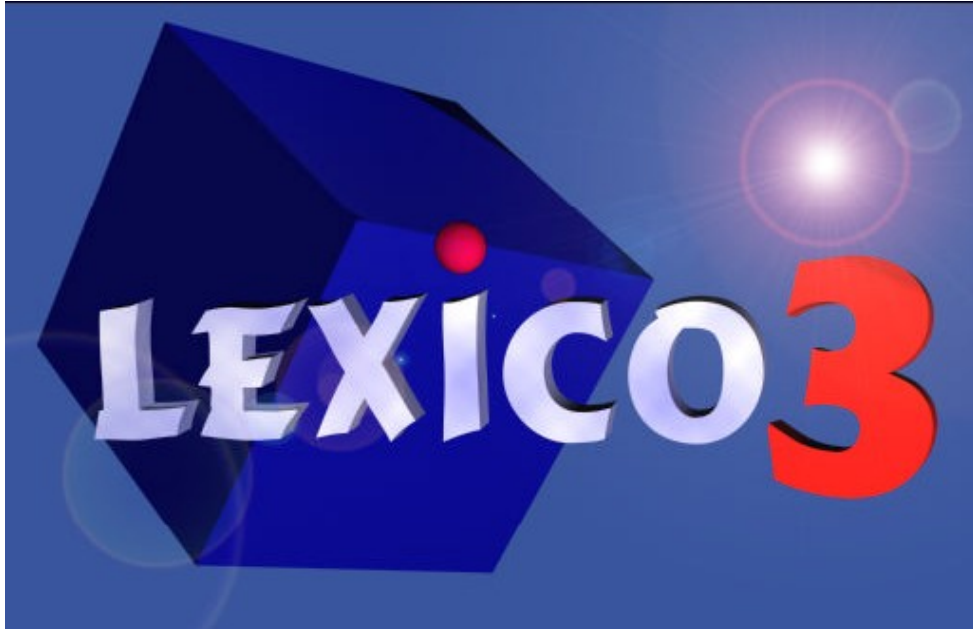


SYLED - CLA2T



Universidade la Sorbonne Nouvelle - Paris 3

Ferramentas de estatística textual

Cédric Lamalle
William Martinez
Serge Fleury
André Salem

Manual resumido de utilização

(Dez primeiros passos com *Lexico3*)

Andrea Kuncova
Aude Maisondieu

Tradução
Dirceu Cleber Conde
(FAFIJAN-UEL-CAPES)

Apresentação

O *Lexico3* é a edição 2001 do sistema *Lexico* cuja primeira versão data de 1990. Suas funções foram melhoradas e a originalidade desse sistema reside na possibilidade que o usuário tem de armazenar os resultados de sua análise lexicométrica desde a segmentação até a edição dos resultados finais. A facilidade de visualizar os dados e de criar diferentes composições, desde a mais simples análise estatística até os cruzamentos de dados fatoriais permite ao pesquisador, cujo objeto é o texto, avançar sobre hipóteses quantificando e qualificando seus dados.

O sistema *Lexico3* ainda não foi traduzido em língua portuguesa; assim, apresentamos apenas os passos para a sua utilização inicial. O manual completo encontra-se em sua língua original no site

<http://www.cavi.univ-paris3.fr/Ilpga/ilpga/tal/lexicoWWW/manuels.htmwww>

E a versão gratuita do software pode ser baixada a partir do endereço:

<http://www.cavi.univ-paris3.fr/Ilpga/ilpga/tal/lexicoWWW/lexico3.htm>

Agradecemos à equipe do SYLED (Systèmes Linguistiques, Enonciation et Discursivité), nas pessoas de seu Diretor, Professor André Salem, e de seu Diretor-Adjunto, Professor Bernard Bosredon. Ao primeiro, por ter nos ensinado a manusear o sistema e ter gentilmente autorizado esta tradução e ao Professor Bosredon, por ser um dos incentivadores de nosso contato com a Textometria e a aproximação entre discurso e materialidade.

Introdução

Originalmente o *Manual Resumido de Utilização* não possui alguns itens que julgamos necessários ao bom andamento dos trabalhos e, por isso, acrescentamos algumas dicas práticas. Assim, teremos uma primeira parte referente à instalação do sistema, uma segunda que orienta como fazer a “balizagem” do *corpus* e finalmente a terceira parte que comporta os 10 primeiros para exploração do *Lexico3* com o maior número de ilustrações possíveis.

Por uma questão de comodidade, usaremos os exemplos do corpus “**Le Père Duchesne**” que já vem com o *Lexico3*, e faz parte de uma das explorações sobre *corpus* com o sistema e que podem ser lidas a partir do sítio

<http://www.cavi.univ-paris3.fr/Ilpga/ilpga/tal/lexicoWWW/rapportsL3.htm>

o qual aconselhamos a visita para maiores informações sobre experiências e métodos de textometria e lexicometria empregados sobre diferentes *copora*.

PARTE I – Instalação

1. Advertência

Apesar de todo o cuidado que tivemos em preparar esta versão, alguns erros ainda subsistem. Nós solicitamos que eventuais erros e anomalias sejam encaminhados a:

Lexico3 / ILPGA
19, rue des Bernardins
75005 Paris, France

Junto com o seu encaminhamento, adicione o *corpus* sobre o qual tenha constatado a disfunção bem como o arquivo *atrace.txt* criado no momento da exploração (que se encontra na mesma pasta onde está armazenado o corpus de análise).

2. Configuração mínima

Windows 95
Processador 486, 4Me de memória RAM
3 Me livres no disco rígido

2.1 Configuração aconselhada

Windows 98
Pentium II, 8Me de memória RAM
3 Me livres no disco rígido

O **Lexico3** funciona em Windows 95, Windows NT 3.51 e 4.0.

Aconselhamos, para o bom andamento dos trabalhos, que o programa e o *corpus* estejam armazenados no disco rígido na mesma seção, ou seja, na mesma pasta.

3. Instalação do sistema

Para instalar o **Lexico3** :

- a) insira o CD-ROM no computador ou faça o download do sistema através do sítio recomendado na apresentação;
- b) Dê duplo-clique sobre o ícone do arquivo **SETUP.EXE** que se encontra no CD-ROM, ou na pasta onde o arquivo foi salvo (observar que o arquivo será transmitido sob uma forma compacta – “zipado”)
- c) Siga as instruções do programa de instalação;

PARTE II – Segmentação do *corpus*

Para que o sistema possa reconhecer as partes do *corpus*, é preciso estabelecer as repartições ou balizas que se compõem de uma estrutura simples : **senal menor, critério, sinal de igual, item do critério, sinal de maior**, como por exemplo

**<ano=1995>
<semana=34>**

Cada baliza terá seu alcance no *corpus* até a próxima baliza do mesmo critério. Assim, ela estabelece intervalos que serão extremamente importantes para o desenvolvimento das análises, pois é a partir desses critérios que se podem estabelecer comparações.

Obs.1 : não pode haver espaços no interior das balizas, como por exemplo

<idade= entre 21 e 25>.

Aconselha-se que os termos sejam, de preferência, os mais simples, e que obedçam a uma ordem numérica ou alfabética, da seguinte forma:

<idade=21-25>, ou então,
<idade=A>

Obs.2 : as balizas podem coexistir próximas com ou sem intervalo: **<sexo=M> <idade=A> <escolaridade=3>**

Obs. 3 : depois de estabelecidas as balizas, os arquivos devem ser salvos como “texto sem formatação”; todas as figuras devem ser retiradas e as tabelas convertidas em textos.

Após esse tratamento, o texto terá a seguinte aparência:

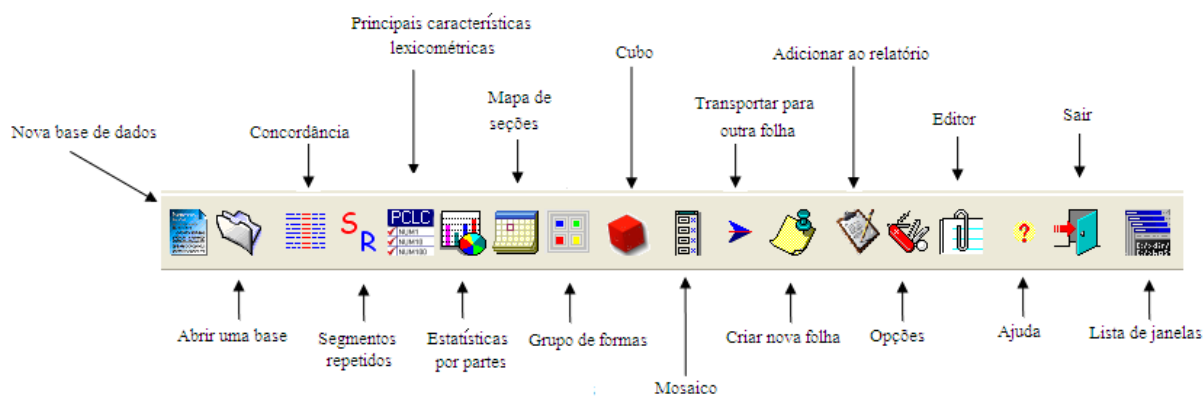
**<mois=01><quinzaine=11><semaine=111>
<Sda=1793><numero=260><edito=0><Epg=1> <Sat=0> § la grande douleur du
*père *duchesne au sujet de la mort de *marat assassiné à coups de couteau par
une garce du *calvados , dont l ' évêque *fauchet était le directeur . ses bons avis
aux braves *sans - culottes pour qu ' ils se tiennent sur leurs gardes , attendu qu ' il
y a dans *paris plusieurs milliers de tondus de la *vendée qui ont la patte graissée
pour égorger tous les bons citoyens**

Dicas:

1. Delimite os “parágrafos” inserindo o símbolo “§” que será de grande valia para as análises futuras ;
2. Para haver o cruzamento entre os critérios de partição, é aconselhável inserir um delimitador a mais fora da balizagem, utilizando o símbolo “\$”, por exemplo:

```
<mois=01> $M01 <quinzaine=11> $Q11 <semaine=111>
<Sda=1793><numero=260><edito=0><Epg=1>
<Sat=0>
§ la grande douleur du *père *duchesne
au sujet de la mort de *marat assassiné à
coups de couteau par une garce du *calvados ,
dont l ' évêque *fauchet était le directeur . ses
bons avis aux braves *sans - culottes pour
qu ' ils se tiennent sur leurs gardes ,
attendu qu ' il y a dans *paris plusieurs milliers
de tondus de la *vendée qui ont la patte
graissée pour égorger tous les bons citoyens
```

Barra de Ferramentas



Para iniciar o sistema, clique sobre o ícone

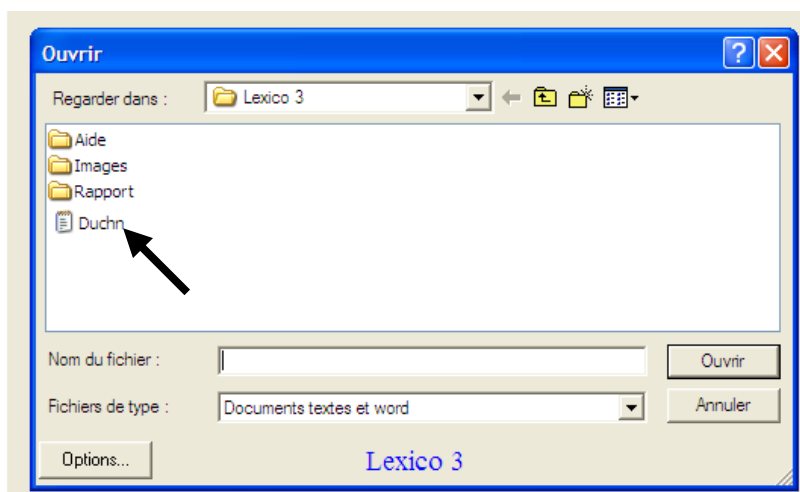


Passo nº 1 : Integrar um texto à base

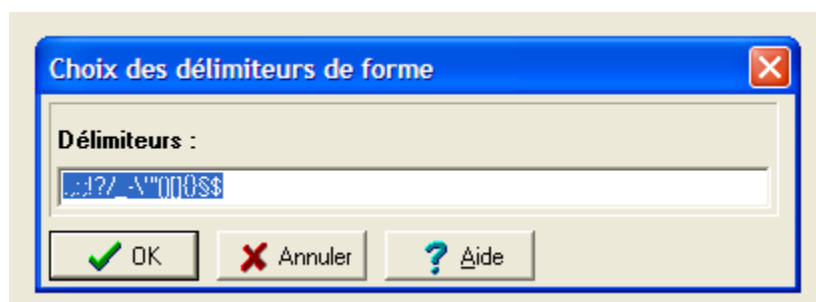
- 1.1 Para segmentar um arquivo-texto, clique sobre o ícone Nova base (Nouvelle base) e uma caixa de diálogo se abrirá, o que lhe permitirá selecionar um texto ;



1.2 Dê duplo clique sobre o arquivo-texto desejado (ou se você desejar, pode verificar o texto *Duch.txt* que é um exemplo de *corpus* em francês)



1.3 O **Lexico3** proporá uma lista de delimitadores por *default*:



1.4. Aceite a lista clicando sobre o botão **OK**

Obs.: Se você utilizou a "Dica 2" da "Parte II", você deverá apagar o sinal de \$ da lista de *default*.

Em seguida, o **Lexico3** segmentará o texto e apresentará uma lista de formas do *corpus* em uma janela à esquerda.

Lexico 3 - [TextPloreur]

Fichier Traitement Fenêtre

Navigation Rapport Dictionnaire

Sélectionnez une couleur: []

Recherche: [diab]

Formes (ordre lexicométrique)	Fréquence
de	6130
les	4748
la	4275
et	3773
le	2764
à	2632
que	2496
qui	2079
des	1944
il	1825
l	1770
pour	1665
en	1309
qu	1298
d	1274
nous	1270
un	1267
est	1232
tous	1195
ils	1187
...	1186

<mois=01><quinzaine=11><semaine=111>
 <Sda=1793><numero=260><edito=0><Epg=1>
 <Sat=0>
 § la grande douleur du *père *duchesne
 au sujet de la mort de *marat assassiné à
 coups de couteau par une garce du *calvados ,
 dont l' évêque *fauchet était le directeur . ses
 bons avis aux braves *sans - culottes pour
 qu' ils se tiennent sur leurs gardes ,
 attendu qu' il y a dans *paris plusieurs milliers
 de tonsus de la *vendée qui ont la patte
 graissée pour égorger tous les bons citoyens .
 <edito=1> § *marat n' est plus , foutre . peuple , gémis ,
 <Epg=2> pleure ton meilleur ami ; il meurt martyr de
 la liberté . c' est le *calvados qui a vomie le
 monstre sous les coups duquel il vient
 de périr . une jeune fille , ou plutôt une

Passo n° 2: Realizar uma concordância



2.1 Clique no ícone « concordância » (*concordance*)

2.2 Selecione a forma (palavra) do dicionário ou segmento repetido que você deseja estudar os contextos, em seguida, arraste o item selecionado para a janela à direita apertando o botão esquerdo do mouse, e solte-o.

Lexico 3 - [Concordances]

Fichier Traitement Fenêtre

Navigation Rapport Dictionnaire

Sélectionnez une couleur: []

Recherche: [diab]

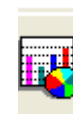
Formes (ordre lexicométrique)	Fréquence
rendre	52
traité	52
tyrans	52
diab	51
eh	51
esclaves	51
fin	51
reste	51
surtout	51
villes	51
aller	50
êtes	50
femmes	50
royalistes	50
ca	49

tous ces chenapans ; mieux vaut tuer le diable que le diable nous tue , comme je vous l' apans ; mieux vaut tuer le diable que le diable nous tue , comme je vous l' ai dit cent lards du *palais va se débattre comme un diable dans un bénitier pour prouver que ce qui lles commères qui ne vous parlent que du diable , avoir sans cesse à son chevet un bougre oisonner de vos liqueurs , gargotiers du diable , vous voulez encore empoisonner l' esprit ' ai souvent dit , il vaut mieux tuer le diable que le diable nous tue . § pourquoi tous it , il vaut mieux tuer le diable que le diable nous tue . § pourquoi tous les jean - foutres seaux cette maudite engeance , et que le diable l' emporte au milieu des rochers ou au roche à vous embêter avec ces journaux du diable , et les feuilles empoisonnées que *pitt livrons en la terre . mieux vaut tuer le diable que le diable nous tue , comme je l' ai grande colère contre les deux avocats du diable , qui ont osé plaider la cause de cette s . n' ai je pas vu ces deux avocats du diable non seulement se démener comme des diables is , roi de *france . depuis ce pacte du diable , les *sans - culottes ont été traités comme de bénédictions ; mais , foutre , si le diable voulait te tenter , songe à *la - *fayette

Discas:

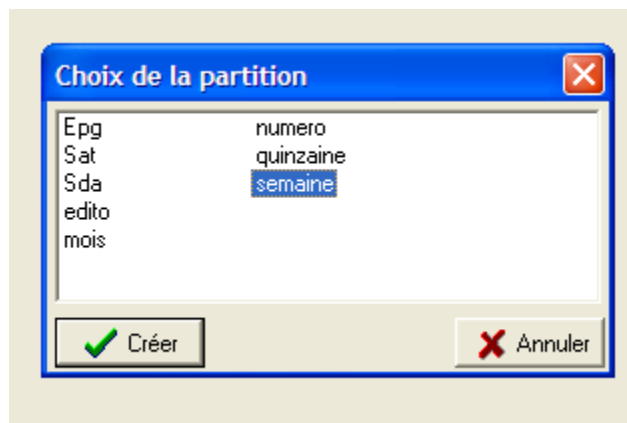
1. É possível selecionar um grupo de formas no dicionário ou de segmentos repetidos; para isso, basta clicar sobre as expressões apertando a tecla “Control” (Ctrl) e arrastar o que foi selecionado para a janela da direita;
2. A barra superior sobre a janela de concordância possui outras ferramentas muito úteis, mas que serão melhor exploradas no manual completo.

Passo n° 3: Criar uma partição de corpus

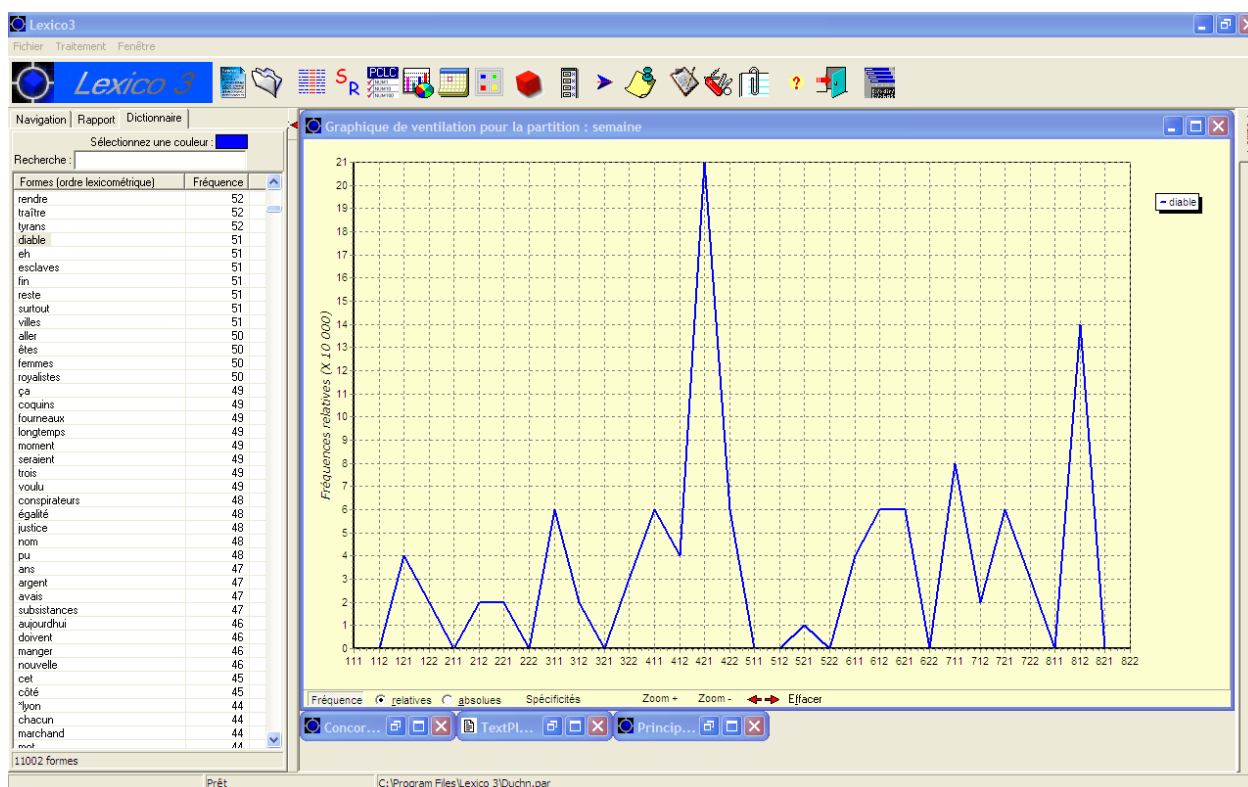


3.1 Clique sobre o ícone “estatísticas por partes” (statistiques pour parties)

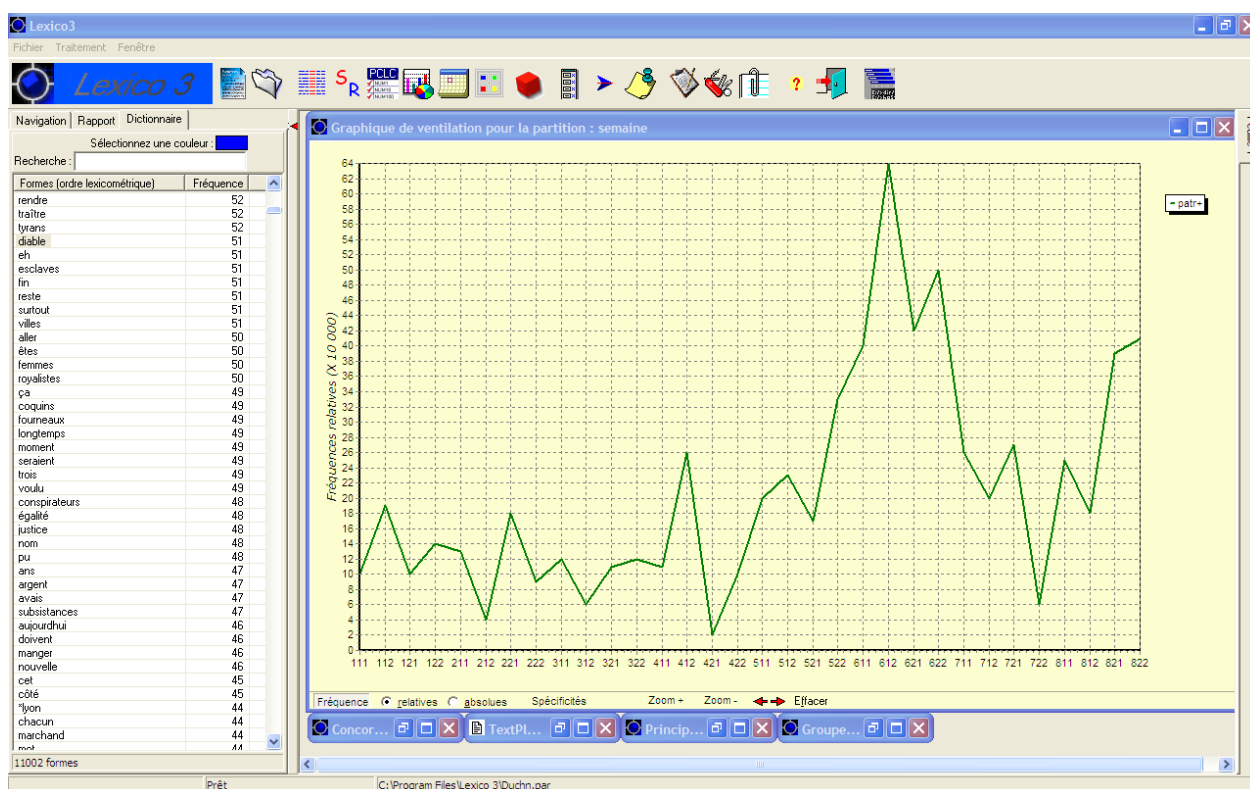
3.2 Uma caixa de dialogo aparecerá e lhe permitirá selecionar uma chave que servirá de critério de divisão do *corpus*. Selecione uma das partes (por exemplo, *semaine*), e depois clique sobre o botão “criar” (Créer)



3.3 Seleccione no dicionário a "forma" (palavra) da qual você deseja estudar os contextos, e arraste-a para a janela direita, depois solte-a sobre o “gráfico de estimativa” (graphique de ventilation) :



Você poderá perceber que na semana 421, houve um uso relativamente maior do termo “diable” do que em outras semanas. É a partir de indícios como esses com os quais o pesquisador pode se deparar, e fazer deles um objeto de reflexão.



Passo nº 5 : Criar um mapa de seções

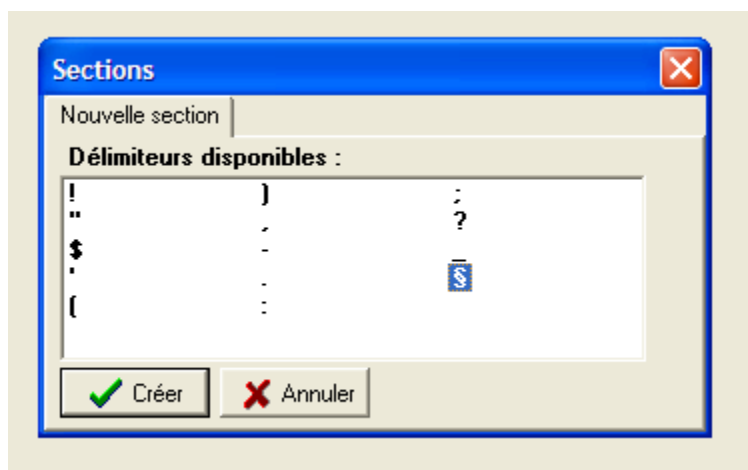


5.1 Clique sobre o ícone “mapa de seções” (*carte des sections*) ; uma caixa de diálogo se abrirá e permitira a seleção dos delimitadores : parágrafo, ponto etc. que dividirão o texto em seções.

5.2 Selecione, por exemplo, o delimitador “parágrafo”.

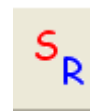
5.3 Clique em seguida sobre o botão **Créer**. Uma janela aparecerá à direita, sobre a qual cada quadrado representa uma seção do texto (você poderá visualizar a divisão em partes selecionando a partição *semaine* na caixa de partições).

5.4 Selecione (no dicionário) a forma que você deseja estudar, arraste-a para a janela direita e solte-a.



Os quadrados coloridos em “azul” representam os parágrafos em que a palavra “diable” apareceu. Clicando sobre cada um dos pontos demarcados, é possível visualizar o parágrafo todo em que ocorreu.

Passo nº 6 : Créer un inventário de segmentos repetidos



6.1 Clique sobre o ícone « Segments repetidos » (*Segments repetes*) e uma caixa de diálogo se abrirá e lhe permitirá selecionar os delimitadores de segmentos.

6.2 **Lexico3** lhe proporá alguns limitadores de seleção e delimitadores por *default*.

6.3 Aceite os limitadores clicando sobre o botão **OK**

6.4 **Lexico3** exibirá uma lista de segmentos repetidos no *corpus* em uma janela à esquerda.

Paramétrage du calcul des segments répétés

Délimiteurs de séquence :

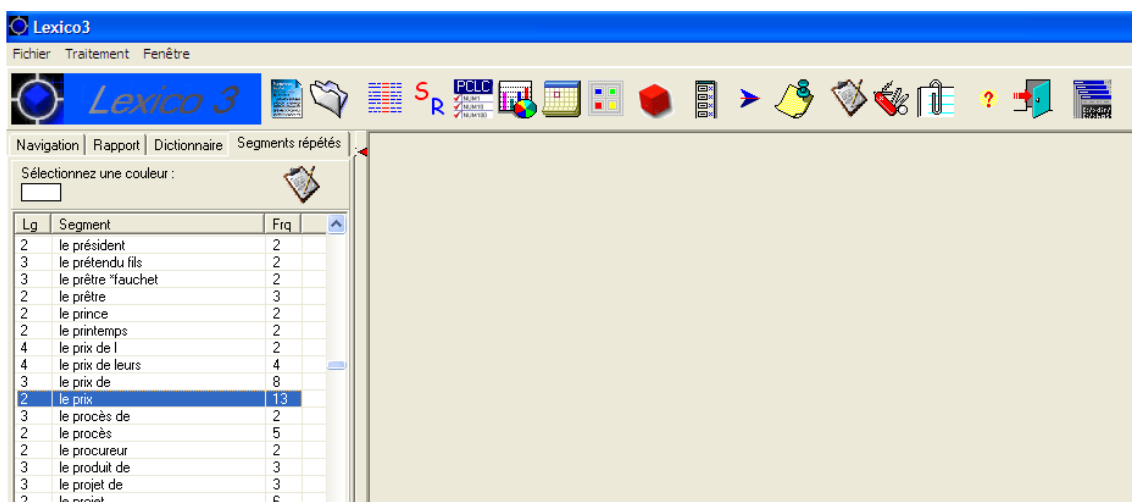
<input checked="" type="checkbox"/> !	<input checked="" type="checkbox"/>)	<input checked="" type="checkbox"/> :
<input checked="" type="checkbox"/> "	<input checked="" type="checkbox"/> ,	<input checked="" type="checkbox"/> ?
<input checked="" type="checkbox"/> \$	<input checked="" type="checkbox"/> -	<input checked="" type="checkbox"/> _
<input checked="" type="checkbox"/> '.	<input checked="" type="checkbox"/> .	<input checked="" type="checkbox"/> \$
<input checked="" type="checkbox"/> (<input checked="" type="checkbox"/> :	

<input checked="" type="checkbox"/> Epg	<input checked="" type="checkbox"/> mois
<input checked="" type="checkbox"/> Sat	<input checked="" type="checkbox"/> numero
<input checked="" type="checkbox"/> Sda	<input checked="" type="checkbox"/> quinzaine
<input checked="" type="checkbox"/> edito	<input checked="" type="checkbox"/> semaine

Seuil de sélection des segments : 10

Dica :

Você poderá escolher a quantidade mínima de segmentos repetidos a partir de 2 (*Seuil de sélection des segments*) ou seja, se o segmento “le prix” aparecer pelo menos 2 vezes no corpus, ele será exibido.



Passo nº 7 : Calculer as especificidades de uma parte



7.1 Relembrando: uma partição em « semaine » foi criada no passo 3

7.2 Clique sobre o ícone « principais características lexicométricas » *PCLC clique*

Sobre o item “semaine” e a janela das « principais características lexicométricas da partição » se abrirá.

7.3 Selecione uma parte do corpus na janela

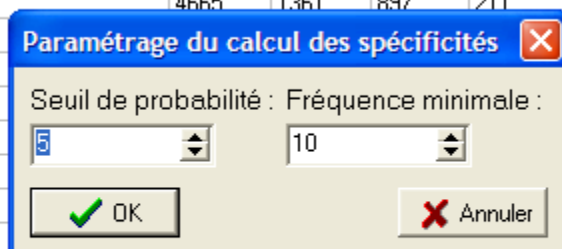
7.4 Clique sobre o botão *Spécifs*

Num	Partie	Occurenc	Formes	Hapax	Finis	Forme
1	111	3968	1255	840	166	de
2	112	5601	1532	1010	268	de
3	121	4909	1415	938	220	de
4	122	4778	1387	929	232	de
5	211	4314	1320	894	177	de
6	212	4857	1473	993	188	de
7	221	4756	1431	946	216	de
8	222	6415	1745	1140	294	de
9	311	4665	1361	897	211	de
10	312	4475	1395	979	214	de
11	321	4313	1329	911	189	de
12	322	5824	1635	1097	239	de
13	411	4352	1413	1003	192	de
14	412	4109	1231	816	169	de
15	421	4656	1432	970	188	de
16	422	4887	1523	1060	204	de
17	511	4333	1287	868	193	de
18	512	2934	966	660	117	de
19	521	5730	1644	1108	212	de
20	522	4817	1404	934	224	de

e uma caixa de diálogo se abrirá que lhe permitirá fixar a quantidade de termos e a sua frequência mínima.

7.5 Aceite a entrada e a frequência propostas clicando sobre o botão **OK**

6415	1745	1140	294	de
4665	1361	897	211	de
4333	1287	868	193	de
2934	966	660	117	de
5730	1644	1108	212	de



7.6 o *Lexico3* exibirá em uma coluna à esquerda as especificidades da forma selecionada por semana

Terme	Frq Tot.	Frq P...	Spécif
*marat	65	16	12
elle	368	34	10
*paris	171	22	9
*calvados	20	6	6
aurait	84	11	6
*montagne	14	5	6
*fauchet	17	5	5
puisque	17	5	5
utile	11	3	4
pouvez	10	3	4
moi	144	11	4
citoyen	27	5	4
cependant	56	7	4
cu	13	4	4
m	206	14	4
était	194	14	4
corps	17	3	3
majesté	17	3	3
un	1267	49	3
confiance	17	3	3
perdu	23	3	3
gardes	20	3	3
plusieurs	20	3	3
*bordeaux	16	3	3

Num	Partie	Occurenc	Formes	Hapax	Fmax	Forme
1	111	3968	1255	840	166	de
2	112	5601	1532	1010	268	de
3	121	4909	1415	938	220	de
4	122	4778	1387	929	232	de
5	211	4314	1320	894	177	de
6	212	4857	1473	993	188	de
7	221	4756	1431	946	216	de
8	222	6415	1745	1140	294	de
9	311	4665	1361	897	211	de
10	312	4475	1395	979	214	de
11	321	4313	1329	911	189	de
12	322	5824	1635	1097	239	de
13	411	4352	1413	1003	192	de
14	412	4109	1231	816	169	de
15	421	4656	1432	970	188	de
16	422	4887	1523	1060	204	de
17	511	4333	1287	868	193	de
18	512	2934	966	660	117	de
19	521	5730	1644	1108	212	de
20	522	4817	1404	934	224	de
21	611	4196	1286	885	174	de

Passo nº 8 :Realizar uma Análise Fatorial de Correspondências

8.1 Relembrando: uma partição em « semaine » foi criada no passo 3

8.2 Clique sobre o ícone « principais características lexicométricas » (*PCLC*)” e a janela das « principais características lexicométricas da partição » se abrirá.

8.3 Clique sobre o ícone *AFC* (à direita da tela)

The screenshot shows the Lexico3 software interface. The main window displays a table titled 'Principales caractéristiques de la partition : semaine'. The table has columns for 'Num', 'Partie', 'Occurenc', 'Formes', 'Hapax', 'Fmax', and 'Forme'. The 'AFC' button is highlighted in the right-hand toolbar.

Num	Partie	Occurenc	Formes	Hapax	Fmax	Forme
1	111	3968	1255	840	166	de
2	112	5601	1532	1010	268	de
3	121	4909	1415	938	220	de
4	122	4778	1387	929	232	de
5	211	4314	1320	894	177	de
6	212	4857	1473	993	188	de
7	221	4756	1431	946	216	de
8	222	6415	1745	1140	294	de
9	311	4665	1361	897	211	de
10	312	4475	1395	979	214	de
11	321	4313	1329	911	189	de
12	322	5824	1635	1097	239	de
13	411	4352	1413	1003	192	de
14	412	4109	1231	816	169	de
15	421	4656	1432	970	188	de
16	422	4897	1523	1060	204	de
17	511	4333	1287	868	193	de
18	512	2934	966	660	117	de
19	521	5730	1644	1108	212	de
20	522	4817	1404	934	224	de

e uma caixa de diálogo se abrirá que lhe permitirá fixar a quantidade de termos e a sua frequência mínima.

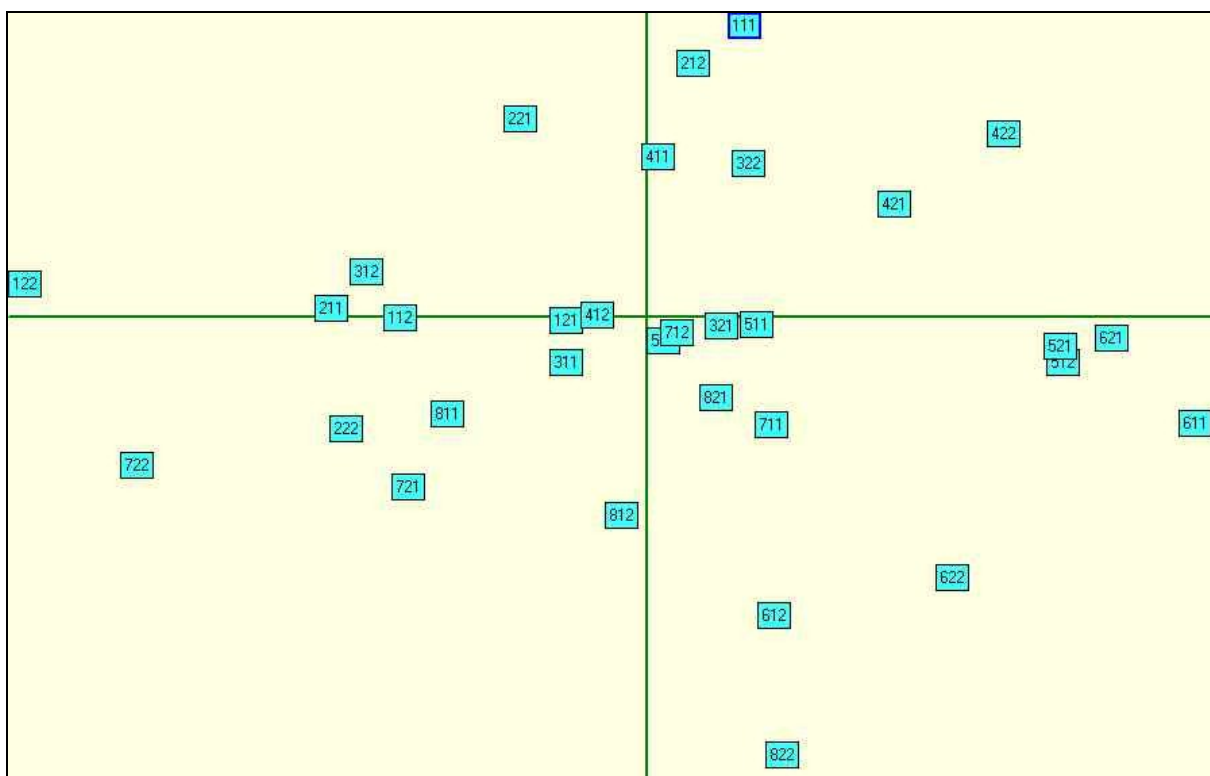
8.4 Aceite a entrada e a frequência propostas clicando sobre o botão *OK*

The screenshot shows the 'Paramétrage de l'AFC' dialog box. The dialog box contains the following fields and values:

- Titre de l'analyse: Duchn : semaine
- NI: 1420
- NJ: 32
- NF: 31
- NI2: 0
- NJ2: 0
- FreqMin: 10

The 'OK' button is highlighted with a green checkmark.

8.5 *Lexico3* exhibirá a tipologia obtida a partir da Análise fatorial de correspondências sobre o plano dos dois primeiros fatores



Passo nº 9 : Navegar entre os resultados

A opção *Navigation* (canto esquerdo ao lado do dictionnaire) lhe permite navegar entre os diferentes resultados produzidos pelo *Lexico3*. Um clique sobre uma das unidades das seções exibe o gráfico correspondente na janela à direita.

Principales caractéristiques de la partition : semaine

Num	Partie	Occurenc	Formes	Hapax	Fmax	Forme
1	111	3988	1255	840	166	de
2	112	5601	1532	1010	268	de
3	121	4909	1415	938	220	de
4	122	4778	1387	929	232	de
5	211	4314	1320	894	177	de

Passo n° 10 : Gravar os resultados em um relatório

- 10.1 Clique sobre a opção *Rapport* na janela da esquerda
- 10.2 Selecione e maximize a janela que você deseja integrar ao relatório
- 10.3 Clique sobre o ícone « Adicionar ao relatório » (*Ajouter au rapport*)



- 10.4 Na opção *Rapport* clique sobre o botão “Gravar” *Enregistrer*.
- 10.5 Após ter gravado o relatório, você poderá consultá-lo clicando sobre o ícone *Editeur*
- 10.6 Um navegador será iniciado (Netscape ou Internet Explorer) e você poderá imprimir ou inserir os gráficos ou quaisquer outros elementos registrados no relatório.

The screenshot shows a web browser window titled 'Rapport - Windows Internet Explorer'. The address bar shows the URL: 'C:\Documents and Settings\CONDE DIRCEU CLEBER\Mes documents\DOUTORADO\psanduiche\Lexicometrie\Duchn_20070212-115754\index.htm'. The browser interface includes a menu bar (Fichier, Edition, Affichage, Favoris, Outils), a search bar (Google), and a toolbar with various icons. The main content area displays a report titled 'Rapport d'analyse' with a table of lexicometric characteristics.

Principales caractéristiques lexicométriques	
Nombre des occurrences	142177
Nombre des formes	10988
Fréquence maximale	6130
Nombre des hapax	5056

Below the table, there is a link: 1. [Analyse factorielle des correspondances - semaine\(1\)](#)

Ao fim dos *dez primeiros passos com Lexico3*, você poderá realizar explorações mais longas. “Bonne chance!”