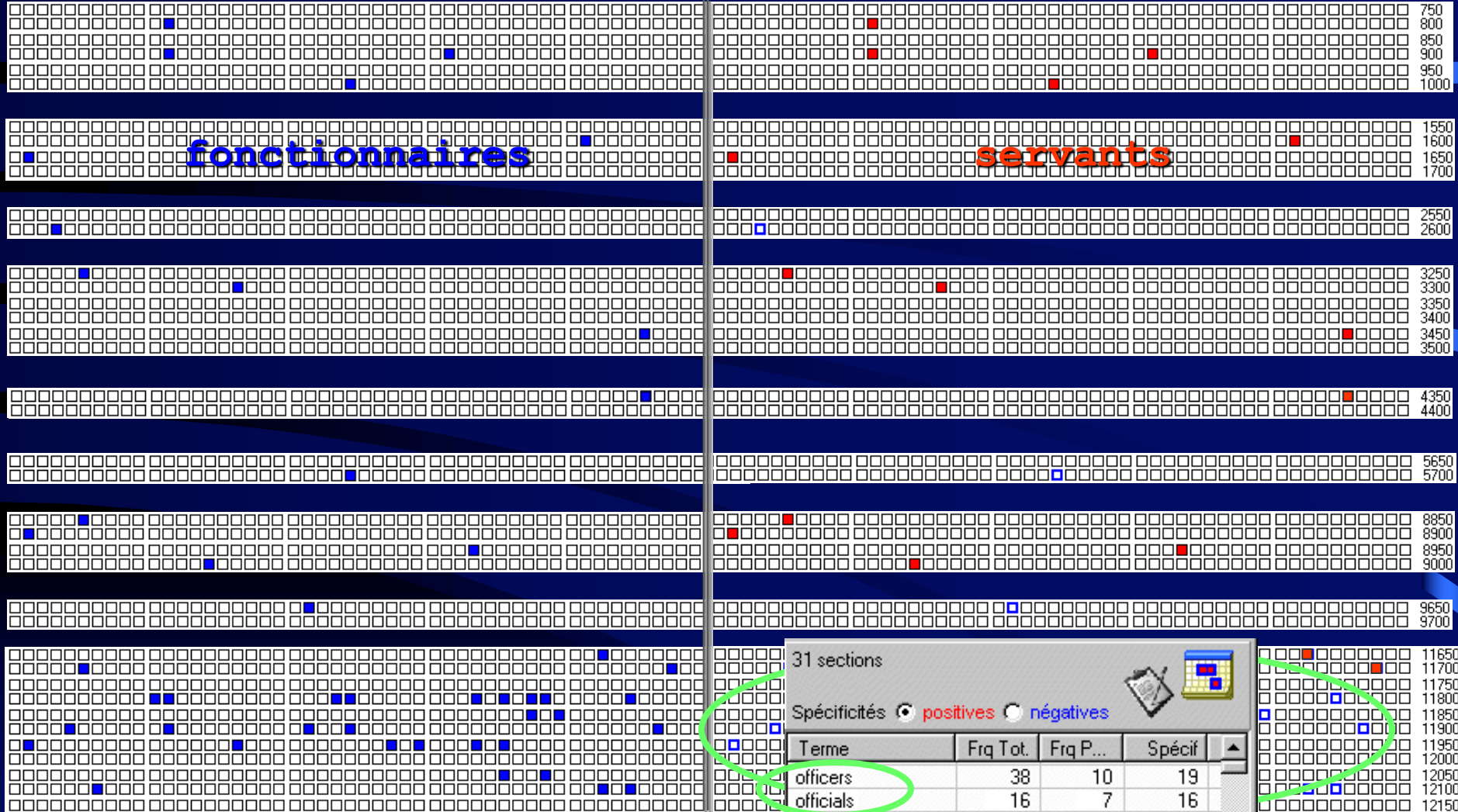
Spécificités  positives  négatives

Terme	Frq Tot.	Frq...	Sp...
servants	50	31	55
civil servants	46	29	52
civil	304	41	40
loyalty	43	14	20
duty	109	15	16
political loyalty	25	10	16
of political	29	10	15
duty of	45	11	15
officers	38	10	14
duty of political loyalty	23	9	14
service	110	14	14
civil service	58	11	13

# Exemple : la carte bi-textuelle (ii)

volet français

volet anglais



# Perspectives ouvertes par l'approche topographique du **bi-texte**

- Navigation interactive dans les textes parallèles ;
- Repérages automatiques d'équivalences de traduction ;
- Découverte de structures lexicales parallèles ;
- Synthèse d'information bi-textuelle ;
- ...

# Pour en savoir plus...

- Lamalle C., Martinez W., Fleury S., Salem A. et al. (2003). *Lexico3 – Outils de statistique textuelle*. Syled-CLA<sup>2</sup>T, Université de la Sorbonne nouvelle – Paris 3 :  
<http://www.cavi.univ-paris3.fr/Ilpga/ilpga/tal/lexicoWWW>
- Martinez W., Zimina M. (2002). Utilisation de la méthode des cooccurrences pour l'alignement des mots de textes bilingues. In *Actes des JADT'02*, Saint-Malo.
- Zimina M. (2002). Repérages lexicométriques des équivalences à basse fréquence dans les corpus bilingues. In Véronis J. (Ed.), *Lexicometrica*, n° spécial Corpus alignés :  
<http://www.cavi.univ-paris3.fr/lexicometrica/thema/thema6.htm>
- Zimina M. (2000). Alignement de textes bilingues par classification ascendante hiérarchique. In *Actes des JADT'00*, Lausanne.