

Tutoriel n°2 : Séries textuelles chronologiques

[STC]

André Salem

salem@msh-paris.fr

Apprendre à :

1. Constituer une série textuelle chronologique
2. Repérer la structure chronologique
3. Repérer les formes prototypiques de l'évolution
4. Utiliser les outils d'analyse chronologique
5. Organiser le retour au texte

Les *séries textuelles chronologiques* sont des corpus constitués par la réunion de textes similaires produits par une même source textuelle au cours d'une période de temps. Comme nous allons le voir, la prise en compte de la dimension chronologique de tels corpus permet de mettre en évidence des variations qui surviennent au cours du temps dans l'emploi du vocabulaire, de mettre en évidence des moments importants dans l'évolution de celui-ci¹. L'analyse des séries textuelles chronologiques s'appuie sur une série de méthodes tout particulièrement adaptées à ces objectifs.

Les 96 livraisons qui constituent le corpus *Duchn* constituent un découpage en unités trop ténues pour permettre aux analyses quantitatives de mettre en évidence des variations chronologiques. Comme nous l'avons vu plus haut, les analyses opérées à partir de ce type de découpage font ressortir des oppositions qui concernent avant tout des différences stylistiques dans la confection des différents numéros. Nous allons voir dans ce qui suit que dès que l'on regroupe les livraisons publiées au cours de la même semaine, de la même quinzaine ou du même mois les analyses textométriques mettent en évidence des évolutions chronologiques liées à l'évolution du vocabulaire au cours du temps.

1 L'analyse des séries textuelles chronologiques : analyses de la série textuelle Duch8M

La partition du corpus *Duchn* en huit parties regroupant chacune les livraisons parues au cours d'un même mois permet de centrer plus précisément nos observations sur l'évolution du vocabulaire au cours de la période couverte par le corpus.

¹ Sur l'analyse des séries textuelles chronologiques on consultera, par exemple [Salem 1993].

1.1 Typologie sur les 8 périodes

L'AFC réalisée à partir de ce regroupement en 8 parties met en évidence la proximité relative des périodes consécutives et leur échelonnement dans le temps.²

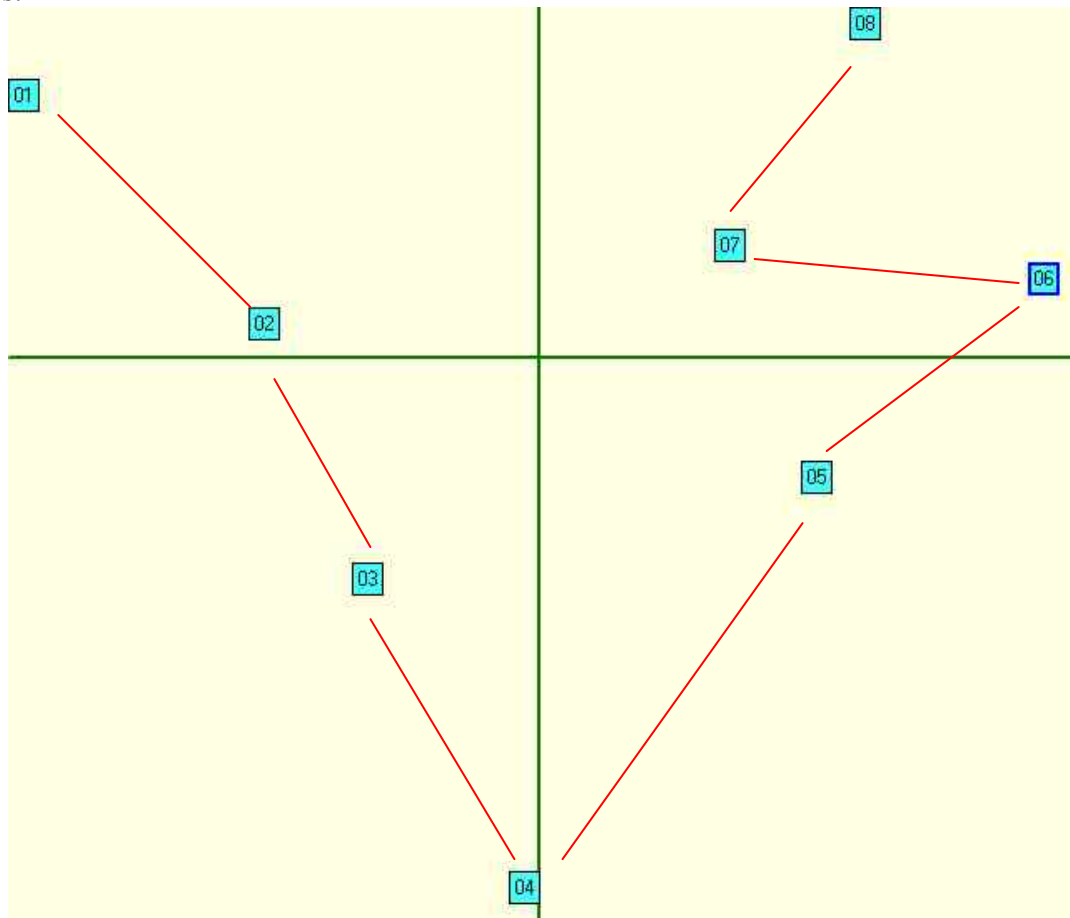


Figure 1

AFC sur le corpus *Duchn*

8 périodes d'un mois x 1420 formes de fréquence ≥ 10

1.2 Interprétation des analyses chronologiques

==== Repère méthodologique ==== *l'effet Guttman*

L'AFC réalisée à partir des séries textuelles chronologiques produit souvent des figures particulières, dont l'interprétation doit être faite selon des règles spécifiques.

L'existence d'une évolution progressive du vocabulaire au fil des parties d'un corpus chronologique induit un système de distances particulier sur l'ensemble des parties soumises à l'analyse :

- les parties consécutives dans les temps se révèlent plus proches entre elles que les parties séparées par une période de temps plus important.

² Les traits rouges qui relient les parties consécutives dans le temps ont été rajoutés par nous sur le résultat de fournis par l'analyse. Signalons que l'on rencontre très régulièrement ce genre de résultats lorsqu'on analyse des corpus chronologiques (cf. [Salem 19xx]).

La décomposition de ces distances selon les différents axes de la représentation factorielle constitue alors une suite d'approximations progressive plutôt plus complexes que le phénomène relativement simple dont elles à rendent compte.

Le schéma théorique correspondant à une telle évolution est le suivant :

- le premier facteur rend compte d'une évolution linéaire.
- les facteurs suivants sont des fonctions de degré plus élevé (2, 3, 4, etc.) de ce premier facteur.

L'analyse des corpus chronologiques réels relève pour partie de ce schéma même si d'autres oppositions peuvent venir perturber le schéma théorique décrit plus haut. :

Pour en savoir plus :

P on consultera par exemple :

Salem, A. : « Approches du temps lexical », in *Histoire et mesures*, **XXXXX**, Klincksieck, 19xx

Guttman 1941]. On trouvera un exposé plus accessible dans [Benzécri 1973] p 192-196, et des applications dans [Van Rijckevorsel 1987]..

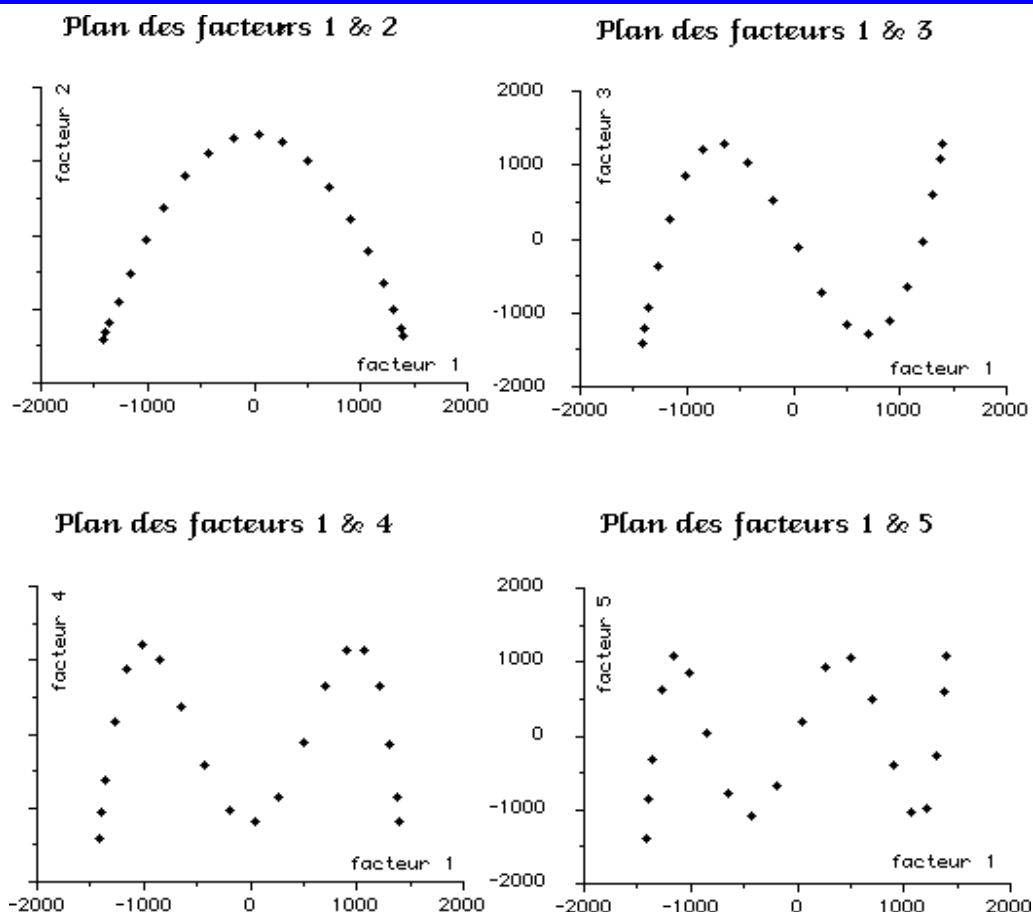


Figure 2

Les facteurs issus de l'analyse du scalogramme 20 x20

Guide de lecture pour la figure 2

On voit sur cette figure les valeurs prises par les cinq premiers facteurs issus de l'analyse du scalogramme 20x20.

Les points dessinent une parabole sur le plan des deux premiers facteurs si on les relie dans cet ordre. L'étude des facteurs suivants montre que ces derniers sont également des fonctions, de degré supérieur, du premier facteur. La prise en compte de facteurs supplémentaires apporte, dans ce cas, une information d'ordre systématique qui est peu intéressante.

Tableau 1

Scalogramme de dimension 20 x20

	C 1	C 2	C 3	C 4	- -	C 18	C 19	C 20
L1	1	1	0	0	-	0	0	0
L2	1	1	1	0	-	0	0	0
L3	0	1	1	1	-	0	0	0
L4	0	0	1	1	-	0	0	0
.....								
L18	0	0	0	0	-	1	1	0
L19	0	0	0	0	-	1	1	1
L20	0	0	0	0	-	0	1	1

Le modèle de l'effet Guttman nous fournit un éclairage important pour comprendre la disposition des parties sur la courbe présentée à la figure 1. Devant des résultats de ce type on évitera donc de commenter séparément les oppositions constatées sur chacun des axes factoriels pris isolément. Le schéma classique de l'interprétation d'une typologie que l'on affine au fur et à mesure par la prise en compte de nouveaux axes factoriels doit faire place à la reconnaissance d'une situation caractéristique globale liée à l'existence et à la dominance d'une gradation progressive.

Au delà de cette mise en évidence, le modèle d'analyse attire notre attention sur les moments particuliers de l'évolution lexicale. La particularité de l'évolution constatée ici à propos de l'évolution de la série **Duch8M** tient au fait que la période M7 ne s'inscrit pas totalement dans l'évolution d'ensemble.

1.3 Retour à l'analyse des tableaux lexicaux

A la différence du scalogramme soumis à l'AFC dans la section qui précède, les tableaux lexicaux tels que celui que nous avons soumis à l'AFC à la section §1 comptent en général plusieurs dizaines de milliers de cellules. De plus, ces tableaux résultent de décomptes réels effectués sur des textes et non d'une construction formelle comme c'était le cas pour le scalogramme. Si les résultats issus de l'analyse factorielle appliquée à ces deux tableaux présentent des similitudes, ils présentent aussi des différences profondes. Dans les tableaux lexicaux, la propriété d'autocorrélation des facteurs ne concerne pas tous les facteurs issus de l'analyse bien que les premiers facteurs possèdent en général cette propriété. De plus, comme on a pu le vérifier sur la figure 1, le schéma des distances que l'on construit sur la base d'un tableau lexical ne présente pas la régularité de celui que l'on obtient à partir d'un scalogramme.

Comme dans le cas du scalogramme, les premiers facteurs issus de l'AFC du corpus **Duch8M** possèdent la propriété de rapprocher les périodes consécutives dans le temps. Cette propriété d'une variable s'appelle l'**autocorrélation**. On vérifie sur la figure XX que tel est bien le cas pour le premier facteur issu de l'analyse du tableau (1420 formes x 8 périodes).

L'expérience prouve que dans le cas de l'analyse des tableaux réalisés à partir de séries chronologiques

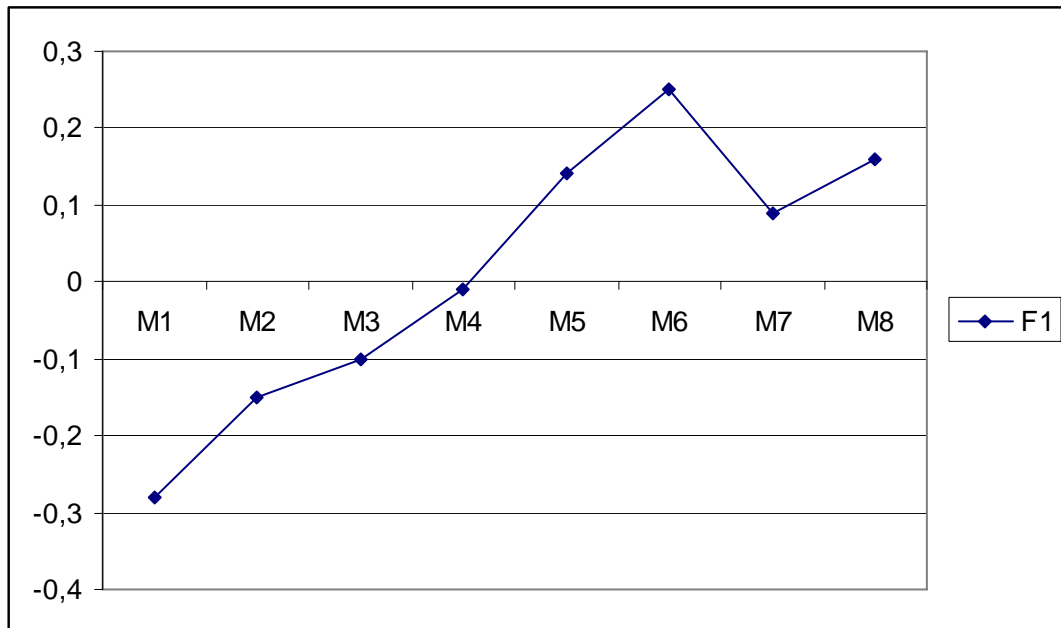


Figure 3

Le premier facteur issu de l'analyse du tableau
8 périodes d'un mois x 1420 formes de fréquence ≥ 10

1.4 Mesures de contiguïté

==== Repère méthodologique ==== le coefficient de Geary

Parmi d'autres coefficients du même type, le *coefficient de Geary* permet de mesurer l'autocorrélation d'une série de nombres : x_1, x_2, \dots, x_n .

Une valeur faible du coefficient indique que les valeurs successives sont « en moyenne » plus proches

Pour en savoir plus :

Lebart, L., Salem, A. : *Statistiques textuelles*, Paris, Dunod, 1994.

2 Rechercher les formes prototypiques de l'évolution

La disposition des 8 parties que l'on voit sur la figure 1, nous laisse penser que l'évolution dans le temps de la série Duch8M n'est pas complètement régulière. Des formes voient leur importance croître au cours du temps mais on peut penser que cette croissance connaît un ralentissement en M7 pour reprendre dans la huitième période.

Une des premières questions qui se posent est celle de mettre en évidence tout ou partie de l'ensemble des formes lexicales dont le profil ressemble à cette évolution. Plusieurs méthodes statistiques peuvent nous aider à mettre ces formes en évidence³. Dans ce qui suit nous avons

³ A partir des résultats de l'AFC, les formes possédant une coordonnée factorielle positive et également pourvues d'une forte contribution relative au facteur constituent de bons candidats pour ce type de sélection.

sélectionné une classe de formes construite par une méthode de classification hiérarchique à partir des profils des formes.

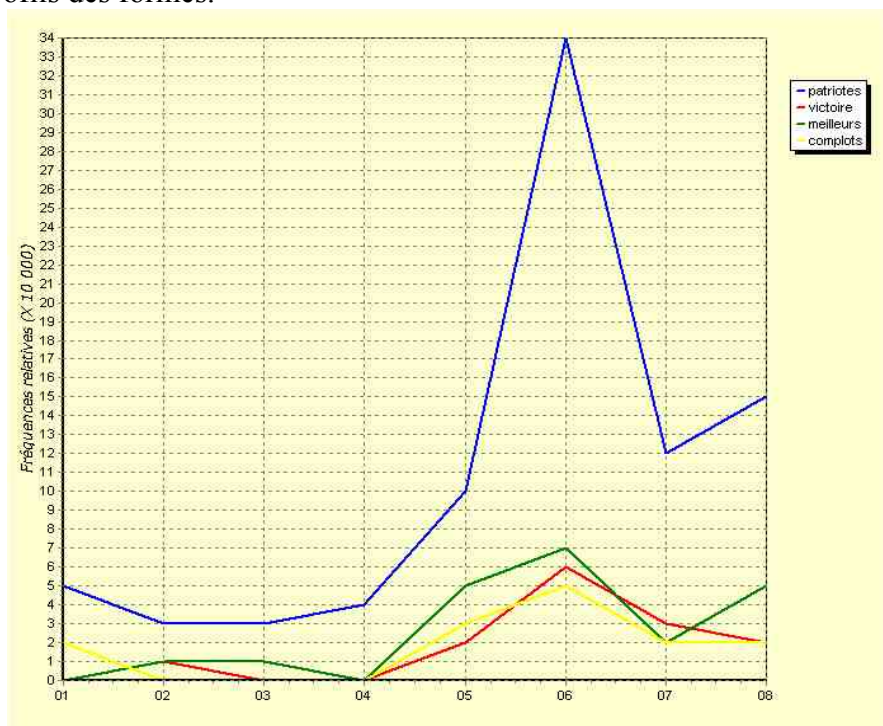


Figure 4

Quelques formes prototypiques de l'évolution de la série Duchn8M

Dans une classe plus large qui contient : *patriotes*, *meilleurs*, *complots*, *surplus*, *accusent*, *victoire*, *réjouissance*, *modérés*, *revient*, *tricherie*.

2.1 Calcul des spécificités chronologiques

Le calcul des *spécificités chronologiques* permet de mettre en évidence les formes qui ont une ventilation particulièrement centrée sur une période ou un groupe de période consécutives du corpus.⁴ Ce calcul signale comme le fait le plus saillant, sur l'ensemble du corpus, la répartition de la forme *patriotes* centrée sur la période M6.

Tableau 2 :

Spécificités chronologiques majeures du corpus *Duchn*

⁴ Sur le calcul des spécificités chronologiques, on consultera [Salem 1993] et [Lebart & Salem 1994]

Formes/SR	F	f	coeff	Partie(s)
patriotes	152	60	18	06
constitution	72	53	16	01 - 02
*paris	171	78	14	01 - 02
vous	1097	705	13	01 - 04
*phélipotin	13	13	13	06
avez	171	55	12	04
*custine	70	23	12	01 - 02
foutre	1100	784	10	07 - 08
phélippotins	14	14	10	07 - 08
nobles	42	21	9	01
elle	368	90	9	01
*courtille	13	3	9	08
*vincent	15	14	9	06 - 07
républicains	129	91	9	05 - 08
tu	296	200	9	03 - 06
on	856	288	9	05 - 06
*marat	65	27	9	01
nouvelle	46	21	9	06
çà	8	8	8	02
*ami	8	8	8	01
je	979	561	8	04 - 07
*cordeliers	17	6	8	08
sociétés	15	10	8	08
ai	202	114	8	04 - 06

Guide de lecture pour le tableau 2

Le tableau fournit les spécificités chronologiques majeures calculées pour le corpus *Duchn* muni de la partition en 8 périodes d'un mois.

- La première colonne du tableau indique le terme pour lequel le diagnostic de spécificité chronologique a été calculé ;
- la seconde *F*. donne la fréquence du terme dans l'ensemble du corpus ;
- la troisième *f* la fréquence de ce même terme dans la période considérée ;
- la quatrième *Coeff.* fournit le coefficient de spécificité chronologique calculé pour le terme et la période ;
- on trouve dans la dernière colonne la période pour laquelle le diagnostic a été calculé.

==== Repères méthodologiques ====

Spécificités chronologiques - Accroissements spécifiques

Pour chaque unité textuelle (forme, segment répété, etc.) d'un corpus divisé en *périodes*, le calcul des *spécificités chronologiques* permet de déterminer l'intervalle temporel (période ou groupe de périodes consécutives) au cours duquel cette unité connaît un emploi spécifique maximal (positif ou négatif). Les spécificités ainsi calculées sont ensuite triées par coefficient de spécificité décroissant (i.e. les plus remarquables en tête) afin de mettre en évidence les distorsions les plus remarquables dans l'emploi des termes par les différentes périodes (ou groupes de périodes consécutives).

Pour chaque période du corpus (à partir de la seconde), le calcul des *accroissements spécifiques* permet de mettre en évidence le vocabulaire et les segments dont l'utilisation connaissent un accroissement particulièrement

important au cours de la période (*accroissements spécifiques positifs*) ou au contraire ceux dont l'utilisation se raréfie brutalement après une utilisation plus intensive (*accroissements spécifiques négatifs*).

Pour en savoir plus :

Pour un exposé et des exemples d'application de l'analyse des spécificités à l'étude des corpus de textes, on consultera par exemple :

Salem, A. : « Approches du temps lexical », in *Histoire et mesures*, **XXXXXX**, Klincksieck, 19xx

Lebart, L., Salem, A. : *Statistiques textuelles*, Paris, Dunod, 1994.

Les diagnostics suivants concernent, dans l'ordre décroissant d'importance : l'abondance des formes *constitution* et *paris* dans les deux premières parties du corpus puis, plus loin dans la liste, d'autres diagnostics dont plusieurs concernent à nouveau la période M6 (phélipotin, *nouvelle*, etc.)

==== **Lexico3** ==== **Liste des spécificités chronologiques**

- ✓ Sélectionner l'icône **PCLC** (5^{ème} icône à partir de la gauche)
- ✓ Sélectionner une partie ou un groupe de parties
- ✓ Appuyer sur le bouton **SpecDia** (à droite de la fenêtre)
- ✓ Les résultats apparaissent dans un onglet spécifique dans la fenêtre centrale. On peut trier ces diagnostics en cliquant sur le titre de chaque colonne.

2.2 Calcul des accroissements spécifiques

Tableau 3

Spécificités majeures pour les formes et les segments répétés dans la période M6 du corpus **Duch8M**

Terme	Frq. Tot.	Fréquence	Coeff.
<i>patriotes</i>	152	60	18
<i>*phélipotin</i>	13	13	13
<i>les patriotes</i>	91	36	11
<i>nouvelle</i>	46	21	9
<i>c est là que</i>	10	8	7
<i>*vincent</i>	15	10	7
<i>est là</i>	36	15	6
<i>là que</i>	12	8	6
<i>c est là</i>	28	13	6
<i>conspirateurs</i>	48	18	6
<i>ai</i>	202	48	6
<i>les</i>	4748	687	6
<i>accuser</i>	10	7	5
<i>diviser</i>	16	9	5
<i>mes</i>	120	30	5
<i>contre</i>	298	60	5
<i>les meilleurs</i>	31	13	5
<i>généraux</i> <i>*sans</i>	10	7	5
<i>culottes</i>			
<i>aristocrates</i>	133	32	5

Forme	Frq. Tot.	Fréquence	Coeff.
patriotes	111	60	21
les patriotes	68	36	13
*phélipotin	13	13	11
*vincent	10	10	9
nouvelle	42	21	8
les	3667	687	8
conspirateurs	35	18	7
*camille	8	8	7
généraux *sans culottes	7	7	7
*bouchotte	9	8	7
c est là	24	13	6
diviser	13	9	6
aristocrates	94	32	6
*phélipotins	6	6	6
est là	30	15	6
c est là que	10	8	6
les meilleurs	22	13	6
ai	179	48	5
victoires	7	6	5
victoire	20	11	5
mes	101	30	5
contre	243	60	5
là que	12	8	5
accuser	9	7	5
en réjouissance de	8	6	5
meilleurs	29	13	5
réjouissance	8	6	5
les conspirateurs	19	10	5
en réjouissance	8	6	5
des patriotes	10	7	5

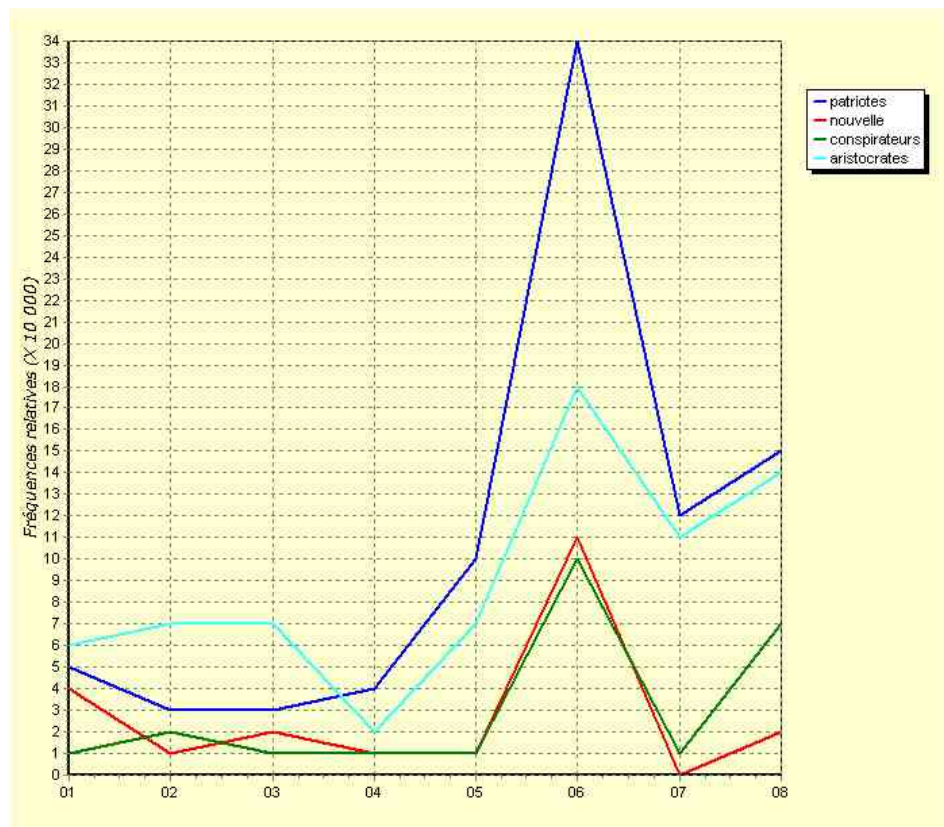


Figure 5

2.3 Un vocabulaire « offensif »

L'analyse des *spécificités positives* (unités textuelles particulièrement abondantes dans chacune des parties du corpus) montre que le caractère le plus marquant de cette évolution lexicale réside dans l'augmentation progressive, particulièrement remarquable dans la sixième partie M6, d'un vocabulaire dont on peut remarquer qu'il est lié au conflit politique et que nous appellerons, dans le cadre de cette étude, *vocabulaire offensif*. Ce vocabulaire, relativement sous-utilisé dans les premières parties du corpus (M1, M2, M3) connaît par la suite une utilisation croissante qui trouve un paroxysme en M6. Du point de vue de cette évolution d'ensemble, la période M7 marque un net recul dans l'utilisation de ce vocabulaire qui sera à nouveau utilisé en M8.

On trouve à la figure 6 la ventilation dans les 8 parties du corpus de quelques-unes des formes les plus caractéristiques de cet ensemble : les formes *patriotes*, *nouvelle*, *conspirateurs*⁵. Comme on le voit sur cette figure, le schéma de l'évolution fréquentielle de ces formes, exprimé à la fois en terme de fréquence relative (à gauche) et en termes de spécificité (à droite), correspond bien au type d'évolution chronologique que nous avons décrit plus haut.

⁵ J.C Deroubaix a proposé d'appeler ces formes, particulièrement représentatives d'un type de distribution remarquable des *parangons* de leurs classes respectives.

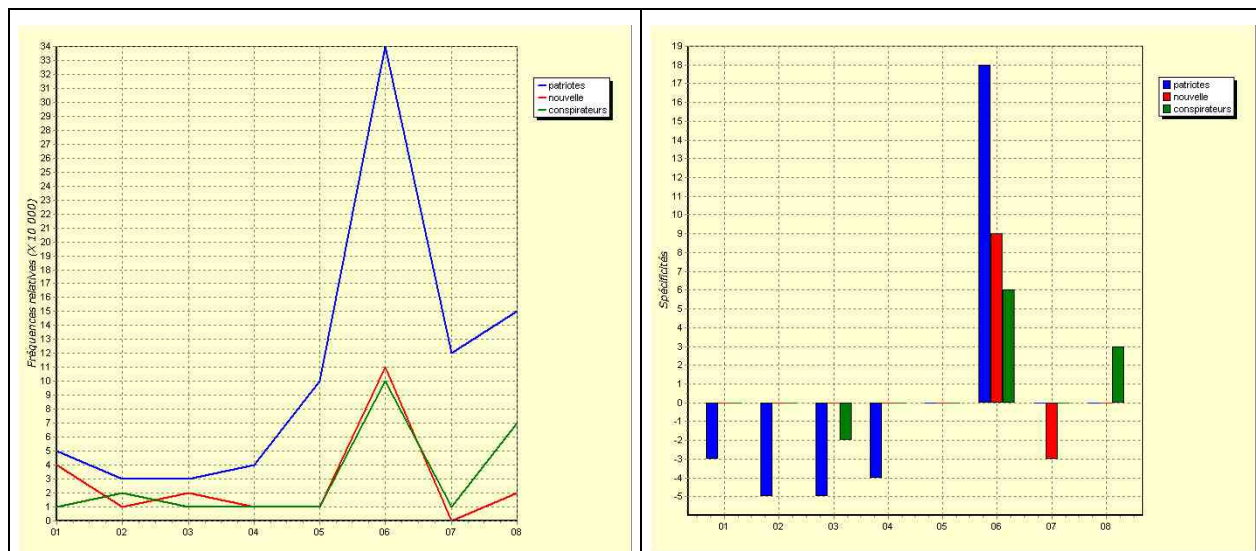


Figure 6

Ventilation des formes *patriotes*, *nouvelle*, *conspirateurs* dans les 8 périodes du corpus *Duch8M*

==== Lexico3 ==== Statistiques par parties

- ✓ Sélectionner l'icône **Statistique par parties** (6 ème icône à partir de la gauche)
- ✓ Sélectionner une clé pour créer une partition du corpus (ici : mois)
- ✓ Traîner sur la fenêtre ainsi créée une ou plusieurs forme ou un groupe de formes
- ✓ Choisir le mode de représentation (*fréquence relative*, *absolue* ou *spécificité*) en sélectionnant sous le graphique le bouton correspondant.

2.4 Vers une sémantique de corpus

Examinons maintenant la liste des formes les plus spécifiques de la partie M6 (tableau 3). Pour plusieurs de ces formes : *conspirateurs*, *accuser*, *diviser*, *fripions*, etc., l'appellation *vocabulaire offensif* paraît tout à fait légitime dans la mesure où ces termes véhiculent dans leur définition et dans tous leurs emplois discursifs une charge polémique évidente. Pour des termes comme *aristocrates* ou *phélipotins* (partisans d'un certain *Phélipot*), une connaissance, même superficielle, de la période historique permet de reconstituer la charge polémique que ces termes ont pu revêtir dans le contexte de la lutte révolutionnaire. La situation est plus délicate lorsqu'il s'agit d'expliquer l'appartenance à cet ensemble de formes comme *nouvelle*, *meilleurs*, etc. auxquelles on a pas coutume d'attacher de valeur polémique dans la langue courante. Pour comprendre la raison de la présence de ces formes dans des situations de conflit, il nous faut retourner à l'ensemble des contextes de leur utilisation.

Sur le tableau 4, les contextes de la forme *nouvelle* à l'intérieur de la partie M6 du corpus ont été triés en fonction de la forme qui suit le pôle. Un rapide examen de ces contextes permet de vérifier que, si l'on excepte les quelques occurrences de la forme qui renvoie à la forme substantivale *une (bonne/mauvaise) nouvelle*, la majorité des contextes renvoie à la forme adjectivale *une nouvelle X*. Dans un second temps, on peut vérifier que les contextes droits de cette forme adjectivale renvoient le plus souvent à des entités auxquelles le contexte

historique permet d'attribuer une valeur négative, ex : *une nouvelle *vendée*⁶, soit à des entités présentées comme négatives par l'emploi de formes explicitement dépréciatives (*une nouvelle clique d'aristocrates, une nouvelle clique de modérés*).⁷

Dans ces contextes, les entités auxquelles on fait allusion renvoient à des ennemis X dont on pensait s'être débarrassé dans un passé récent. L'adjectif *nouvelle X* introduit l'idée que la lutte n'est pas terminée et qu'il convient de la poursuivre.

Tableau 4
Concordances pour les formes *nouvelle* et *meilleurs*
dans la période M6 du corpus *Duch8M*

Partie : 06, Nombre de contextes : 21

lipoterie ; s ' il propose de former une	nouvelle	*vendée en mettant en liberté ses bons
recruter tous les brigands et former une	nouvelle	*vendée en proposant d ' ouvrir les pri
e est venu à son secours . " formons une	nouvelle	*vendée , lui a dit l ' animal cornu ,
andelle ; je savais qu ' il existait une	nouvelle	clique d ' aristocrates , de modérés ,
colère du *père *duchesne . de voir une	nouvelle	clique de modérés , de feuillants , d '
gtemps que je furetais pour découvrir la	nouvelle	conspiration qui est maintenant sur le
courriers extraordinaires ont apporté la	nouvelle	d ' une nouvelle victoire , on a annon
lus . la convention vient de recevoir la	nouvelle	de la destruction totale des brigands ;
bonnets rouges . ces *sans - culottes de	nouvelle	fabrique se sont répandus dans les sect
l débite gratis . au dénouement de cette	nouvelle	farce contre - révolutionnaire , nous d
ont unis , ne sera pas le jouet de cette	nouvelle	intrigue . après avoir sauvé la *france
' est pas assez , il faut savoir quelle	nouvelle	liste civile paye les rhapsodies qu ' i
a *vendée , ou plutôt pour en former une	nouvelle	, mais suivons pied à boule notre endor
rahisons , serait le connétable de cette	nouvelle	majesté égorgeante ; *brissot serait so
ers desseins . quand je découvre quelque	nouvelle	manigance contre les *sans - culottes ,
: victoire , vive la république . grande	nouvelle	, mes amis , leur ai je dit ; femmes ,
ommères , en réjouissance de cette bonne	nouvelle	, qui donne le coup de grâce aux aristo
parole . voilà , mon vieux , une grande	nouvelle	qui va faire gueuler , dans tous les co
uronnés , quand ils vont apprendre cette	nouvelle	. sa grande colère contre certains jean
dinaires ont apporté la nouvelle d ' une	nouvelle	victoire , on a annoncé que les lignes
détruire , lui donnent au contraire une	nouvelle	vie . ils tiennent les patriotes en hal

Partie : 06, Nombre de contextes : 13

vention même ; crie , gueule contre les	meilleurs	*montagnards , accuse le comité de sal
s manigances des *phélipotins contre les	meilleurs	citoyens , et surtout contre les génér
u plus d ' audace ? celui qui traite les	meilleurs	patriotes de bourreaux , d ' assassins
s osent lever la crête , et accusent les	meilleurs	patriotes de tous les coups de chien q
nt de sauver la république ; dénonce les	meilleurs	patriotes . " § le monstre *phélipotin
à la porte de ce comité pour accuser les	meilleurs	patriotes . les amis de la royauté , l
artes à la convention , en dénonçant les	meilleurs	patriotes pour faire revenir sur l ' e
elle branches se raccrocher accusent les	meilleurs	patriotes pour qu ' on les oublie , ou
, et tous les nains qui asticotaient les	meilleurs	patriotes sont rentrés cent pieds sous
ents de *pitt et *cobourg , accusent les	meilleurs	républicains , et les *sans - culottes
vu les plus vils intrigants accuser les	meilleurs	républicains ; je me doutais bien que
s sa fureur de mordre et de déchirer les	meilleurs	républicains on l ' étouffera comme un
ssassiner la patrie , qui ont égorgé les	meilleurs	républicains subirent tous le supplice

L'examen des contextes de la forme *meilleurs* dans cette même partie M6, montre que cette forme introduit une distinction, particulièrement insistante dans la période M6, parmi des catégories d'actants présentés jusqu'alors comme positives dans leur totalité. L'apparition de l'adjectif *meilleurs* devant chacun de ces groupes induit la question de l'hétérogénéité de la catégorie et la possibilité de l'existence d'intrus parmi les gens susceptibles de s'en réclamer. Pour plus de clarté, *les meilleurs X* sont d'ailleurs présentés comme faisant l'objet d'intrigues,

⁶ La Vendée est une province française qui a connu un soulèvement armé contre le pouvoir révolutionnaire. Ce soulèvement a fait l'objet d'une répression très dure, qui apparaît comme terminée à l'époque où sont produits ces textes.

⁷ Un examen similaire sur les contexte de la forme *nouveau* montre un fonctionnement identique de la forme (*nouveau complot, nouveau coup de chien*), cette fois dans la partie M5 du corpus.

de manigances, d'accusations diverses de la part d'ennemis plus difficiles à identifier. Dans ce sens, la spécificité de la forme *meilleurs* dans la partie M6 peut tout à fait être rattachée au vocabulaire « offensif » signalé plus haut.

2.5 Le cas de la forme *les*

La forme *les*, que la méthode des spécificités signale comme particulièrement représentée dans la période M6 du corpus *Duchn8M*, mélange, compte tenu de la méthode de segmentation en *formes graphiques* utilisée depuis le début de cette expérience, les occurrences de l'article défini pluriel *les* et celles du pronom personnel homographe. De plus, comme dans la plupart des textes rédigés en français que nous avons étudiés, cette unité est très fréquente dans le corpus (4 748 occurrences). Quelle signification peut-on accorder à la présence abondante des occurrences de cette forme dans les parties qui recourent fortement au vocabulaire que nous avons qualifié d'*offensif* ?

L'analyse des segments les plus fréquents commençant par la forme *les* nous donne quelques indications. Comme on pourra le vérifier en annexe B, en se reportant à une liste plus fournie de ces segments, l'article défini *les* introduit souvent les actants, en majorité hostiles, que nous avons évoqués plus haut en les constituant comme des groupes dont l'unité ne peut pas être remise en cause au moment de l'énonciation : *les patriotes* (91 occ.), mais aussi : *les traîtres* (86 occ.), *les aristocrates* (75 occ.), *les fripons* (64 occ.), *les brigands* (63 occ.), etc..

Par conséquent, la condensation des occurrences de , la ventilation de cette forme graphique constitue également un indicateur de l'accroissement du discours offensif que nous avons évoqué plus haut. Ce que l'on peut vérifier dans le paragraphe présenté ci-dessous à titre d'exemple.

[§ 487] /.../ **les** *brissotins, comme **les** jean-foutres de conseillers de *capet, ont persécuté **les** patriotes et **les** ont accusé également d'être des anarchistes. on se sert aujourd'hui d'un autre mot pour avilir **les** ardents républicains. on **les** appelle des ultra-révolutionnaires, parce que **les** mots de factieux et de désorganisateur sont usés; mais foutre, le peuple ne prendra pas le change, ce sont encore **les** apôtres du modérantisme, **les** amis de la royauté et **les** aristocrates déguisés qui **les** accusent. /.../

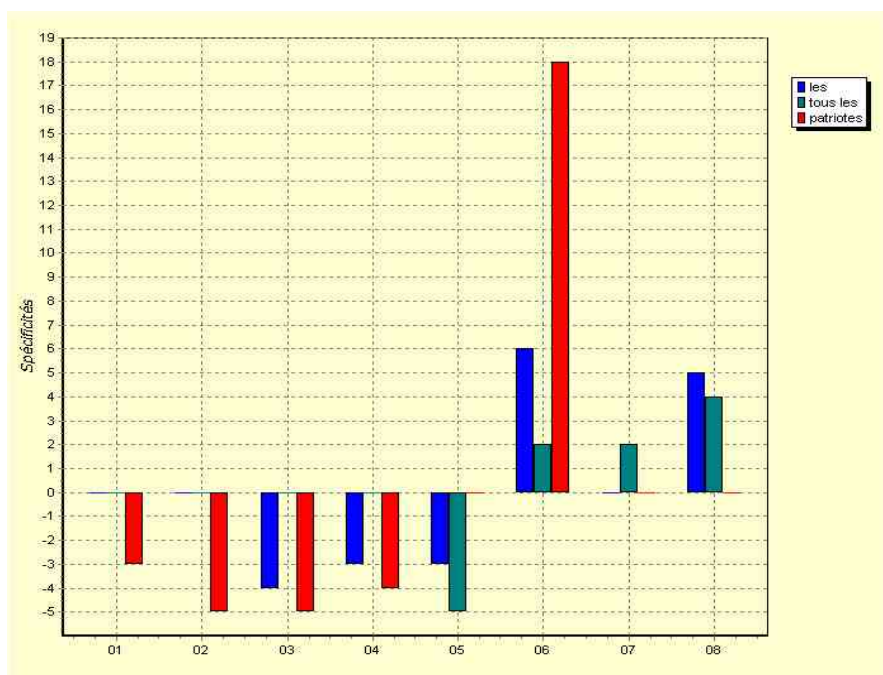


Figure 7

Ventilations des termes *patriotes, les* et *tous les*
dans les 8 périodes du corpus *Duch8M*

3 Faire varier l'échelle d'analyse

Comme on l'a vu plus haut, la division du corpus *Duchn* en 8 périodes d'un mois permet de mettre en évidence une évolution du vocabulaire employé au fil du temps qui n'était pas repérable sur le corpus découpé en unités trop fines (les 96 numéros). On peut voir sur la figure XX, le résultat d'une analyse portant sur le même corpus découpé cette fois en 16 parties correspondant chacune à un intervalle de temps de deux semaines. Les quinzaines ont été numérotées à l'aide de deux chiffres. Le premier de ces chiffres correspond au mois de parution également mentionné dans l'analyse en 8 parties. Le second indique s'il s'agit de la première ou de la seconde quinzaine.

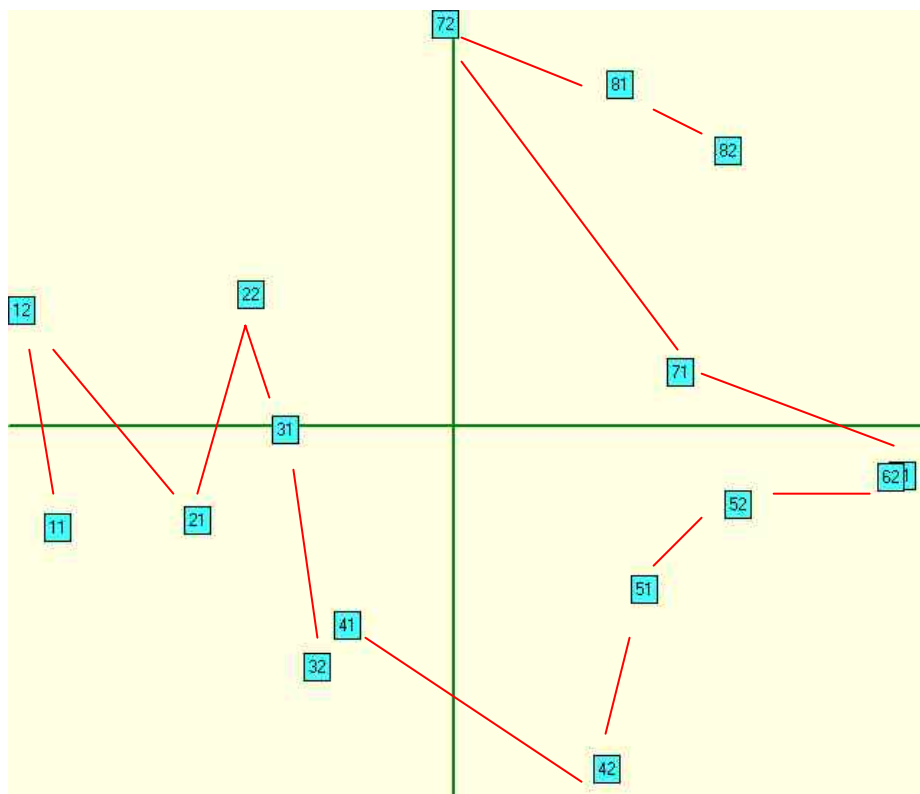


Figure 8

AFC sur le corpus *Duchn*

16 périodes d'une semaine x 1420 formes de fréquence ≥ 10

Comme on le voit sur cette seconde analyse, la partition du corpus en unités plus fines amène des résultats un peu moins lisibles, du fait de la plus grande variabilité de ces parties. En revanche, l'analyse permet cette fois de porter des jugements sur des périodes de temps plus courtes et par la même de tenter de produire un commentaire qui colle de plus près à la chronologie des événements.

Tout le problème devient alors de faire le départ entre les variations qui traduisent des modifications chronologiques survenues sur des périodes courtes, hautement pertinentes pour le chercheur, et des variations qui sont à mettre sur la réduction de la taille des unités

statistiques dans lesquelles sont effectuées les décomptes.⁸ On note que pour cette seconde analyse opérée sur le découpage en quinzaines, les parties proches dans le temps occupent toujours des positions voisines sur le premier plan factoriel. Si l'on conserve le modèle d'interprétation chronologique proposée à partir de l'analyse en 8 parties, on peut préciser le schéma d'évolution proposé plus haut en situant désormais dès la première quinzaine de la partie 7 le repli opéré par le Père Duchesne par rapport à l'évolution chronologique d'ensemble du vocabulaire.

4 Organiser le retour au contexte

Il nous reste à vérifier en retournant au contexte que l'hypothèse de la montée d'un vocabulaire offensif rend bien compte de l'évolution lexicale au sein du corpus étudié.

Sur le tableau XX nous avons projeté le type constitué par les formes les plus spécifiques de la partie M6. Le logiciel a ensuite calculé, pour chaque section, la spécificité attachée à ce type dans chacun des paragraphes du corpus. Un seuillage de ce diagnostic a permis de colorier de manière différenciée les paragraphes. Les paragraphes en couleur sombre contiennent en principe un nombre important de formes et de segments appartenant au type étudié (ici, le type : formes les plus fréquentes du vocabulaire spécifique de la partie M6). Remarquons qu'un certain nombre de ces parangons n'appartiennent pas à la partie M6. On en conclura que ces parangons extérieurs anticipent (pour ceux qui sont situés dans la période M5, par exemple) l'expression particulière qui va avoir lieu dans la période M6. Le parangon situé dans la période M8 constitue au contraire un écho de cette expression

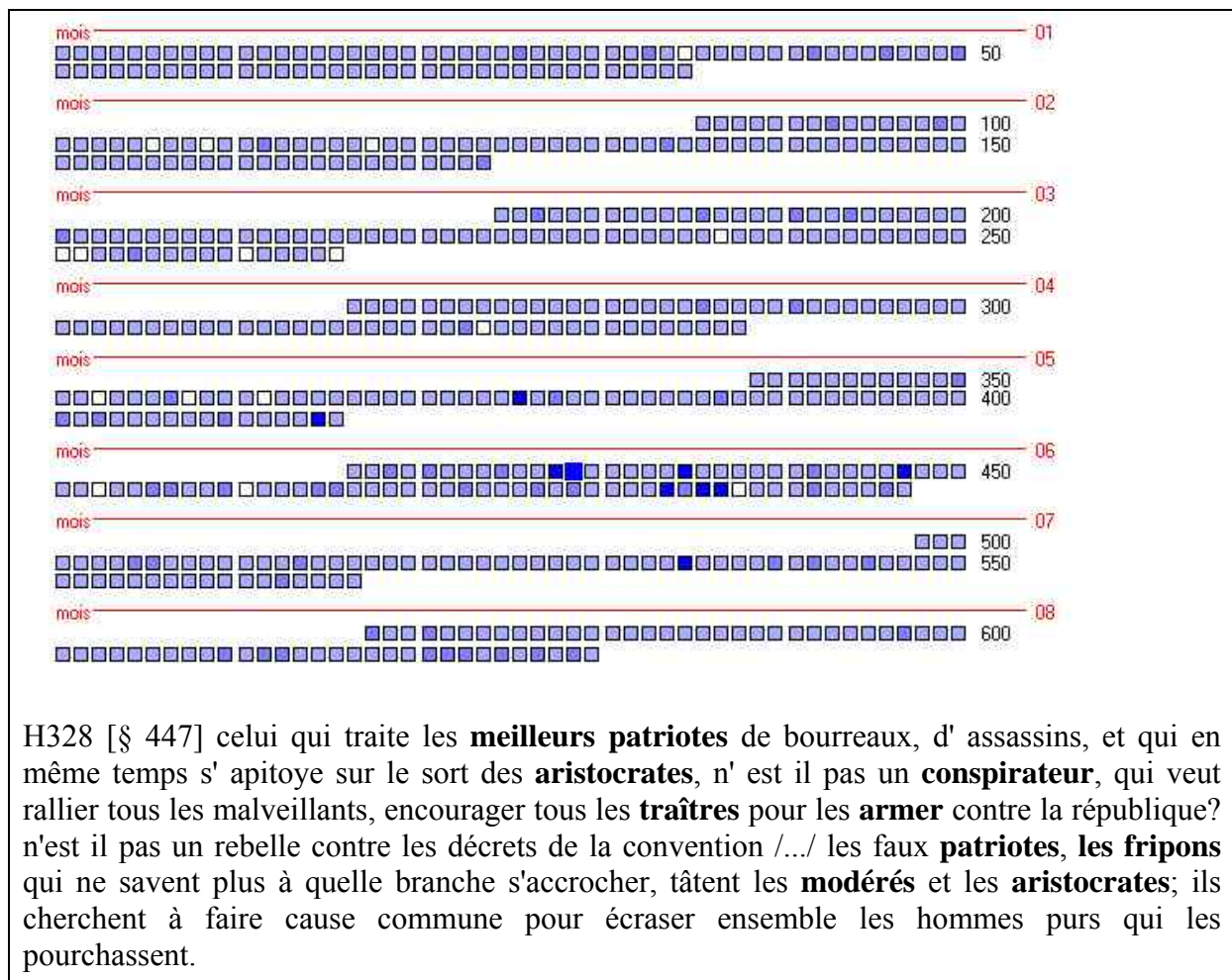
Parmi d'autres parangons signalés en couleur sombre, le paragraphe sélectionné au tableau XX révèle une tonalité particulièrement offensive. Aux formes qui ont servi à construire le type viennent s'ajouter plusieurs formes comme : *rebelles, malveillants, bourreaux, assassins, écraser* que leur faible fréquence a écarté de notre sélection mais qui auraient tout à fait leur place dans liste des termes polémiques.

==== Lexico3 ==== Repérage des parangons pour un type donné

- ✓ Construire le type que l'on désire spatialiser (forme, groupe de formes, liste ...)
- ✓ Construire une carte des sections en fixant un délimiteur de section (§ 2.1)
- ✓ Traîner le type sur la carte des sections
- ✓ Cocher la case « seuillage » (en haut sur la carte des sections) après avoir éventuellement modifié le niveau des seuils (bouton à gauche de cette case)

⁸ Plusieurs modèles statistiques ont été proposés par les statisticiens pour aider les chercheurs face à ce type d'interrogation.

Tableau 5 :
Repérage des parangons pour un type donné



4.1 Les types caractéristiques

Une *sélection* ou *zone textuelle* est un ensemble d'occurrences sélectionnées dans le corpus. Une zone peut être constituée par un ensemble de parties, de paragraphes ou de phrases consécutives ou non ou encore par une sélection arbitraire d'occurrences éparpillées dans le texte. Pour une zone donnée, on appellera *type caractéristique*, ou *type spécifique* toute liste des formes constituées par les formes les plus caractéristiques de la zone.

Dans l'exemple ci-dessus, les formes les plus caractéristiques prélevées parmi les spécificités majeures de la partie M6, dont nous avons analysé la ventilation dans les paragraphes du corpus constituent donc un type spécifique pour la zone M6.

Le type caractéristique présenté plus haut nous a permis de localiser les parangons de la partie M6 (paragraphes particulièrement saturés en formes spécifiques de la partie M6). Le graphique présenté à la figure 9 permet d'apprécier l'évolution de ce que nous avons appelé le vocabulaire offensif dans un découpage plus fin du corpus.

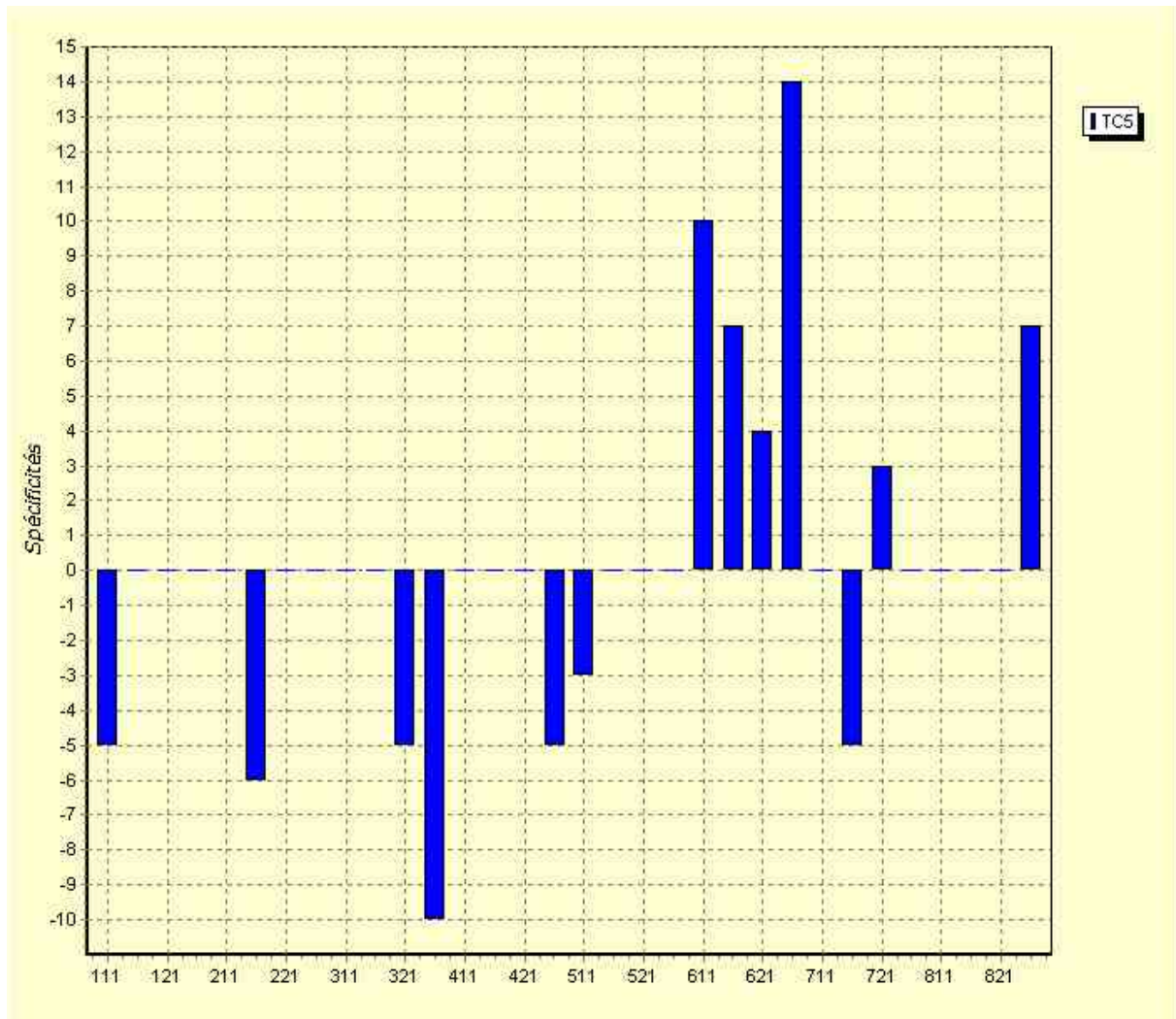


Figure 9 :

Ventilation dans les 32 semaines du corpus *Duchn* du type spécifique T5 (M6)
(formes de spécificité $<1/10^5$ dans la partie M6)

L'histogramme présenté à la figure XX permet avec plus de précision l'apparition et l'augmentation du vocabulaire *offensif* introduit plus haut.

==== *Lexico3* ==== Etude de la variation des types caractéristiques

- ✓ Construire le type que l'on désire étudier (forme, groupe de formes, liste ...)
- ✓ Construire une partition en choisissant une clé
- ✓ Traîner le type sur le graphique
- ✓ Choisir le mode de représentation (fréquence relative, absolue ou spécificité)

5 Proposer une interprétation

Dans le cadre de ce seul paragraphe, nous tenterons d'opérer un rapprochement entre les particularités observées dans la répartition des formes lexicales mises en évidence par les analyse textométriques et les principaux événements survenus au plan politique dans la

période couverte par le corpus. En d'autres termes, d'esquisser une tentative d'interprétation au plan historique des phénomènes constatés⁹.

Les premières périodes du corpus (M1-Mx) voient un renforcement des positions défendues par l'auteur du *Père Duchesne* au détriment des ennemis politiques qu'il ne cesse de dénoncer (*Brissotins*, i.e. partisans de Brissot, etc.). Les 31 députés girondins qui constituent le parti des *modérés*, parmi lesquels Brissot, seront exécutés le xxx qui correspond à la période Mxx du corpus. Après cette victoire politique, Hébert entre en opposition avec le mouvement jacobin et plus particulièrement avec les montagnards du Comité de salut public (Robespierre, St Just, etc.). Dans le but d'influencer l'action du Comité de salut public, il redouble d'ardeur dans la dénonciation des *conspirateurs*, des *faux patriotes*, des *nouvelle(s) clique(s) de modérés*, de ceux qui veulent allumer une *nouvelle Vendée* et appelle à l'union des *meilleurs patriotes*.

Les jacobins réagiront d'abord par la voix de Camille Desmoulin qui attaquera Hébert dans son journal *Le vieux cordelier* (période M7)¹⁰. Le *Père Duchesne* tentera dans un premier temps de reprendre l'offensive en faisant à nouveau appel aux thèmes qui ont assuré sa popularité des premières périodes : lutte pour la déchristianisation, lutte contre la présence de nobles à la tête des armées. Devant l'échec de cette tentative, il reprendra, sans succès au cours de la dernière période M8 qui verra son arrestation et son exécution, l'offensive amorcée en M6.

6 Conclusion

L'exploration du corpus *Duchn*, à l'aide des méthodes textométriques met en évidence une importante évolution du vocabulaire au cours des huit mois sur lesquels s'étend le corpus.

Les analyses quantitatives sur la partition en 96 livraisons, mettent en évidence des différences stylistiques liées à une alternance de genre entretenue par l'auteur du corpus. De ce fait, elles ne permettent pas d'apprécier l'évolution lexicale du corpus.

Un regroupement des livraisons en périodes de 30 jours consécutifs permet par sa part de cerner l'évolution lexicale de manière nettement plus satisfaisante. Les méthodes quantitatives permettent alors tout à la fois : de mettre en évidence un vocabulaire offensif qui trouvera un emploi particulièrement remarquable dans la période M6. Le retour au contexte permet de préciser ces analyses.

7 Références

Lamalle C., Salem A., « Types généralisés et topographie textuelle dans l'analyse quantitative des corpus textuels », in *actes des 6emes journées d'analyse statistique des données textuelles*, Inria, St Malo, 2002

<http://www.cavi.univ-paris3.fr/lexicometrica>, 1997

⁹ Cette tentative d'interprétation est née dans le cadre d'une collaboration étroite et ancienne entre l'auteur de ces lignes, nullement spécialiste de ces problèmes et des historiens spécialistes de la période concernée, tout particulièrement J. Guilhaumou, connaisseur des langages de la Révolution française.

¹⁰ On trouve d'ailleurs, dans le journal de Hébert, des traces de cette polémique en M6 et M7 qui concernent, le /gentil /renégat/ benêt/ niais/ Camille.

8 Principales fonctionnalités Lexico3 utilisées

<i>N°</i>	<i>Fonctionnalité</i>	<i>Résultat</i>
<i>2</i>	Partition (clé a, pour année)	
<i>5</i>	Principales car lexicom (PCLC)	<i>Tableau 2</i>
<i>5.6</i>	Accroissement du vocabulaire (corpus)	<i>Figure 1</i>
<i>5.6</i>	Accroissement du vocabulaire (P92, P93)	<i>Figure 2</i>
<i>4</i>	Segments Répétés (seuil minimal =2)	
<i>8</i>	Sélection d'un Type (occurrence de SR long>10)	
<i>7</i>	Carte des sections (paragaphes, présence SR de long>10)	<i>Figure 3</i>

Annexe

