

L'analyse de corpus hétérogènes

*Séminaire Pratiques de corpus numériques :
enjeux et méthodes interdisciplinaires*

Magali BIGEY

ELLIADD-ADCOST EA 4661
Université de Franche-Comté

Trois types de corpus

- Corpus littéraire de 50 romans
- Corpus issu du web : analyse de réception de la lecture
- Corpus d'entretiens sur la mobilité résidentielle



Corpus littéraire

Analyse par thématiques, avec le calcul des spécificités

Analyse avec des dictionnaires électroniques spécifiques

- dictionnaire des verbes de sentiments
- dictionnaire des parties du corps

Travail sur les concordances, puis sur les segments
répétés des concordances

Analyse avec l'Analyse Factorielle des Correspondances

Analyse par thématiques

- Recherche des spécificités :

The screenshot displays the Lexico3 software interface. The main window shows a list of terms with their total frequency, frequency in the selected part, and a specificity indicator. A secondary window, titled 'Principales caractéristiques de la partition : T', provides a summary of the partition's statistics and a table of the most frequent terms.

Principales caractéristiques de la partition : T

Nombre d'occurrences: 2085644 Nombre de formes: 4890
Nombre d'hapax: 17552 Fréquence maximale: 7768

Num	Partie	Occurenc	Form
✓ 1	01	41805	614
✓ 2	02	41268	560
✓ 3	03	36223	465
✓ 4	04	40421	571
✓ 5	05	40993	601
✓ 6	06	43185	603
✓ 7	07	36939	582
✓ 8	08	44731	617
✓ 9	09	36420	620
✓ 10	10	42968	668
✓ 11	11	39590	672
✓ 12	12	42105	631
✓ 13	13	40076	660
✓ 14	14	40288	587
✓ 15	15	35340	677

Princ. Caract. Lexicométriques

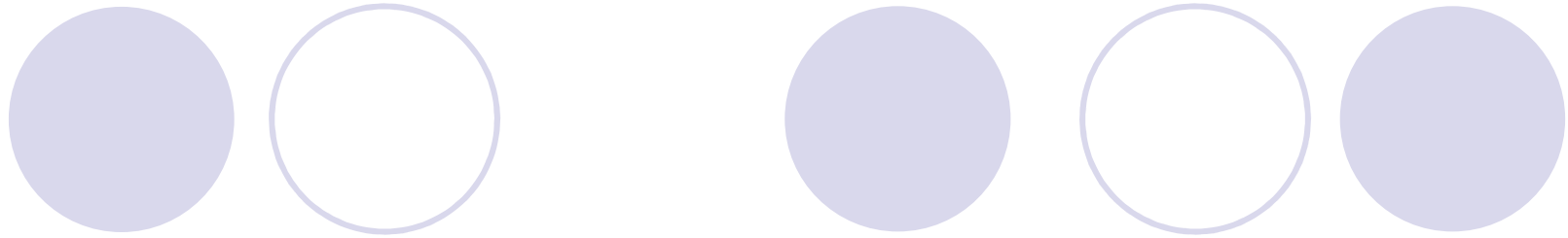
Dictionnaires électroniques spécifiques

- un dictionnaire des verbes de sentiments

Sujet				V "concret"	Adjectif				Comp. direct				
N _{hum}	N _{nr}	le fait O <u>u</u> P	V ¹ O		N ₀ V	a = ant	a = able	a = ix	N _{hum}	N _{hum}	le fait O <u>u</u> P	N ₁ se V de ce O <u>u</u> P	N ₁ se V après de N _{hum} de ce O <u>u</u> P
+	+	+	+	-	+	+	-	-	+	-	-	-	-
+	+	+	+	+	+	-	-	-	+	-	+	-	-
+	+	+	+	+	+	-	-	-	+	-	-	-	-
+	+	+	+	-	+	+	-	-	+	-	-	-	-
-	+	+	+	+	+	+	-	-	+	-	-	-	-
+	+	+	+	-	+	-	-	-	+	-	-	-	-
+	+	+	+	+	+	-	-	-	+	-	-	-	-

```

dic Vsent - Bloc-notes
Fichier Edition Format Affichage ?
abasourdir, v+sent+4
abattre, v+sent+4|
abîmer, v+sent+4
abrutir, v+sent+4
absorber, v+sent+4
abuser, v+sent+4
accaparer, v+sent+4
accrocher, v+sent+4
activer, v+sent+4
adoucir, v+sent+4
affadir, v+sent+4
affaiblir, v+sent+4
affecter, v+sent+4
  
```



- Analyse générale et retour au texte pour l'interprétation.
- Analyse des concordances du dictionnaire des sentiments

Concordance des verbes de sentiment

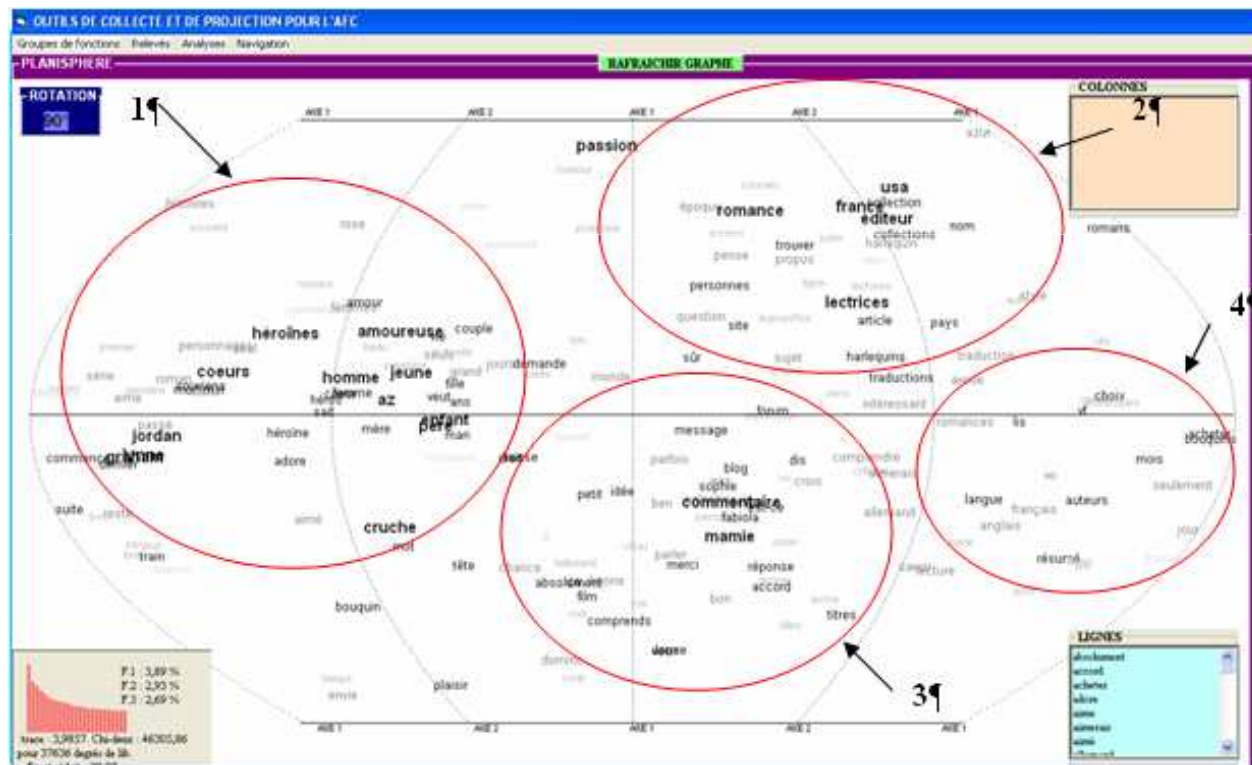
The screenshot shows the Concord VSENT 10 mots software interface. The main window displays a concordance search for the word 'aim'. The search parameters are set to 'characters' (selected), 'before, and 5 after', and 'Display: Matches' (checked). The results are displayed in a table with columns for 'Text', 'Before', 'Seq.', and 'After'.

Text	Before	Seq.	After
	chacun a le droit d'	aim	er ce qu'il veut, ce
	sûre de l'orthographe), ils	aim	ent et c'est leur droit
	n'est meilleure façon d'	aim	er toute chose que se dire
	en fait. Bien que n'	aim	ant guère la chick lit, j
	n'est meilleure façon d'	aim	er toute chose que se dire
	j'expliquais que je n'	aim	ais pas la chick lit, à
<TU>	Maintenant, pourquoi je n'	aim	e pas la chick lit commerciale
	n'est meilleure façon d'	aim	er toute chose que se dire
	une journée de boulot, j'	aim	e assez me délasser devant un
que vous estimez "sociologiquement intéressant"		aim	ons lire les romans sentimentaux, un
<TU>	sérieux, la demoiselle n'	aim	e pas harlequin bah c'est
	peine de traiter ceux qui	aim	ent de tarés ou autres. </TU>
	critique sur notre manière d'	aim	er ces livres, le forum pour
<TU>	bref oui moi j'	aim	e lire les harlequin et pourtant
	'une sardine. mais oui j'	aim	e ce genre de livre et

Query 46/46

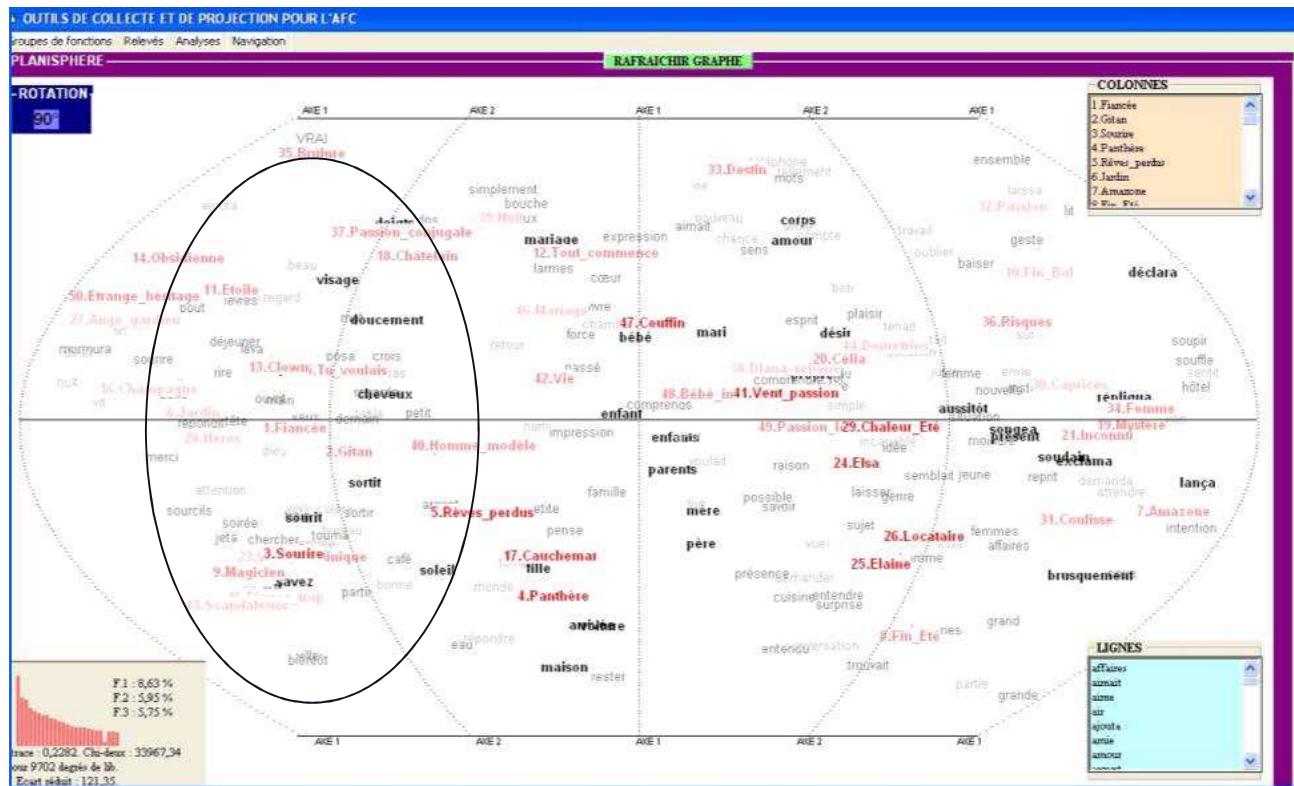
Corpus issu du web

- Réception de la lecture : A.F.C.

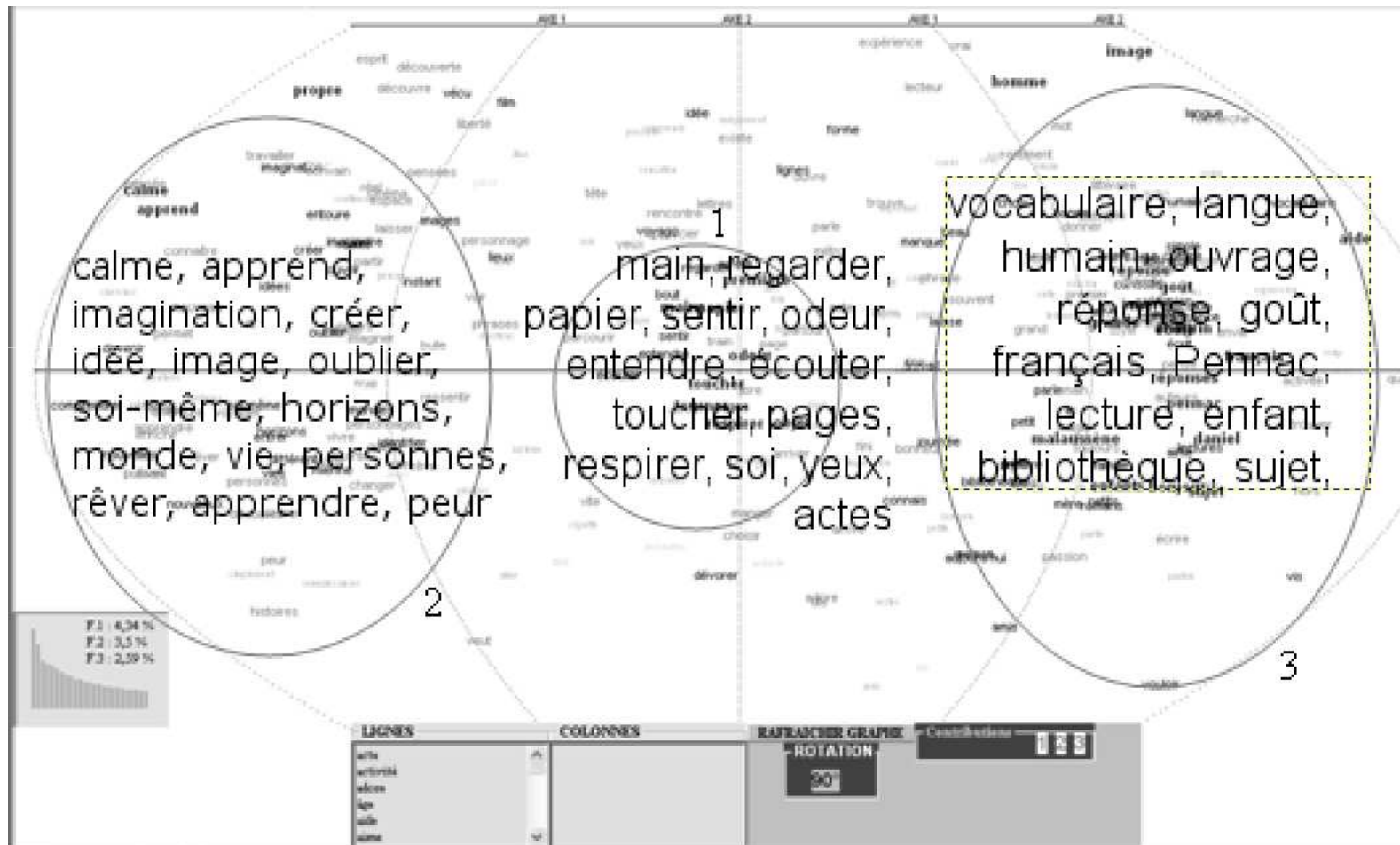


Corpus issu du web

- Réception de la lecture : A.F.C.



Corpus issu du web



Corpus d'entretiens

- Corpus d'entretiens dans le cadre d'une ANR :
Questionnement sur la mobilité résidentielle :
Base de travail : entretiens en face à face de deux heures.

Etape préparatoire : étudier ces 12 entretiens pour écrire le questionnaire final : travailler sur les concordances.

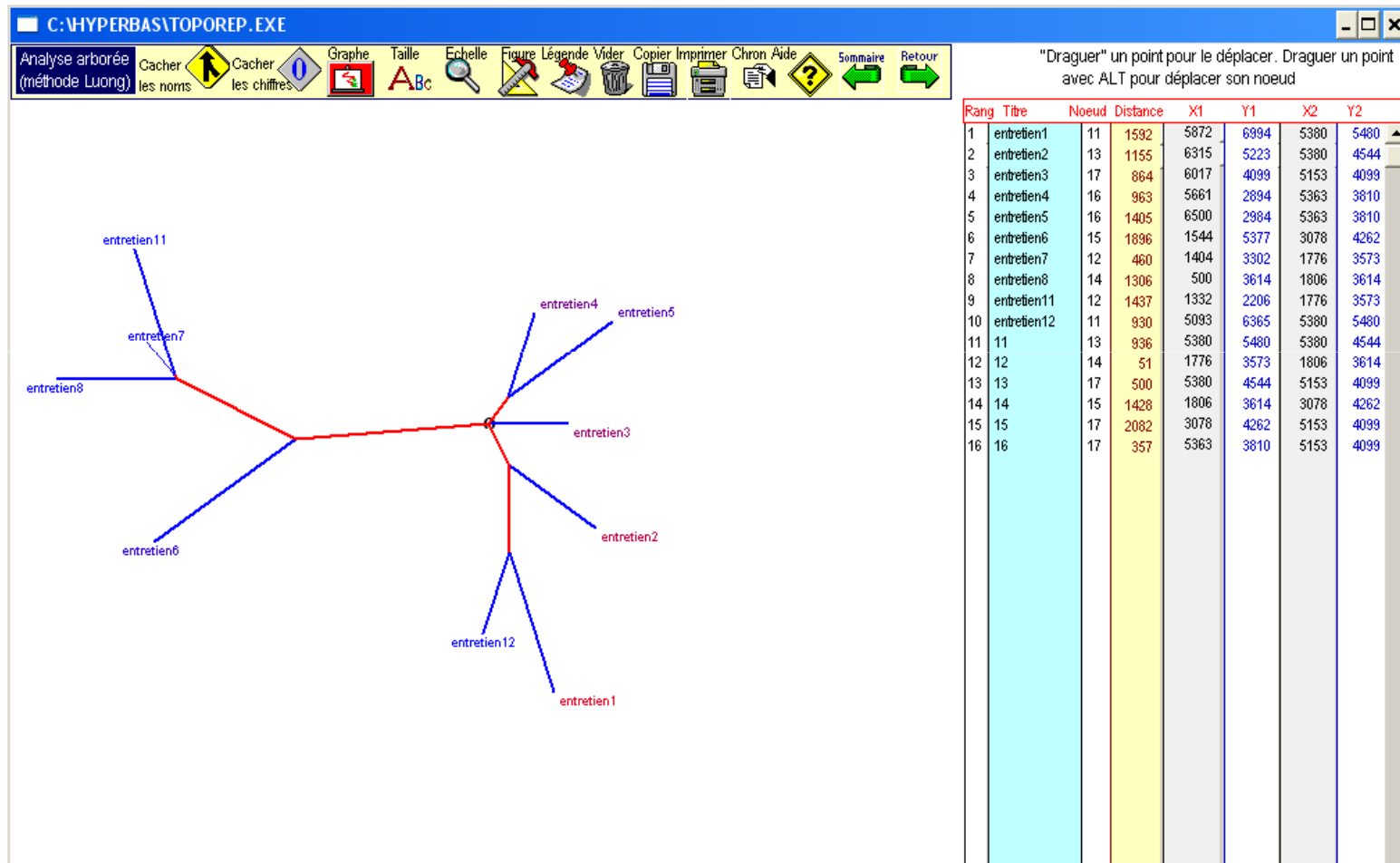
Lors de la transcription des entretiens, remplacer tous les termes de lieux par des termes génériques non ambigus, et tous les termes de temps de la même manière (ex : locus _ tempus)

Corpus d'entretiens

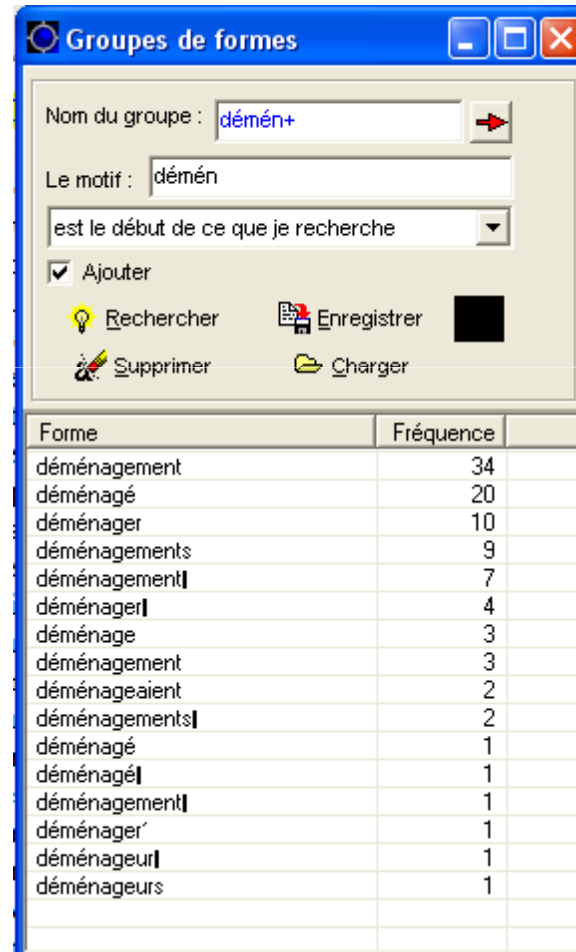
- analyse de concordances :

Puis concordances de concordances (plus de précision qu'AFC, car corpus d'une taille où le retour au texte est aisé)

Distribution des textes






Les groupes de formes





Nom du groupe : →

Le motif :

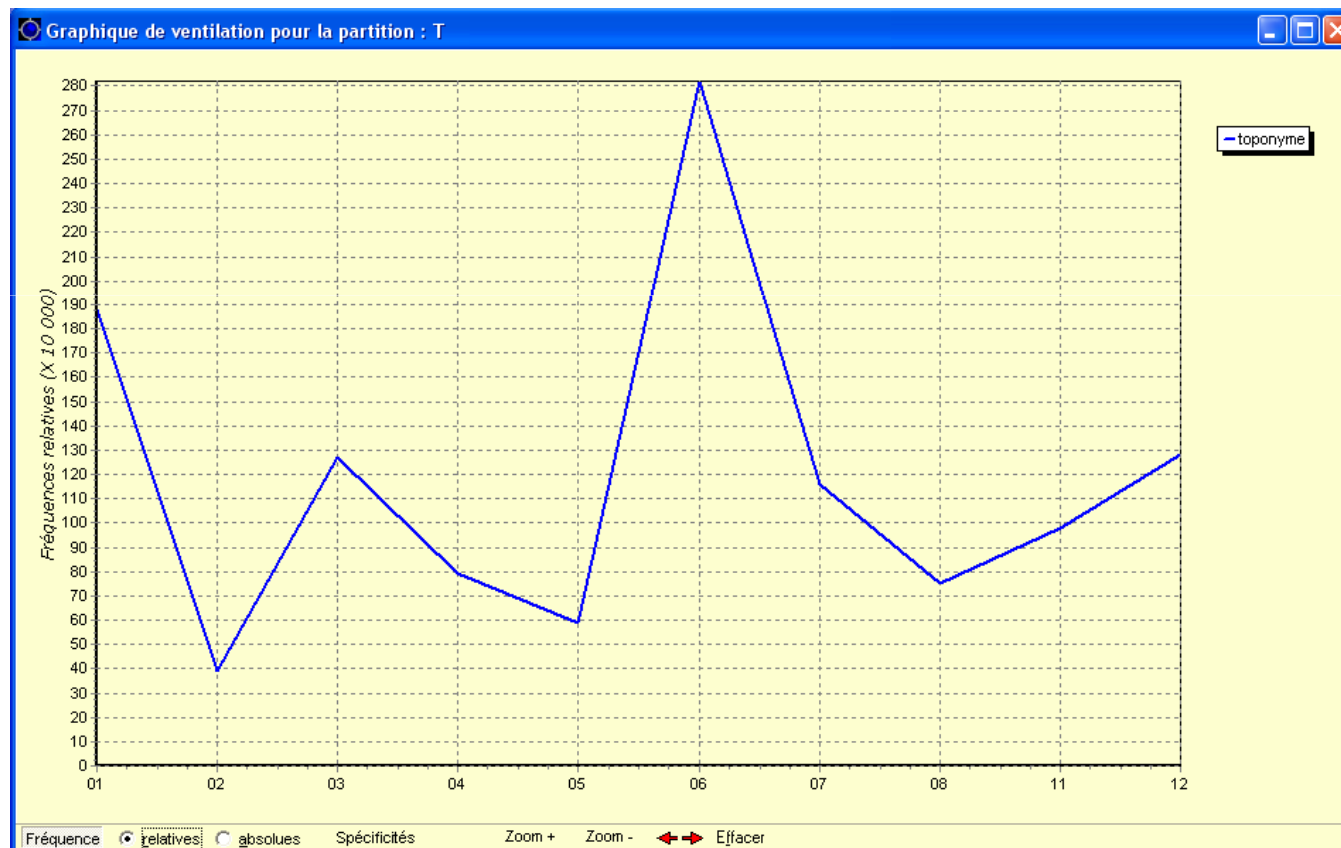
Ajouter

 Rechercher  Enregistrer 

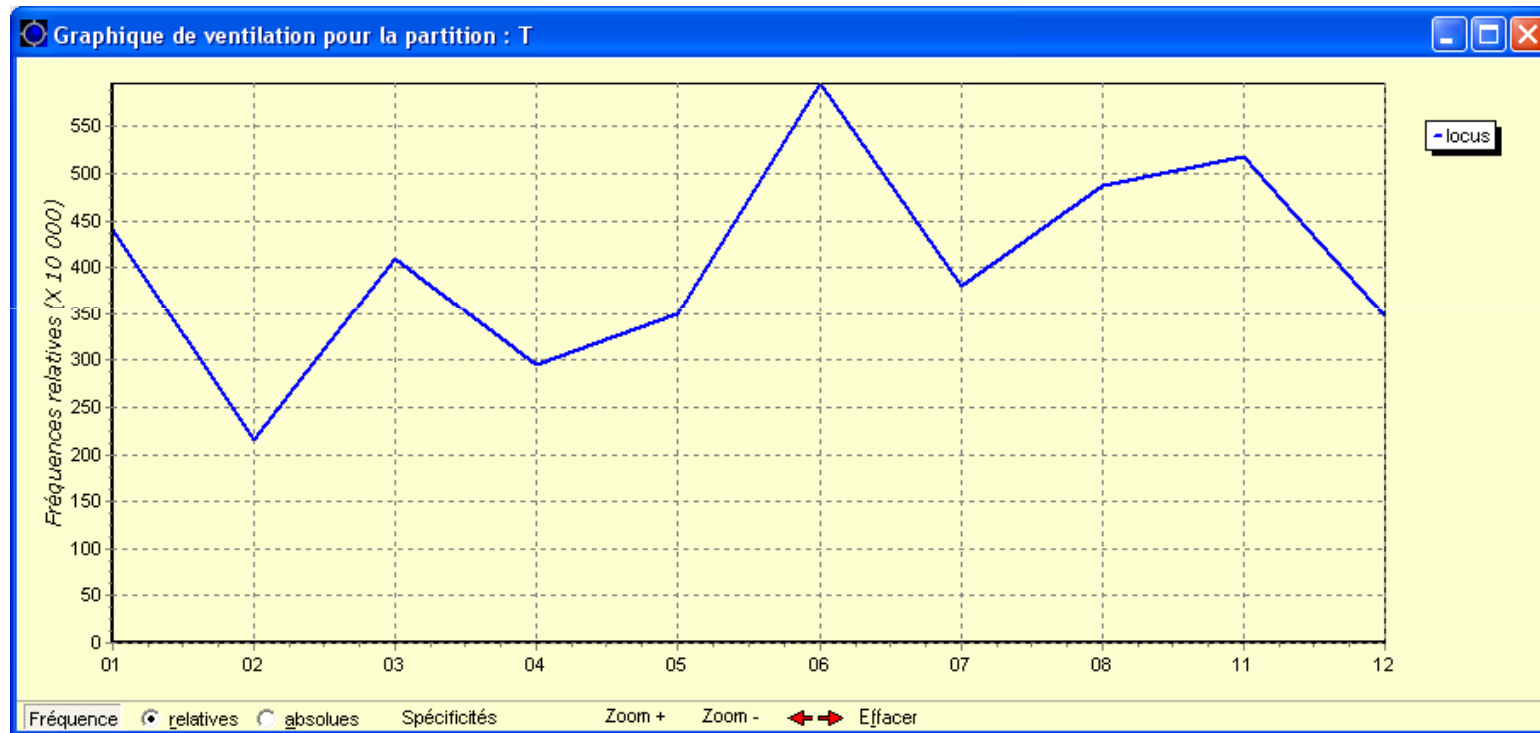
 Supprimer  Charger

Forme	Fréquence
déménagement	34
déménagé	20
déménager	10
déménagements	9
déménagement	7
déménager	4
déménage	3
déménagement	3
déménageaient	2
déménagements	2
déménagé	1
déménagé	1
déménagement	1
déménager'	1
déménageur	1
déménageurs	1

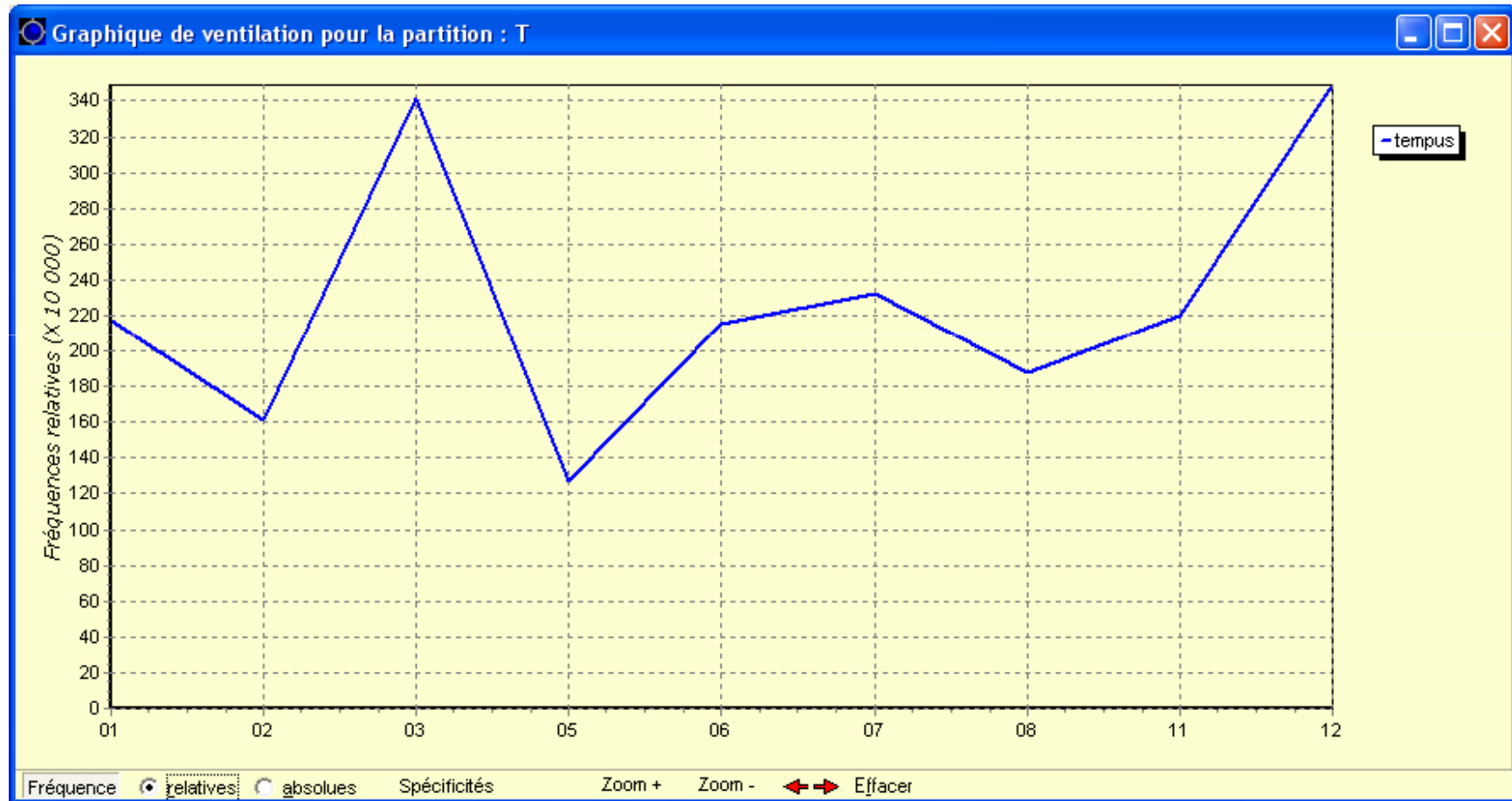
Graphique de ventilation mot *toponyme*



Devant les mots-outils, on trouve *locus*



Tempus, 3^e position après de et je





● Merci de votre attention