

# Modélisation textométrique des textes

*Text modelisation in  
textometry*

Bénédicte PINCEMIN  
ICAR, CNRS & Université de Lyon

*9<sup>èmes</sup> Journées internationales d'Analyse statistique des Données Textuelles (JADT 2008)  
Lyon, 12-14 mars 2008*

# Theory - Textometric procedures need :

UNITS

PARTS

obtained by

DELIMITATION

(what is the item composed of)

IDENTIFICATION (which type the item is an instance of)

## Implementations :

Déroulement syntagmatique du texte (positions)

	→				
<i>forme</i>	.	<b>ouvrirait</b>	<b>les</b>	<b>vannes</b>	.
<i>lemme</i>	.	verbe	article	subst.	.
<i>catégorie</i>	:	ouvrir	le	vanne	:
<i>id. texte</i>	:	37	37	37	:
	.	.	.	.	.
	.	.	.	.	.

Figure 1 : Weblex - Text as Table

Fréqtot Gramr

1	Con	<b>donc</b>
2	Proper	<b>je</b>
1	Vconj	<b>pense</b>
1	Vconj	<b>suis</b>

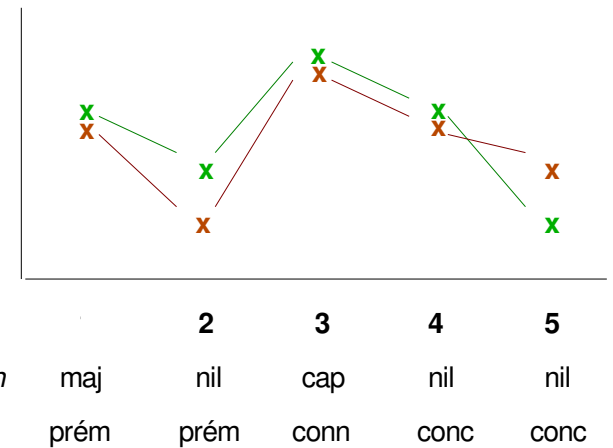


Figure 2 : SATO - Text as Graph

## Some proposals :

- more flexibility :
  - defining new types by combination of elementary types
  - for each stage of the analysis, a different type set could be considered
- unification of DELIMITATION and IDENTIFICATION procedures through tags

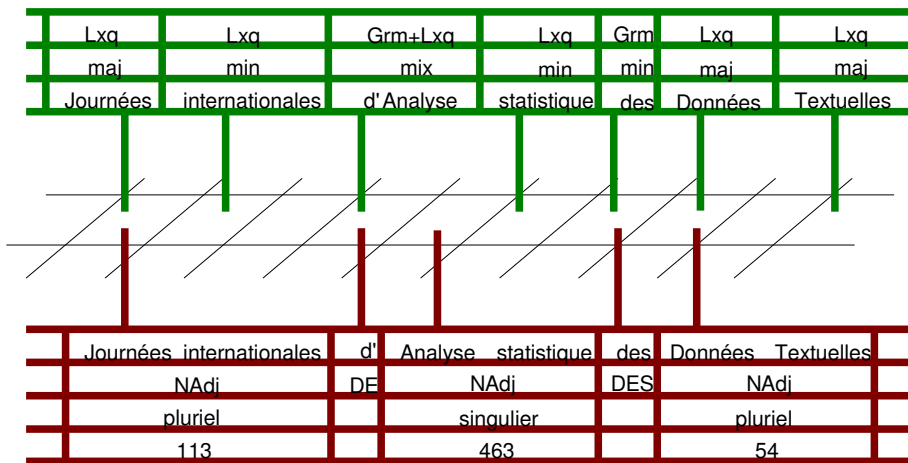


Figure 3 : Le Trameur – Text as a Weaving

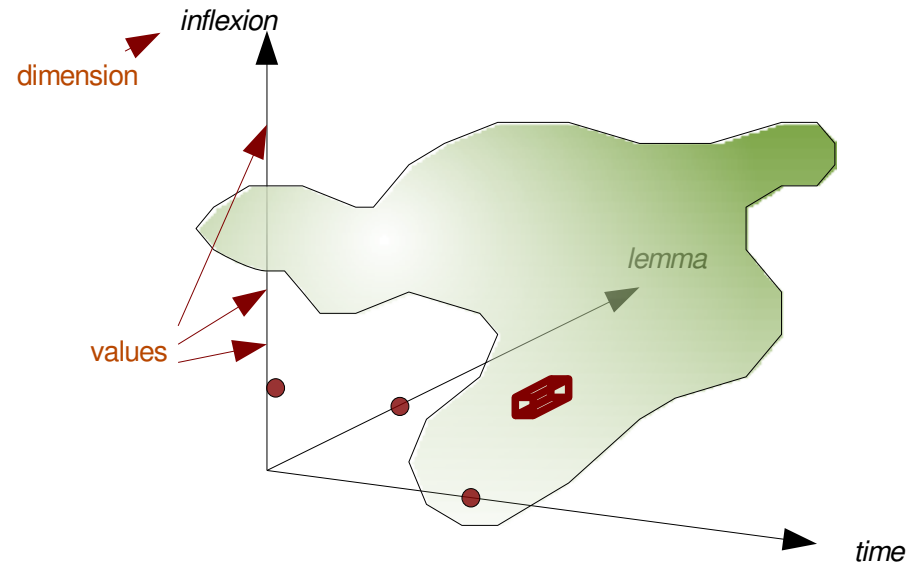


Figure 4 : Proposal – Text as shape