

***Dix années de recherche
sur une unité textuelle englobante,***

le « motif » :

acquis et perspectives

Dominique LONGRÉE, LASLA – Université de Liège

0. (Un peu plus de) 10 années d'une recherche collective , BCL - LASLA

- recherche collective sur le « motif » :
 - UMR 7320 – BCL « Bases, corpus, langage » - CNRS –Université de Nice Sophia Antipolis – S. Mellet
 - LASLA (Laboratoire d'Analyse Statistique des Langues Anciennes) de l'Université de Liège.
- des fichiers lemmatisés et annotés morphosyntaxiquement :
 - fichiers latins du LASLA
- des méthodes et des outils informatiques
 - développés par BCL : Hyperbase-Latin (S. Mellet – E. Brunet)
 - Hyperbase Web Edition
 - Logiciel d'Analyse arborée de X. Luong

0. (Un peu plus de) 10 années d'une recherche collective , BCL - LASLA

- un point de départ : une recherche sur les « sous-genres » au sein du corpus des historiens latins classiques

Longrée D. (2003), « Tacite et Suétone : étude linguistique comparative et genres littéraires », in G. Lachenaud et D. Longrée (éds), *Grecs et Romains aux prises avec l'histoire, Représentations, récits et idéologie*, (Collection Histoire), Rennes, P.U.R., Vol. 1, pp. 315-326.

Longrée D. (2004), « Une approche statistique de la concurrence entre démonstratifs chez les historiens latins (César, Salluste, Tacite) », in C. Bodelot (éd.), *Anaphore, cataphore et corrélation en latin, Actes de la Journée d'étude de Linguistique latine, Université Blaise Pascal - Clermont-Ferrand 2, 7 janvier 2003*, (Collection « Erga », Recherches sur l'Antiquité, 6), Clermont, Presses Universitaires Blaise Pascal, pp. 157-178.

Longrée D. (2005), « Temps verbaux et spécificités stylistiques chez les historiens latins : sur les méthodes d'analyse statistique d'un corpus lemmatisé », in G. Calboli (éd.), *Papers on Grammar, IX, 2, Latina Lingua !, Proceedings of the Twelfth International Colloquium on Latin Linguistics (Bologna, 9-14 June 2003)*, Roma, pp. 863-875.

0. (Un peu plus de) 10 années d'une recherche collective , BCL - LASLA

➤ un objectif : calculer des distances entre textes, et les caractériser de manière endogène , en vue :

- d'une comparaison,
- d'une classification,
- d'une confrontation avec les « sous-genres » établis a priori

➤ déterminer l'origine des variations entre textes : pouvant être liées

- aux sous-genres littéraires

mais aussi :

- à la nature ou au statut narratif des faits rapportés
- à des particularités d'écriture / d'expression des auteurs
- à une évolution diachronique

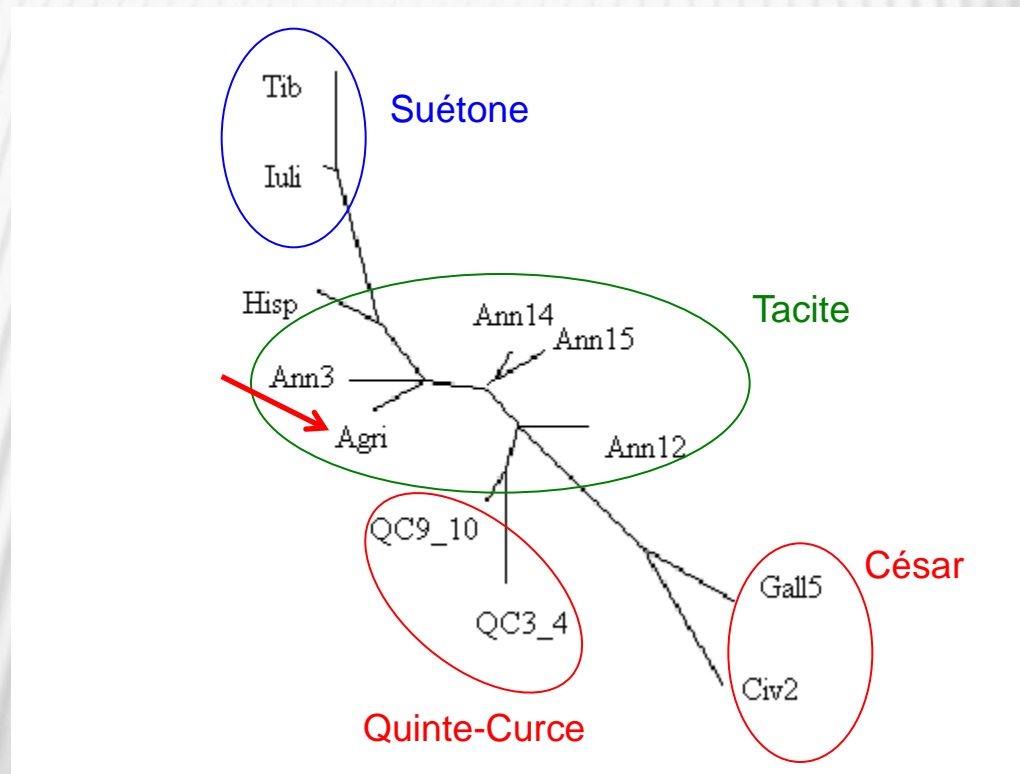
0. (Un peu plus de) 10 années d'une recherche collective , BCL - LASLA

- en faisant appel à des méthodes d'analyse statistique des données textuelles

- en considérant tout d'abord , selon une méthode éprouvée, le texte comme un « sac de mots » au sein duquel effectuer des dénombrements

- une étude des « fréquences » :
 - de formes/mots
 - de lemmes
 - de codes grammaticaux

1. Atouts et limites de l'étude des fréquences



Distribution des temps de l'indicatif dans un corpus de 12 textes historiques

2. Le texte : un espace ordonné (et multidimensionnel)

Un texte n'est pas un sac dans lequel seraient rassemblées en vrac ses unités constitutives, ni seulement un réseau :

4 passés simples 5 imparfaits : Imp. Imp. PS PS Imp. PS. PS. Imp. Imp.

Je n'**hésitais** point. Je **connaissais** une hôtellerie d'amour non loin de ma demeure, et j'y **courus**, et j'y **entraï** comme font ces gens qui se jettent à l'eau pour voir s'ils savent encore nager. Je **nageais**, et fort bien. Et je **demeurai** là longtemps (...). Puis je me **trouvai** dans la rue (...). Je me **sentais** maintenant calme et sûr de moi (...), et prêt encore, me **semblait-il**, pour des prouesses (Maupassant, *Toine*).

4 passés simples 5 imparfaits : PS PS PS PS Imp. Imp. Imp. Imp. Imp.

Toute cette cavalerie, étendards et trompettes au vent, (...) **descendit** la colline de la Belle-Alliance, **s'enfonça** dans le fond redoutable (...), et **disparut** dans la fumée, puis, sortant de cette ombre, **reparut** de l'autre côté du vallon (...). Ils **montaient**, graves, menaçants, imperturbables; dans les intervalles de la mousqueterie et de l'artillerie, on **entendait** ce piétinement colossal. Etant deux divisions, ils **étaient** deux colonnes : la division Wathier **avait** la droite, la division Delord **avait** la gauche (Hugo, *Les misérables*).

2. Le texte : un espace (ordonné et) multidimensionnel

formes : j'y entrai comme font ces gens qui se jettent à l'eau

lemmes : JE Y ENTRER COMME FAIRE CE GENS QUI SE JETER A LE EAU

POS : Pron. Adv. Verbe Conj.Sub. Verbe. Dét. Nom. Rel. Pron. Verbe Prép. Art. Nom

Morph.: Pron.1èrepers.sing. Adv. VerbeIndImprfaitActif1èreperssing.....

Syntaxe : Sujet Compl. Lieu Verbe Sub. Verbe Cdét. Sujet.....

Syntaxe : Princ. Subord. Comparative.....

etc.

- Comment prendre en compte ces deux caractéristiques du texte ?

2. Le texte : un espace ordonné (et multidimensionnel)

Long	Limite	Effectif	Numeré	Lon	Clc sur une ligne > graphique	Fréqenc	Segments sélectionnés (cliquer sur l'un d'eux pour voir le texte)
L8	2	78	272	1	8	13 4	*a111 *a211 *a211 *a312 *l11 *d112 *s *t
L7	6	2	18	2	8	3 4	*a132 *s *c68*1 *b6*1111300 *p *a212 *a341 *s
L6	8	6	90	3	8	3 4	*a161 *b2*2*3123ga *s *k*32 *r *g*31 *b1*2*3113ms *a432
L5	10	38	706	5	8	13 4	*a211 *a111 *a211 *a211 *a312 *l11 *d112 *s
L4	12	688	14745	6	8	10 1 12 1 15 1 36 1	*a211 *a211 *a211 *a211 *a211 *a211 *a211 *a211
L3	14	1897	51054	8	8	13 4	*a211 *a211 *a312 *l11 *d112 *s *t *l*52
L2	16	2636	115141	9	8	3 4	*a212 *a341 *s *c1121 *a212 *a541 *b5*2*3113aa *s
Tout --	5345	182026		10	8	3 4	*a212 *a541 *b5*2*3113aa *s *a141 *a511 *a311 *s
				11	8	3 3 26 1	*a212 *b1*2*1113zd *s *l*32 *c1621 *r *a162 *b3*2*1113zd
				14	8	3 4	*a241 *a132 *s *c68*1 *b6*1111300 *p *a212 *a341
				17	8	3 4	*a331 *a241 *a132 *s *c68*1 *b6*1111300 *p *a212
				19	8	3 4	*a341 *s *c1121 *a212 *a541 *b5*2*3113aa *s *a141
				21	8	3 4	*a362 *b6*11113mg *m***1 *l11 *a311 *b6*1111300 *l*41 *j*11
				23	8	3 3 26 1	*b1*2*1113zd *s *l*32 *c1621 *r *a162 *b3*2*1113zd *m***1
				26	8	3 4	*b2*2*3123ga *s *k*32 *r *g*31 *b1*2*3113ms *a432 *b5*2*3113ms
				32	8	3 4	*b6*1111300 *p *a212 *a341 *s *c1121 *a212 *a541
				33	8	3 4	*b6*11113mg *m***1 *l11 *a311 *b6*1111300 *l*41 *j*11 *b6*111413n
				36	8	16 1 26 2 29 1	*c1113 *l11 *a511 *c1521 *a352 *a241 *d2113 *b5*1111300
				37	8	3 4	*c1121 *a212 *a541 *b5*2*3113aa *s *a141 *a511 *a311
				38	8	3 3 26 1	*c1621 *r *a162 *b3*2*1113zd *m***1 *e*12 *r *a361
				39	8	3 4	*c68*1 *b6*1111300 *p *a212 *a341 *s *c1121 *a212
				44	8	3 4	*l*31 *s *a331 *a241 *a132 *s *c68*1 *b6*1111300
				49	8	3 4	*k*61 *a161 *b2*2*3123ga *s *k*32 *r *g*31 *b1*2*3113ms
				50	8	3 4	*k*62 *s *k*61 *a161 *b2*2*3123ga *s *k*32 *r
				51	8	3 3 26 1	*l*32 *c1621 *r *a162 *b3*2*1113zd *m***1 *e*12 *r
				55	8	3 4	*m***1 *l11 *a311 *b6*1111300 *l*41 *j*11 *b6*111413n *m***1
				59	8	3 3 26 1	*n *a212 *b1*2*1113zd *s *l*32 *c1621 *r *a162
				62	8	3 4	*p *a212 *a341 *s *c1121 *a212 *a541 *b5*2*3113aa
				63	8	3 3 26 1	*r *a162 *b3*2*1113zd *m***1 *e*12 *r *a361 *b2*2*111100
				65	8	3 4	*r *k*62 *s *k*61 *a161 *b2*2*3123ga *s *k*32
				67	8	3 4	*s *a331 *a241 *a132 *s *c68*1 *b6*1111300 *p
				68	8	3 4	*s *c1121 *a212 *a541 *b5*2*3113aa *s *a141 *a511
				69	8	3 4	*s *c68*1 *b6*1111300 *p *a212 *a341 *s *c1121
				72	8	3 4	*s *k*32 *r *g*31 *b1*2*3113ms *a432 *b5*2*3113ms *s
				73	8	3 4	*s *k*61 *a161 *b2*2*3123ga *s *k*32 *r *g*31
				74	8	3 3 26 1	*s *l*32 *c1621 *r *a162 *b3*2*1113zd *m***1 *e*12
				75	8	3 3 26 1	*s *n *a212 *b1*2*1113zd *s *l*32 *c1621 *r
				77	8	3 4	*t *e*51 *a512 *b6*2*3113dd *r *a231 *m***1 *l*12

Un rappel : les segments répétés (A. Salem) – formes - lemmes - codes syntaxiques mais contigus

Les segments répétés

Commencer par l'opération "Préparation" qui n'a lieu qu'une fois pour la même base et fournit les données aux autres fonctions. Patience! Le programme "Préparation" demande du temps...

3. Une première approche : les séquences

- s'applique essentiellement à des données morphosyntaxiques
- avec une réduction du texte à une suite d'annotations morphosyntaxiques d'où ≠ segments répétés

par exemple, à la succession des temps verbaux dans les propositions principales :

4 passés simples 5 imparfaits : Imp. Imp. PS PS Imp. PS. PS. Imp. Imp.

Je n'**hésitais** point. Je **connaissais** une hôtellerie d'amour non loin de ma demeure, et j'y **courus**, et j'y **entraï** comme font ces gens qui se jettent à l'eau pour voir s'ils savent encore nager. Je **nageais**, et fort bien. Et je **demeurai** là longtemps (...). Puis je me **trouvai** dans la rue (...). Je me **sentais** maintenant calme et sûr de moi (...), et prêt encore, me **semblait-il**, pour des prouesses (Maupassant, *Toine*).

3. Une première approche : les séquences

La distribution des temps verbaux : un maillage fondamental de la narrativité

Début du livre 3 de la *Guerre Civile* de César

11 12 **14 14 14** 15 11 11 11 11 15 **14 14**

Début du livre 4 des *Annales* de Tacite:

12 12 12 **14** 15 15 **14 14** 34 **14** 12 11 11 12 **14 14 14** 12 11 12 **14 11**

Début de la *Vie d'Auguste* de Suétone

11 12 12 12 **14 14 14 14 14 14** 11 11 11 **14 14 14 14 14 14 14 14 14**

Début de la *Vie d'Othon* de Suétone

14 14 14 14 14 14 14 14 14 14 14 12 **14 14 14 14 14 14 14 11 14 14**

(avec 14 = code de parfait, 11 = code de présent, 12 = code d'imparfait,
15 = code de plus-que-parfait)

3. Une première approche : les séquences

Séquences cumulées

Tacite, *Agricola*

... parf. parf. parf. .parf. imparf. prés. imparf. parf. parf. parf. ...

5 séquences de 2 parfaits
/ 7 parfaits

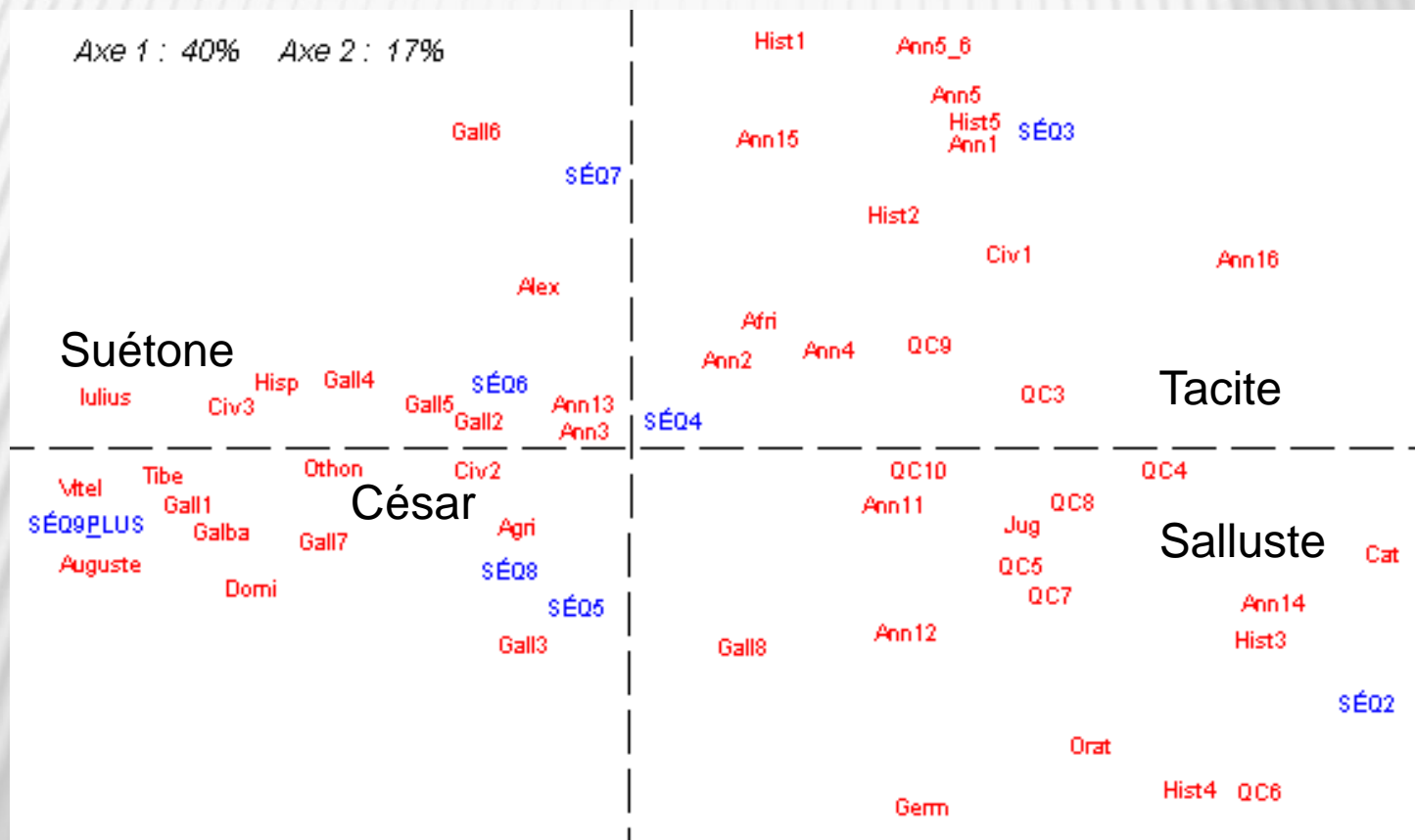
Séquences non cumulées

Tacite, *Agricola*

... parf. parf. parf. .parf. imparf. prés. imparf. parf. parf. parf. ...

1 séquence de 4 parfaits
1 séquence de 3 parfaits
/7 parfaits

3. Une première approche : les séquences non cumulées (de parfaits)



3. Une première approche : les séquences - bilan de l'analyse

- - mise en évidence de faits de style
- caractérisation d'un auteur

Longrée D.& Luong X. (2003), « Temps verbaux et linéarité du texte : recherches sur les distances dans un corpus de textes latins lemmatisés », in X. Luong (éd.), *La distance intertextuelle = Corpus*, 2, pp. 119-140.

Longrée D.& Luong X. (2005), « Spécificités stylistiques et distributions temporelles chez les historiens latins : sur les méthodes d'analyse quantitative d'un corpus lemmatisé », in G. Williams (éd.), *La Linguistique de Corpus*, (Rivages Linguistiques), Rennes, P.U.R., pp. 141-152

mais

- - un cadre unidimensionnel (comme pour les segments répétés)

d'où l'introduction d'un nouvel outil : le motif

utilisable à la fois de manière unidimensionnelle et multidimensionnelle

4. Le motif - définition

➤ Premières apparitions du concept :

Legallois D. (2006) : « Des phrases entre elles à l'unité réticulaire de textes », *Langages* 163, pp. 56-70

Longrée D., Mellet, S. et Luong, X.(2008), « Les motifs : un outil pour la caractérisation topologique des textes », dans S. Heiden et B. Pincemin (éds), *Actes des JADT 2008, 9èmes Journées internationales d'Analyse statistique des Données Textuelles, Lyon, 12-14 mars 2008*, Presses ENS-Lyon, pp. 733-744.

➤ une définition formelle : un motif =

- l'association récurrente de n éléments de (T) muni de sa structure linéaire
- un sous-ensemble ordonné de (T)

➤ un exemple :

- si $(T) \ni x$ occurrences de A, B, C, D, E, F
- motifs = récurrence de ABD, de AA, etc.

4. Le motif - définition

- une unité intégrée à une représentation topologique du texte *stricto sensu*

Mellet S. & Barthélemy J.-P. (2007) : « La topologie textuelle : légitimation d'une notion émergente », *Lexicometrica* (numéro spécial *Topographie et topologie textuelles*), revue en ligne, article consultable à l'adresse <http://www.cavi.univ-paris3.fr/lexicometrica/numspeciaux/special9/mellet.pdf>

Barthélemy J.-P., Longrée D., Luong X. & Mellet S. (2009) : « Représentation du texte pour la classification arborée et l'analyse automatique de corpus : application à un corpus d'historiens latins », *Mathematics and Social Sciences* 187 (3), 107-121.

- fournissant une base solide à des traitements automatiques

Evert S. (2008) : "Corpora and collocations", in A. Lüdeling & M. Kytö (eds), *Corpus Linguistics. An International Handbook*, article 58, vol. 2, Berlin: Mouton de Gruyter, 1212-1248.

4. Le motif - définition

- une notion élaborée dans une perspective discursive, d'où l'importance de la propriété de récurrence.

- admettant
 - la présence de variables en son sein
ici j'aimerais/voudrais/souhaiterais toutefois préciser, etc.
 - la permutation de deux éléments :
ici, toutefois, j'aimerais préciser
 - la présence vs absence d'un élément (opérations de suppression ou d'ajout) :
∅ j'aimerais toutefois préciser
ici j'aimerais ∅ préciser.
 - des variations sur les catégories grammaticales :
ici je souhaite toutefois préciser

- une construction lexico-grammaticale (Gledhill & Frath 2007) associée à un nombre restreint de formes

4. Le motif - définition

- constituant un « cadre collocationnel » accueillant un ensemble de paramètres à définir
- mais en fonction des possibilités des logiciels
- recherches sur des motifs syntaxiques : « cadratifs » et « rallonges »

Mellet S. & Longrée D. (2009) : “Syntactical Motifs and Textual Structures”, *Belgian Journal of Linguistics* 23 (*New Approaches in Textual Linguistics*), 161-173.

5. Les « premiers motifs » étudiés : unidimensionnels et syntaxiques

Les motifs « cadratifs »

César, *De bello gallico*, II, 25, 3 :

/ Cuius aduentu, **spe inlata** (T) militibus ac **redintegrato** (T) **animo**, cum pro se quisque in conspectu imperatoris etiam extremis suis rebus operam nauare **cuperet** (T), paulum hostium impetus **tardatus est** (&).

/ A l'arrivée de celui-ci, **l'espoir ayant été rendu** (T) aux soldats et **leur moral ayant été retrouvé** (T), **alors que** chacun pour lui-même **désirait** (T) apporter sa contribution sous le regard du général en chef, même si sa situation était la plus critique, l'assaut des ennemis **fut** (&) un peu **ralenti**.

/ TTT &

5. Les « premiers motifs » étudiés : unidimensionnels et syntaxiques

Les motifs « de rallonge » »

Tacite, *Annales*, I, 2, 1 :

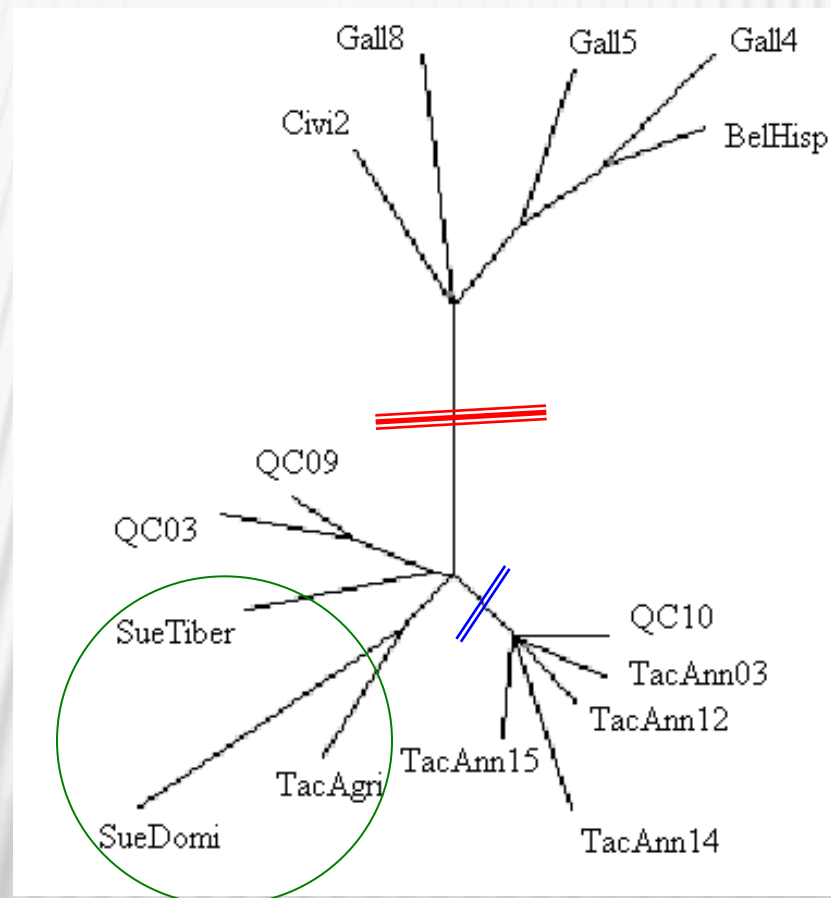
...insurgere paulatim, munia senatus magistratuum legum in se **trahere (&)**, **nullo aduersante (T)**, **cum** ferocissimi per acies aut proscriptione **cecidissent (T)**, ceteri nobilium, quanto quis seruitio promptior, opibus et honoribus **extollerentur (T)** ac nous ex rebus aucti tuta et praesentia quam uetera et periculosa **mallent (T) /**

...il **commença** à s'élever peu à peu, à **attirer (&)** à lui l'autorité du sénat, des magistrats, des lois, **nul** ne lui **résistant (T)**, **alors que** les plus acharnés **avaient péri (T)** par la guerre ou la proscription, que tous les autres nobles, en fonction de l'empressement de chacun à servir, **s'élevaient (T)** en richesses et honneurs et que, ayant gagné par ce changement politique, ils **préféraient (T)** la sécurité présente aux périls passés /.

& T T T T /

5. Les « premiers motifs » étudiés : unidimensionnels et syntaxiques

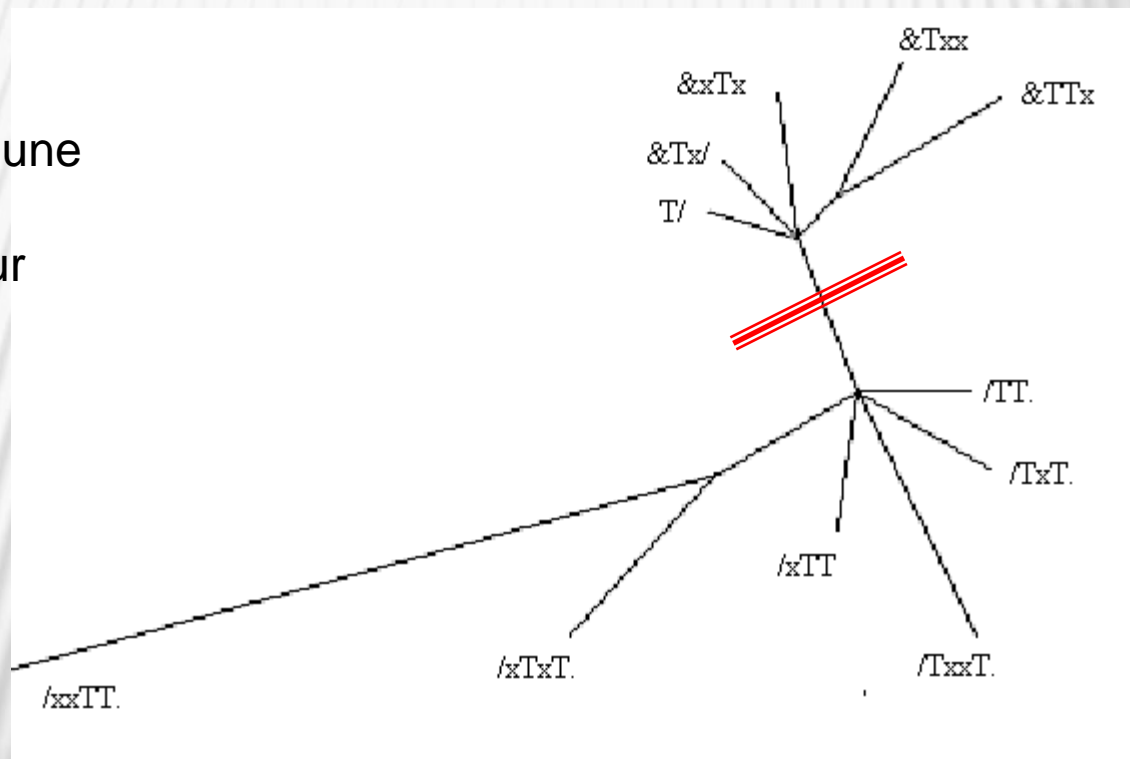
Caractérisation et classification des textes par l'emploi des « motifs » cadratifs et de « rallonge »



Classification de 15 textes d'historiens selon leur emploi de 11 motifs temporels cadratifs ou de rallonge

5. Les « premiers motifs » étudiés : unidimensionnels et syntaxiques

Confirmation : une bipartition des motifs selon leur emploi dans le corpus

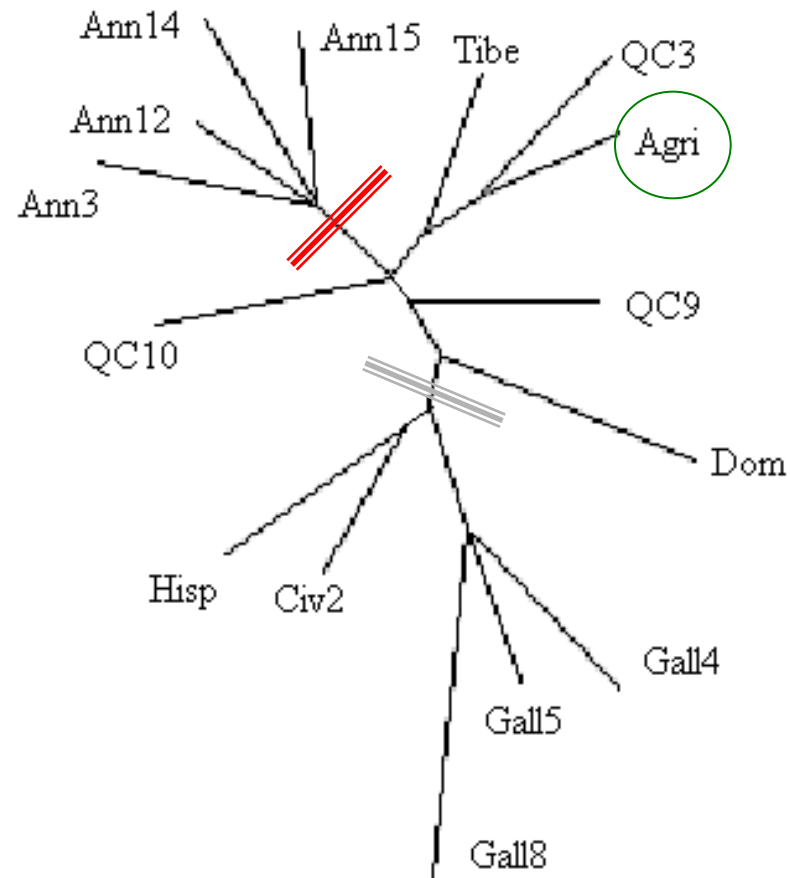


Classification des 11 motifs selon leur distribution dans les 15 textes du corpus

5. Les « premiers motifs » étudiés : unidimensionnels et syntaxiques

Confirmation : un résultat comparable avec des « motifs » cadratifs et de rallonge fondés sur deux propositions temporelles particulières

Classification des 15 textes du corpus en fonction de leur usage des 18 motifs associés aux circonstancielles ablatif absolu et *cum* + subjonctif



6. Le motif « textuel »

➤ une rencontre avec la phraséologie (Conscila 2011) menant à deux constats
Longrée D. & Mellet S. (2013) : « Le motif : une unité phraséologique englobante ? Etendre le champ de la phraséologie de la langue au discours », *Langages* 189, pp. 65-79.

1. le motif constitue un « cadre collocationnel » accueillant un ensemble de paramètres à définir, susceptible d'englober toute forme idiomatique non entièrement figée :

- collocations non figées (Sinclair 1991 : 111-112)
- « phrasèmes pragmatiques » (Mel'čuk 1998) ou « phrasèmes » (Depecker 1999)
- « idiomes morpho-syntaxiques » de Fillmore, Kay & O'Connor (1988)

2. On peut distinguer cette notion « englobante » de « motif » de celle de « motif textuel », une unité discursive

- qui n'est pas stockée « en langue » (dans le vocabulaire d'une langue)
- caractérisée à la fois par sa forme et par sa fonctionnalité,
 - fonction structurante
 - fonction caractérisante

6. Le motif « textuel » multidimensionnel

➤ fonction structurante d'un motif cohésif résumptif

Ablatif absolu en tête de phrase, composé d'un relatif de liaison ou d'un démonstratif suivi du substantif *res* (« choses », « faits », « événements ») et du participe passé passif d'un verbe signifiant soit « faire, achever, réaliser », soit « apprendre, connaître, découvrir »

Quibus rebus cognitis / auditis / nuntiatis, [...] : « et une fois ces choses connues, apprises, annoncées »

Qua re cognita, audita, animadversa [...] : « et une fois cette chose connue, apprise, remarquée »

Quibus rebus perfectis / confectis, [...] : « et une fois ces choses achevées, terminées »

His rebus gestis / ita gestis / ita actis « ces événements ayant (ainsi) eu lieu »

6. Le motif « textuel » multidimensionnel

➤ objectif : faire progresser la narration

Quibus rebus cognitis, media nocte silentio profectus, ad hostium castra mane peruenit (César, *Guerre des Gaules* 7, 18, 2)

« A cette nouvelle, César partit au milieu de la nuit et parvint au matin près du camp ennemi. »

Qua re nuntiata, Caesar celeriter sarcinas conferri, arma expediri iussit (César, *Guerre des Gaules* 7, 18, 4)

« Quand César l'apprit, il fit promptement rassembler les sacs et prendre la tenue de combat. »

His rebus ita gestis, P. Dolabella praetor interdixit (...) DE VI HOMINIBVS ARMATIS (...) (Cicéron, *Pro Caecina*, 23, 1 : partie narrative d'une plaidoirie)

« A la suite de ces événements, le préteur P. Dolabella délivra l'interdit sur la 'violence commise avec des gens armés'. »

6. Le motif « textuel » multidimensionnel

➤ fonction structurante d'un motif transitionnel narratif :

➤ changement de décor et ouverture d'un nouvel épisode narratif

[*Dum* (« pendant que ») + démonstratif neutre au nominatif + locatif + *geruntur* (« se déroulent »)] - *Dum haec in Gallia geruntur* : « Tandis que ces événements se déroulent en Gaule »

Dum haec Romae geruntur, [...] : « Tandis que ces événements se déroulent à Rome »

Dum haec ad Ilerdam geruntur, [...] : « Tandis que ces événements se déroulent près d'Ilerda »

Dum haec per provincias a Vespasiano ducibusque partium geruntur, [...] :
« Tandis que ces événements se déroulent dans les provinces sous la conduite de Vespasien et des chefs de partis »

Haec dum apud hostes geruntur, [...] : « Tandis que ces événements se déroulent du côté des ennemis »

Dum ea [Ø] geruntur/parantur, [...]

6. Le motif « textuel » multidimensionnel

➤ **fonction structurante d'un motif transitionnel, résumptif et argumentatif** :
quae cum ita sint - « dans ces conditions »

Quae cum ita sint, Catilina, perge quo coepisti : egredere aliquando ex urbe
(Cicéron, *Catilina* I, 10)

« Puisqu'il en est ainsi, Catilina, poursuis ce que tu as entrepris : sors une bonne fois de la ville »

Quae cum ita sint quaero abs te quam ob causam recusaris. (Cicéron, *Pro S. Roscio Amerino oratio*, 120, 1)

« Puisqu'il en est ainsi, je te demande pour quelle raison tu te récules »

Quae cum ita sint non enim dubito quin (...) (Cicéron, *In L. Calpurnium Pisonem oratio*, 94, 1)

« Dans ces conditions, je ne doute pas en effet que (...) »

Quae cum ita sint, de Caesare satis hoc tempore dictum habebō (Cicéron, *Philippiques* V, 52, 1)

« Dans ces conditions, j'aurai assez parlé de César pour l'instant »

6. Le motif « textuel » multidimensionnel

- **fonction caractérisante** : modalités expressives largement conditionnées
 - par les genres textuels et les registres de langue
 - par le style individuel d'auteur

- un exemple : [pronom relatif ou subordonnant comparatif + adverbe d'antériorité intradiégétique + verbe déclaratif au passé] – à traduire par « que nous avons évoqué plus haut », « comme je l'ai dit auparavant »

- variations :
 - alternance entre le relatif et le subordonnant comparatif
 - présence ou absence de l'adverbe anaphorique intradiégétique
 - forme de l'adverbe : *ante(a)* « auparavant » et *supra* « ci-dessus »
 - présence ou absence de l'adverbe *paulo* « peu »
 - choix du lexème verbal au sein d'un paradigme : *dico* « dire », *demonstro* « démontrer, argumenter, présenter », *memoro* « rappeler »
 - variations de la personne verbale : première personne du singulier; première personne du pluriel; troisième personne du passif impersonnel (« comme il a été montré plus haut »)

6. Le motif « textuel » multidimensionnel

ut ante dictum est / quem supra memorauit

Variantes du motif :

Comp_x_ante_1s_dico

Comp_x_ante_P_dico

Comp_supra_1s_dico

Comp_supra_P_dico

Comp_supra_1p_demonstro

Comp_supra_P_demonstro

Rel_x_ante_1s_dico

Rel_x_ante_P_dico

Rel_x_ante_1s_memoro

Rel_supra_1p_dico

Rel_supra_1p_demonstro

Corpus

Cicéron, *Discours*

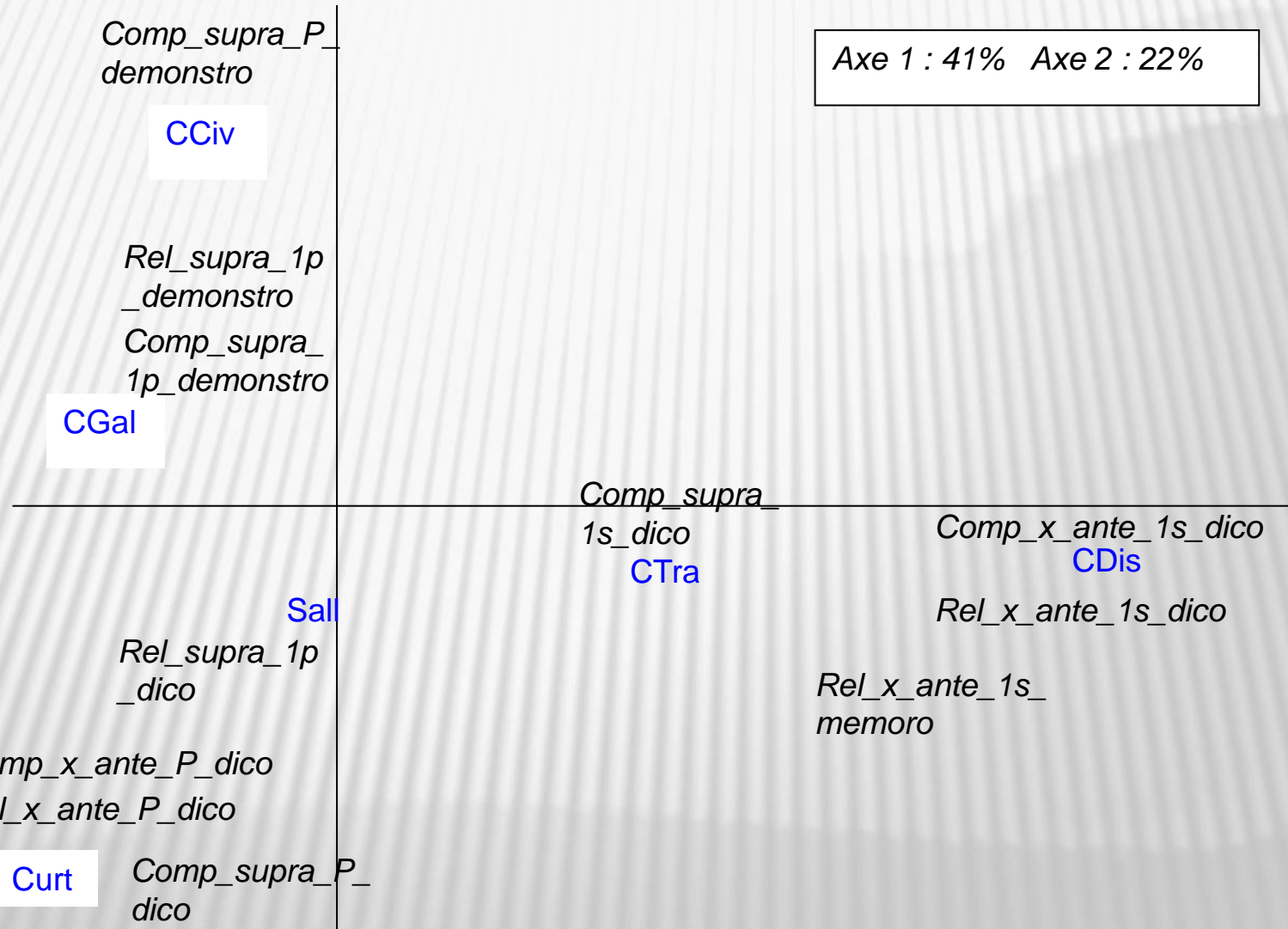
Cicéron, *Traité*s

César, *Guerre des Gaules*

César, *Guerre Civile*

Salluste, *Catilina et Jugurtha*

Quinte-Curce, *Histoires*



6. Le motif « textuel » multidimensionnel

- un jeu subtil entre
 - la forte unité d'un schéma abstrait sous-jacent, qui en tant que tel fournit un paramètre de caractérisation des textes (sa distribution permet un certain type de regroupements) et
 - la multiplicité des variantes qui offre à chaque écrivain une liberté expressive sur la base de laquelle d'autres regroupements, plus subtils, deviennent possibles.
- qui se marque d'autant moins que les motifs sont d'autant plus figés

Mellet S. & Longrée D. (2012) : « Légitimité d'une unité textométrique : le motif », in A. Dister, D. Longrée et G. Purnelle (éds) JADT 2012, Actes des 11èmes Journées internationales d'Analyse statistique des Données Textuelles, Liège, 716-728.

Longrée D & Mellet S. (2015) : « Les variantes des motifs chez les prosateurs latins, Entre récurrence générique et spécificité d'auteur, des formes révélatrices et caractérisantes », D. Longrée, S. Fialon et P. Pietquin (éd.), Langues anciennes et analyse statistique : cinquante ans après – Distances textuelles et intertextualités, LEC, 2015, pp. 65-88.

7. La question des fondements cognitifs

- si le motif textuel présente des fonctions structurantes ou caractérisantes, cela présuppose sa reconnaissance et donc sa mémorisation,
- ce qui pose la question des processus cognitifs en jeu :
 - est-ce seulement la récurrence du motif qui assure la mémorisation ?
 - est-ce que le type de fonction textuelle exercée par le motif la favorise, voire la conditionne ?
 - la représentation en mémoire du motif est-elle purement linéaire et de surface (séquence des formes constitutives) ?
 - ou repose-t-elle sur un schème construit par abstraction et généralisation ?
- pour tenter de répondre partiellement à ces questions, mise au point d'une expérimentation avec F. Lavigne, psychologue de BCL

Lavigne F., Longrée D. & Mellet S. (en préparation) : « Semantic Integration by Pattern Priming : Experiment and Cortical Network Model ».

7. La question des fondements cognitifs

- Le point de rencontre avec les psychologues :
Quibus rebus cognitis
- Un motif pour le latiniste, un “triplet” pour un psychologue
- constitué de deux amorces successives : *quibus* et *rebus*
- et d’une cible : *cognitis*
- les deux amorces pouvant susciter une attente plus ou moins grande (une activation neuronale, a “priming effect”, de la cible)
- avec plusieurs questions :
 - la cible est-elle amorcée plus fortement par *quibus rebus* que par *quibus* et *rebus* séparément ?
 - la fréquence du triplet augmente-t-elle le “priming effect” ?
 - le “priming effect” est-il plus important pour des unités phraséologiques que pour des motifs textuels ?

7. La question des fondements cognitifs

- Pour répondre à ces questions, nous avons mis au point un test de reconnaissance de mots latins :
- après avoir vu deux mots associés (amorces) sur un écran d'ordinateur, les participants ont du préciser si sur une chaîne de caractères apparaissant sur l'écran suivant était bien un mot latin ou non.
- Participants :
30 étudiant de Master en Langues et Littératures classiques de l'ULg et de l'ULB (formation identique avec des corpus d'apprentissage proches)
- Objectif : voir dans quelle mesure les temps de réponse allaient être moindres pour reconnaître la cible, par exemple *cognitis*, quand celle-ci apparaissait après les amorces *quibus rebus* que quand elle apparaissait après une séquence de deux autres mots ne formant pas "motif" avec elle.

7. La question des fondements cognitifs

- Utiliser des apprenants d'une langue seconde "morte" présentait paradoxalement des avantages :
 1. un corpus d'apprentissage clos, relativement similaire pour tous les participants
 2. la possibilité de calculer des forces d'association entre les amorces et la cible en fonction de la fréquence d'apparition des motifs dans le corpus (en s'appuyant sur les données du LASLA) et de la fréquence des collocations des amorces et de la cible en dehors du motif :
 - chaque amorce fortement reliée individuellement à la cible
 - chaque amorce faiblement reliée individuellement à la cible
 - la première fortement et la deuxième faiblement
 - la première faiblement et la deuxième fortement
- 144 triplet différents présentés à chaque participants

7. La question des fondements cognitifs

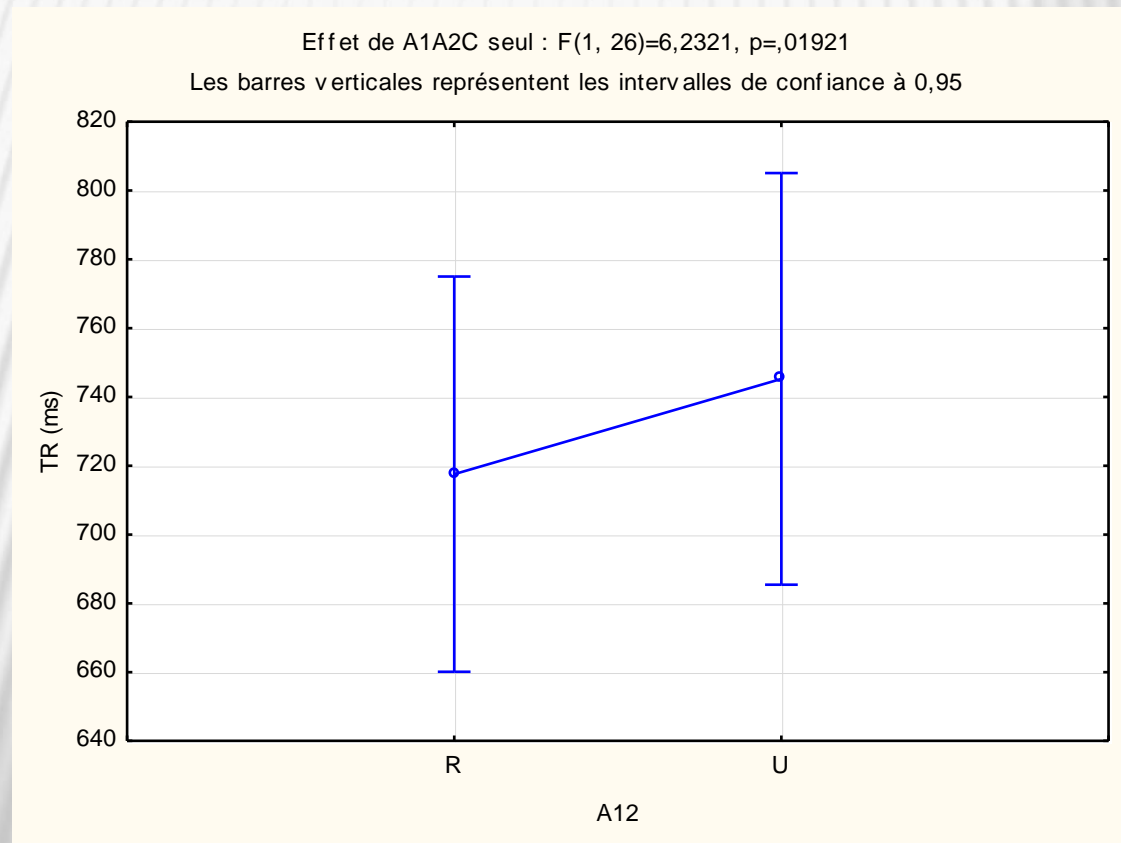
- Pratiquement :
 - - 144 paires de 2 mots reliés
 - ❖ 72 suivies par un pseudo-mot (chaîne de caractères) “cible”
 - 36 avec une cible “non reliée” – phonétiquement éloignée du mot attendu
ex. Auribus teneo – pausit
 - 36 avec une cible “reliée” – phonétiquement proche du mot attendu
ex quod erat – degestrandum
 - ❖ 72 suivies par un “vrai” mot latin
 - 36 motifs textuels : ex. Quibus rebus
 - 36 unités phraséologiques : ex. excessit e
- présentés avec des cibles reliées ou non reliées
- 36 motifs textuels : ex. Quibus rebus cognitis/recitatis
 - 36 unités phraséologiques : ex. excessit e uita/superiore

7. La question des fondements cognitifs

- si A1XC est fréquent, on obtient un temps de réponse court même si A1A2C n'est pas fréquent.
- quand A2C apparaît essentiellement dans le triplet A1A2C, les temps de réponse sont plus courts que quand A2C apparaît également ailleurs
- dans tous les cas, les temps de réponse sont d'autant plus courts que A1A2C est fréquent (717 ms contre 755 ms).

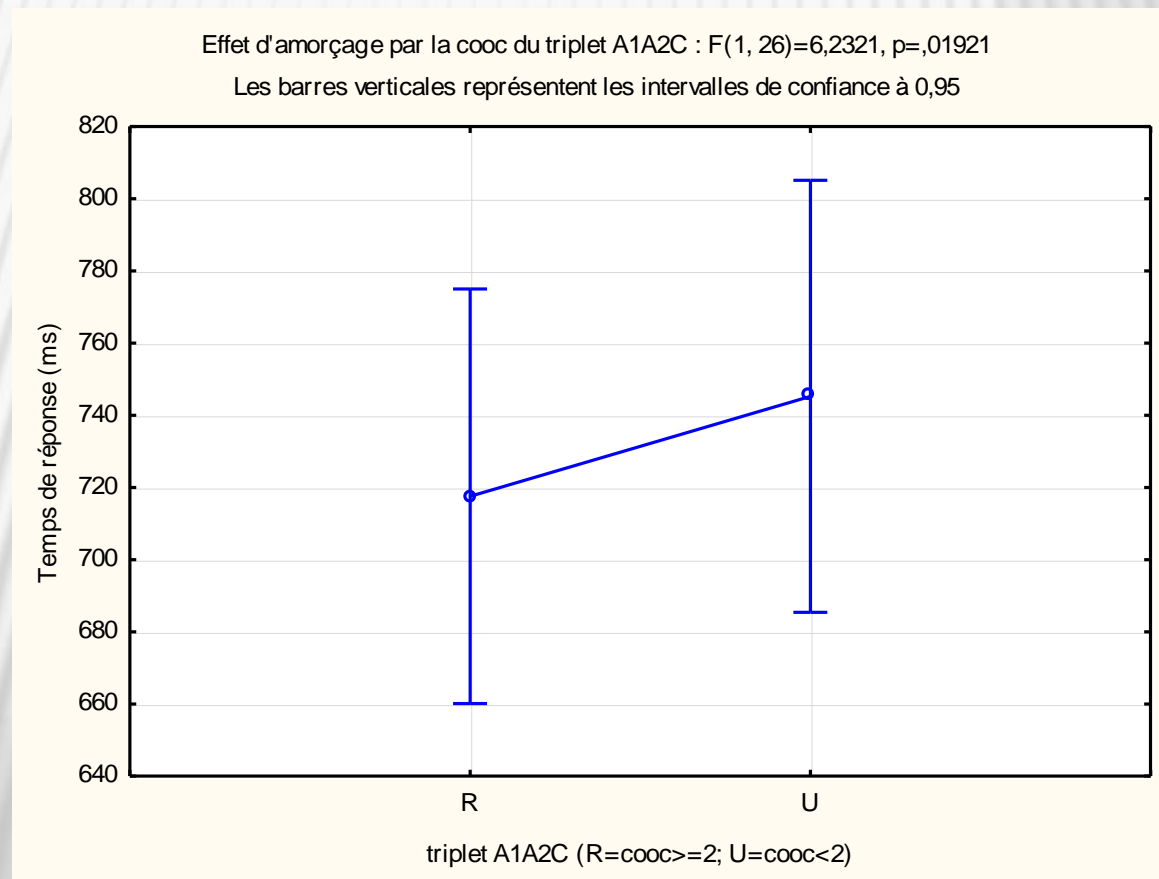
7. La question des fondements cognitifs

Temps de réponse pour une cible reliée (R) ou non reliée (U) dans un triplet A1A2C, quand A1XC et XA2C ne sont pas fréquents



7. La question des fondements cognitifs

Temps de réponse pour une cible reliée (R) ou non reliée (U) dans un triplet A1A2C, quand A1XC et XA2C sont fréquents



7. La question des fondements cognitifs

- L'effet relevé en peut pas être attribué au seul fait de la fréquence de la cible C.
- En effet, le nombre total des occurrences de mots cibles est plus élevé pour les triplets rares (en moyenne 7040 occurrences) que pour les triplets fréquents (en moyenne 1128 occurrences). Si la simple fréquence lexicale jouait un rôle, les temps de réponse devraient être plus courts pour des mots plus fréquents.
- Dans des « motifs » du type *Quibus rebus cognitis*, l'attente de *cognitis* est donc bien induite par la double amorce *Quibus rebus*.
- L'expérimentation démontre donc bien que la reconnaissance du « motif » repose sur une mémorisation liée à sa récurrence.

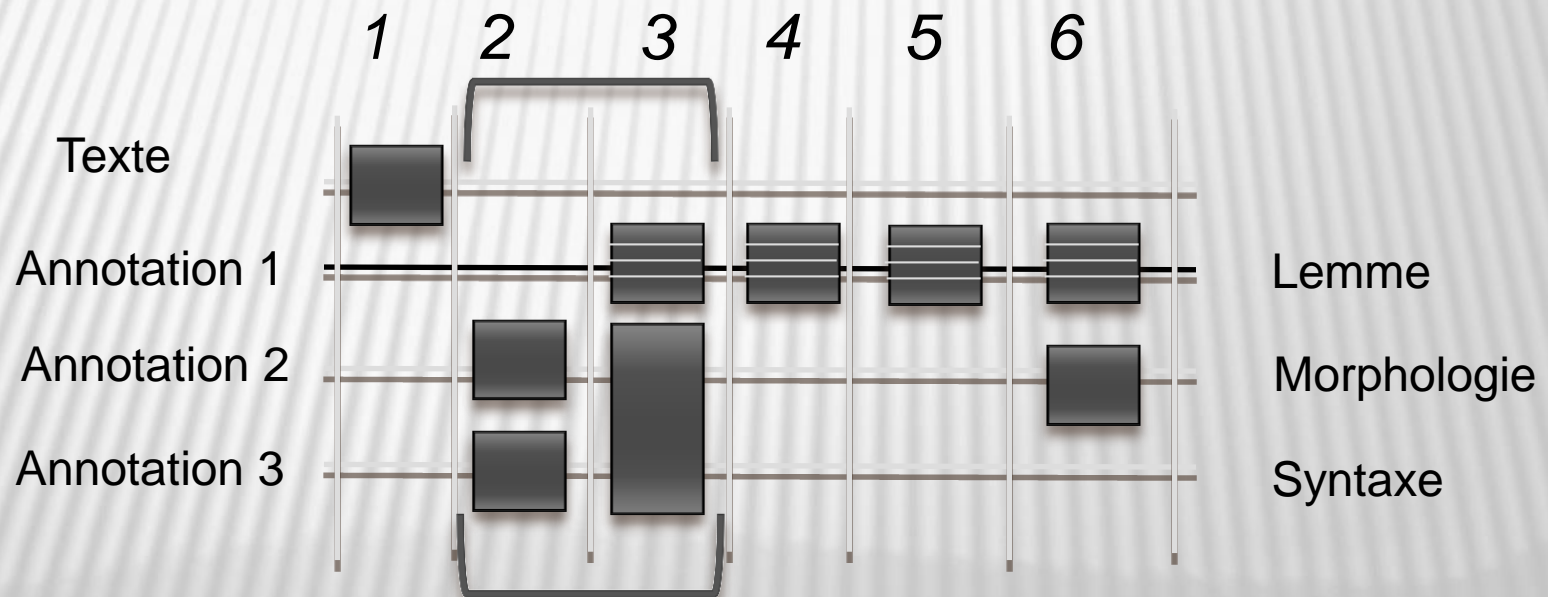
7. La question des fondements cognitifs

- On notera que les résultats de l'expérience sont en outre similaires pour les « motifs » et les unités phraséologiques,
- ce qui pose une série de questions :
 - les mécanismes de mémorisation sont-ils les mêmes pour les motifs et les unités phraséologiques ?
 - en quoi les résultats obtenus sont-ils liés à une expérimentation sur une langue seconde ? aurait-on des résultats similaires
 - pour une langue seconde vivante
 - pour une langue maternelle (problème des forces d'association) ?

8. L'extraction des motifs : non – semi – supervisée

- La difficulté liée à la multi-dimensionnalité

ici (1) j' (2) aimerais (3) toutefois(4) encore (5) préciser (6)



8. L'extraction des motifs

- L'existence de logiciel d'extraction semi-supervisée :
un exemple SDMC (Sequential Data Mining under Constraints)

Legallois D., Quiniou S., Cellier P. & Charnois Th. (2012) : « What About Sequential Data Mining Techniques to Identify Linguistic Patterns for Stylistics ? », in *Lecture Notes in Computer Science*, Springer.

Quiniou S., Cellier P., Charnois Th. & Legallois D. (2012) : « Fouille de données pour la stylistique : cas des motifs séquentiels émergents », in Anne Dister, Dominique Longrée et Gérald Purnelle, *JADT 2012, Actes des 11èmes Journées internationales d'analyse statistique des données textuelles*. Liège, pp. 821-833.

- Plusieurs difficultés à surmonter :
 - une multidimensionnalité plus importante que les simples parties du discours
 - la nécessité d'un tri entre les types de « motifs », significatifs ou non (le motif le plus fréquent n'est pas nécessairement le plus significatif !)

9. La distribution des « motifs » au fil du texte

- si le motif est structurant,
 1. il participe à la dynamique du texte;
 2. il interagit probablement avec d'autres marqueurs de la structuration textuelle
- d'où deux défis méthodologiques : rechercher des outils
 1. pour caractériser des distributions au fil des textes
 2. pour comparer les distributions et relever les complémentarités éventuelles
- une première solution : faire appel à la notion de « voisinage »
 - définir la taille d'une « fenêtre coulissante » = empan de texte successifs organisés autour d'un pivot
 - mesurer dans chaque empan la fréquence d'un phénomène donné

9. La distribution des « motifs » au fil du texte

Chaîne linéaire des codes temps-modes affectés à chacun des verbes principaux :

11 12 11 11 11 11 11 11 12 12 12 12 12 12 12 12 12 12 (11 11 11 11 11 15
15 12 15 15 14) 11 11 15 11 11 11 11 11 11 11 15 15 15 11 12 12 15 14 12
12 12 12 12 12 14 14 14 14 11 11 14 14 12 14 14 14 14 14 14 14 14
14 12 12 12 12 14 (14 14 14 14 14 14 14 14 14 14 14) 15 14 14 12 12 12
12 etc.

Topologie établie à partir des mesures de voisinage de chaque unité (taille du voisinage = 11, propriété du voisinage = nombre d'occurrences du code 14) :

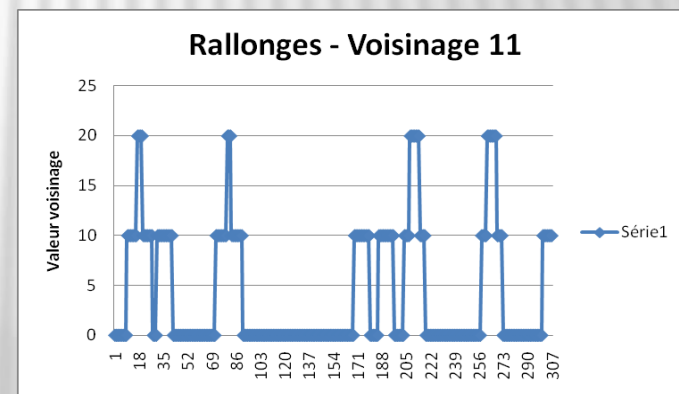
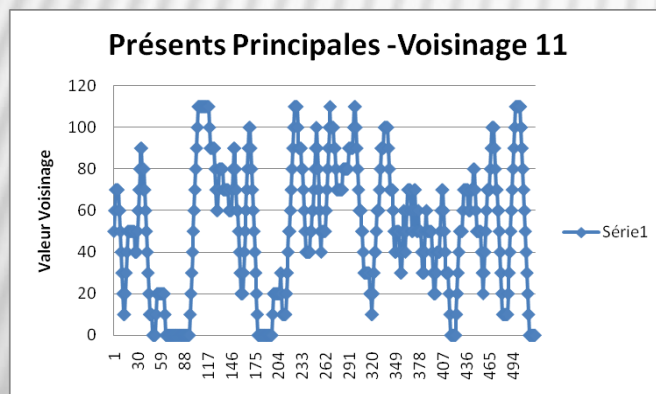
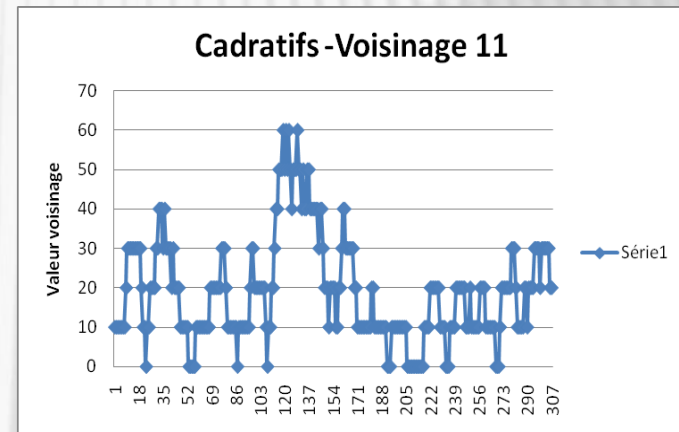
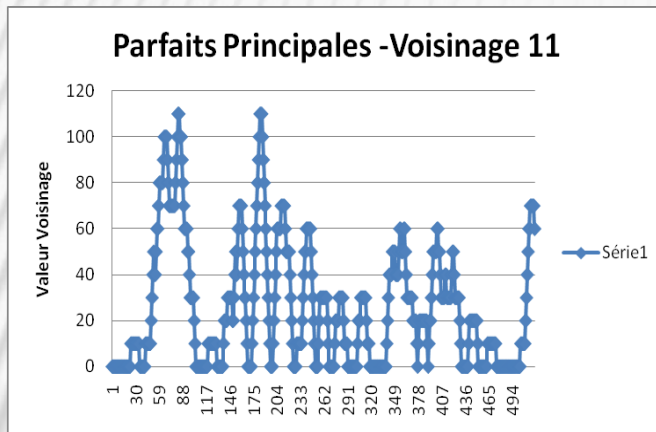
X X X X X 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1
1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 1 1 1 1 1 2 3 4 5
4 4 5 6 6 7 8 8 8 8 8 9 10 10 10 10 9 8 7 7 7 7 7 7 7
8 9 10 11 11 10 10 10 9 8 7 6 6 6 6 etc.

9. La distribution des « motifs » au fil du texte

- la mesure de voisinage : un instrument souple permettant
 - d'étudier des objets variables et complexes
 - d'établir des seuils de fréquence

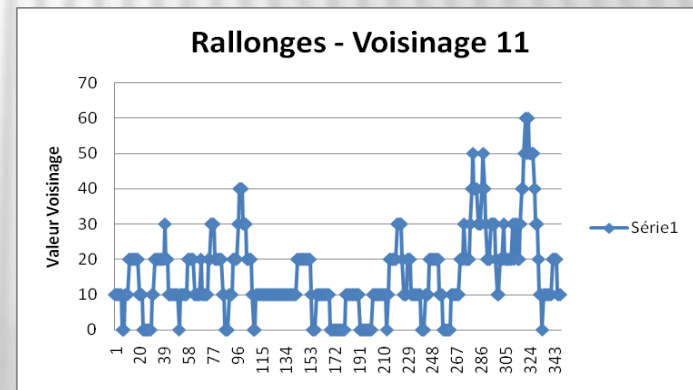
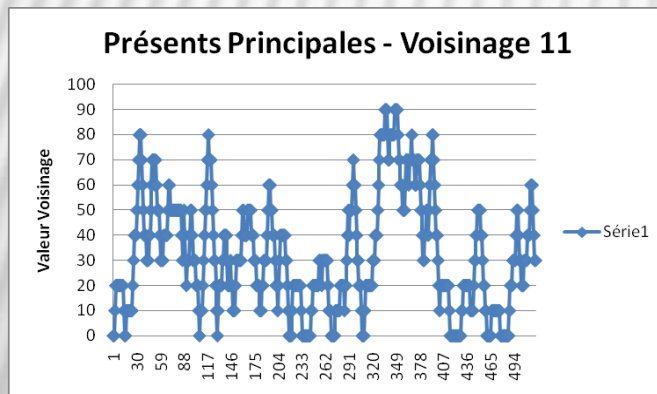
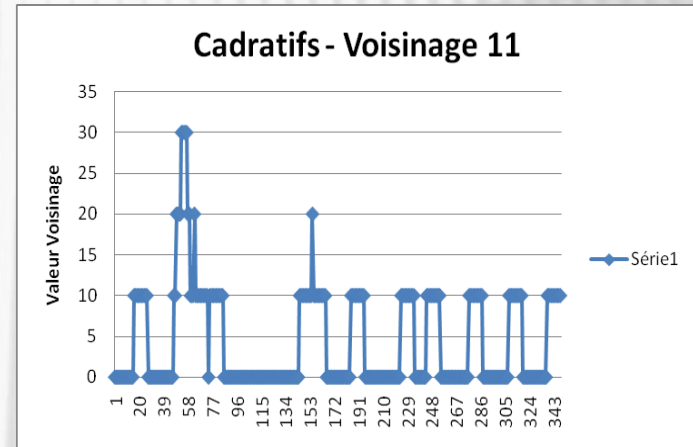
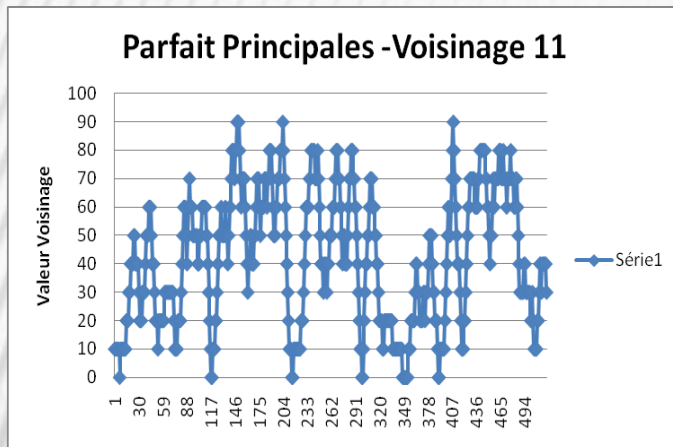
- avec une combinaison de divers paramètres simples ou complexes
 - > d'où possibilité de combiner avec le motif

9. La distribution des « motifs » au fil du texte



Distribution des 4 indicateurs textuels en *Civ2*

9. La distribution des « motifs » au fin du texte



Distribution des 4 indicateurs textuels en Ann. 12

9. La distribution des « motifs » au fil du texte

- acquis :

- une vision qualitative de la dynamique et de la structuration du texte
- une aide visuelle pour déterminer
 - les interactions entre motifs
 - un partitionnement du texte

Longrée D. & Mellet S. (à paraître) : "A Text Structure Indicator and two Topological Methods: New Ways for Studying Latin Historic Narratives"

- limites :

- pas d'outils pour une évaluation quantitative du poids dans la structuration du texte
 - pour chaque motif
 - pour leur combinaison
- une méthode effective seulement pour les phénomènes suffisamment représentés
- pour des phénomènes plus rares : utilisation de la méthode dite des "rafales"

9. La distribution des « motifs » au fil du texte

- la méthode des rafales de Lafon (1981) :
vise à fournir une évaluation statistique fondée sur 'la configuration dessinée par les occurrences d'une forme dans la séquence ordonnée d'un texte ', en termes de régularité ou d'irrégularité de la distribution ;
- exemple : si un texte = 1000 mots et occurrences d'un mot $M = 10$
 - distribution d' 1 occurrence de M tous les 100 mots :
particulièrement régulière
 - 8 occurrences entre les 20^e and 30^e mots du texte :
particulièrement irrégulière ("*en rafales*")
- une formule mathématique pour évaluer une régularité/irrégularité avec un indice (X) qui rend compte de la distribution dans un texte

9. La distribution des « motifs » au fil du texte

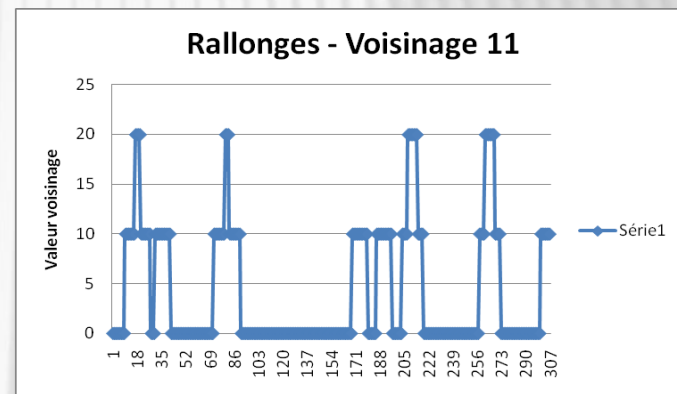
Application aux “rallonges” dans *Civ2*

T (nombre de phrases) = 307

f (nombre de rallonges) = 12

$X = 1663$

$\text{Exp}(X) = 556$



-interprétation :

- configuration claire en “rafales”
- significatives comme le montre le retour au texte : correspondent toujours à des moments où l’armée de César est en difficulté et où l’action militaire piétine (mise en place du siège d’une ville, installation d’un camp sur un terrain défavorable)

10. Des perspectives pour les 10 années à venir ?

De nouvelles recherches

- sur les fondements cognitifs du « motif »:
 - langues premières et langues secondes
 - au-delà du « triplet » ?
 - unités phraséologiques et motifs textuels
- sur des méthodes et des outils d'analyse:
 - logiciels de détection prenant mieux en compte la multidimensionnalité
 - logiciel permettant des tris et classements automatiques
- sur la distribution des motifs textuels au fil du texte et sur la complémentarité des distributions,
 - en utilisant d'autres outils statistiques
 - en s'intéressant à d'autres corpus (littéraires – politiques)
- un espoir : des instruments de caractérisation et de classification plus fins