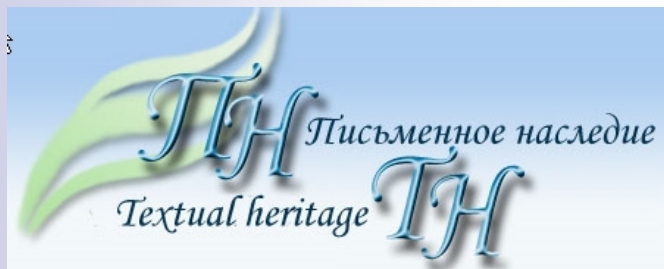




)))(((icar

A gentle introduction to Textometry & TXM

Serge Heiden &
Alexei Lavrentiev
(Lyon University, Франция)
textometrie@ens-lyon.fr



Petrozavodsk workshop - <http://textualheritage.org>, September 4-5, 2012



Textual corpora analysis with Textometry tools & applications

- Qualitative tools

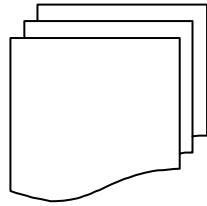
- Full text patterns search, Kwic concordances, Text edition browsing, NLP...

- Quantitative tools

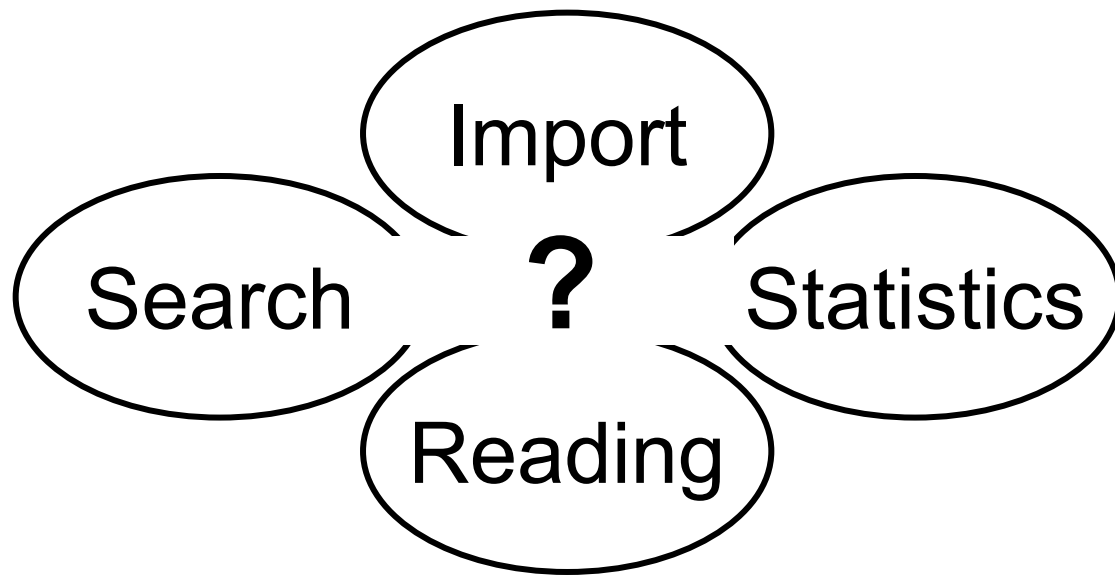
- Frequency lists, Specificity analysis, Factorial analysis, Classification, Collocates, N-grams...

- Applications : discourse analysis (80's), linguistics, literary studies, geography, philosophy, history, etc.

Computerized methodology



Corpus format = TXT < XML < XML-TEI



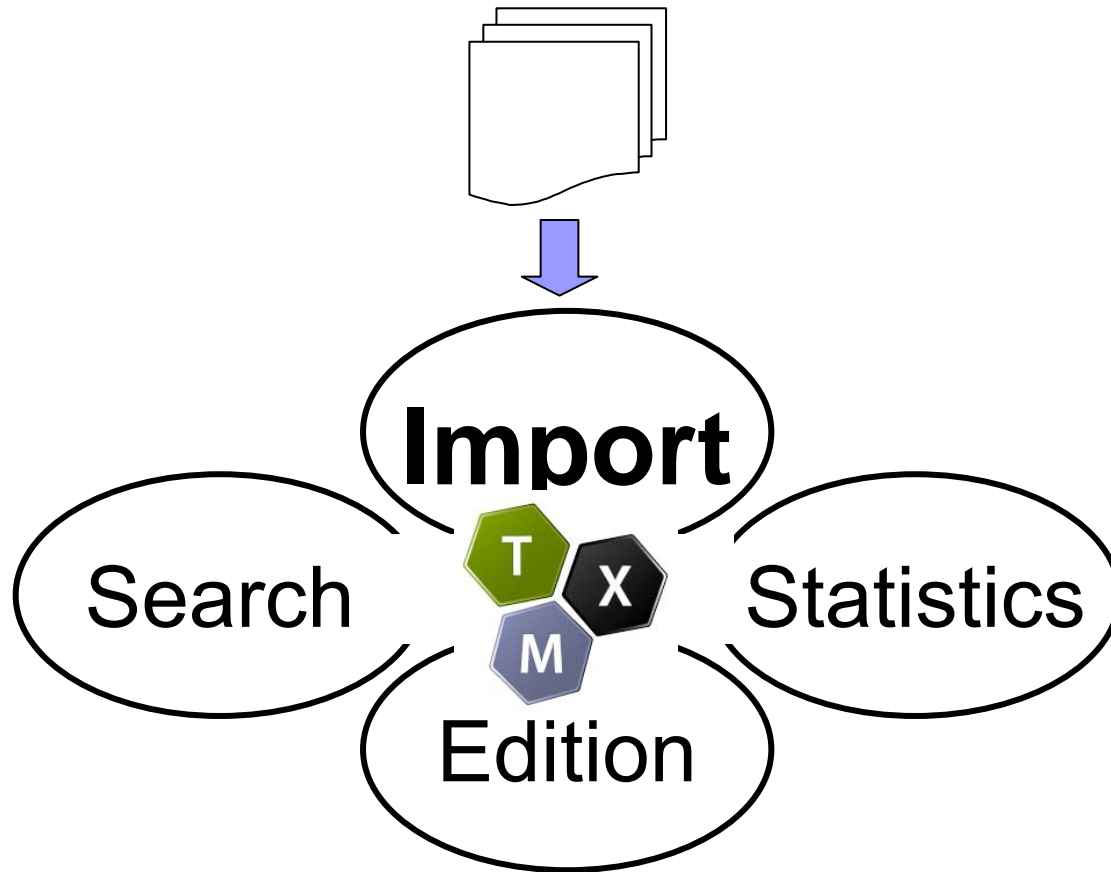
**Software
to
interact
with
any texts**



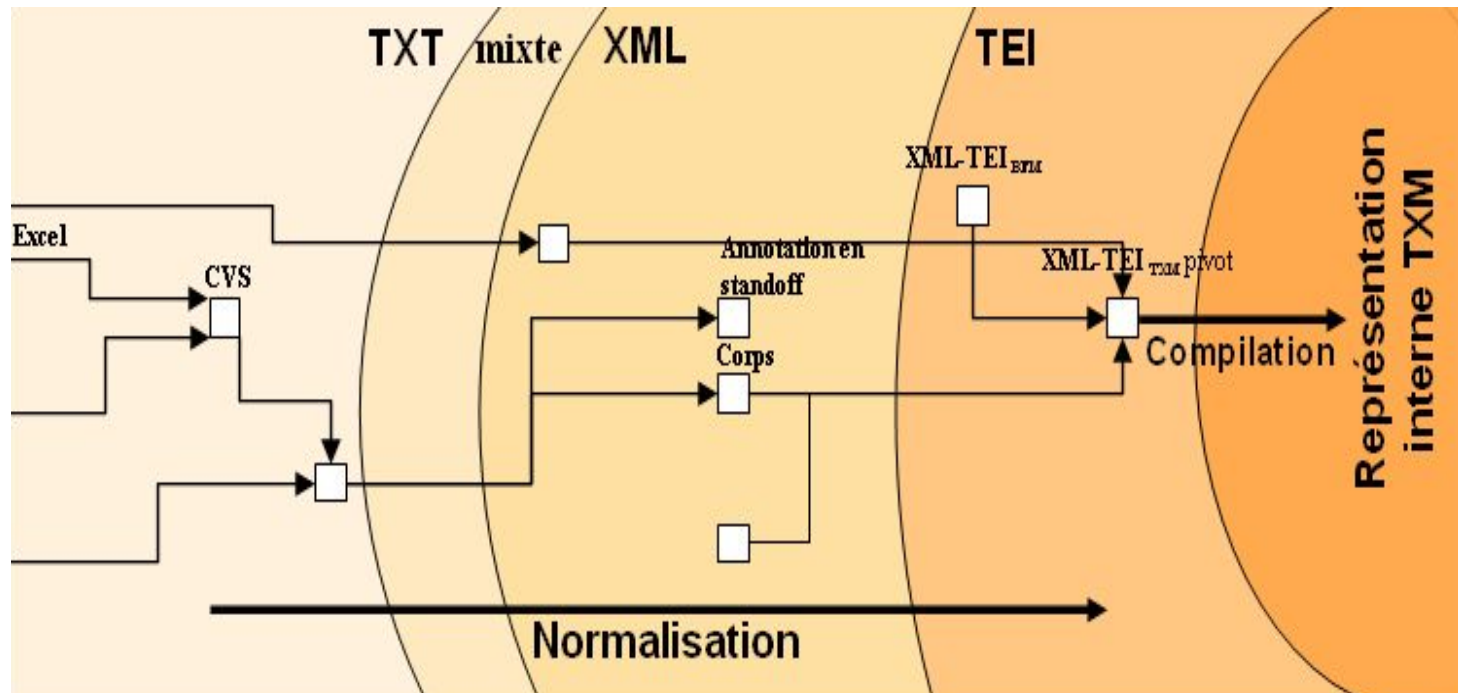
platform - <http://sf.net/projects/txm>

- 2007-2010 **kick-off** French Research Agency grant
 - 6 partner laboratories (4 FR, 1 UK, 1 CA)
- **Standard software** architecture (Java+OSGi+J2EE)
 - Key open-source components : R (statistics) & CQP (search engine)
- **Standard corpus formats :**
 - Unicode (TXT)
 - XML
 - XML-TEI compatible
- **Free & Open-source** GPL software
 - **TXM Windows**, Mac & Linux desktop application
 - **TXM Web Portal**: online access (membership, control)
- Efficient : 1 billion words

TXM import

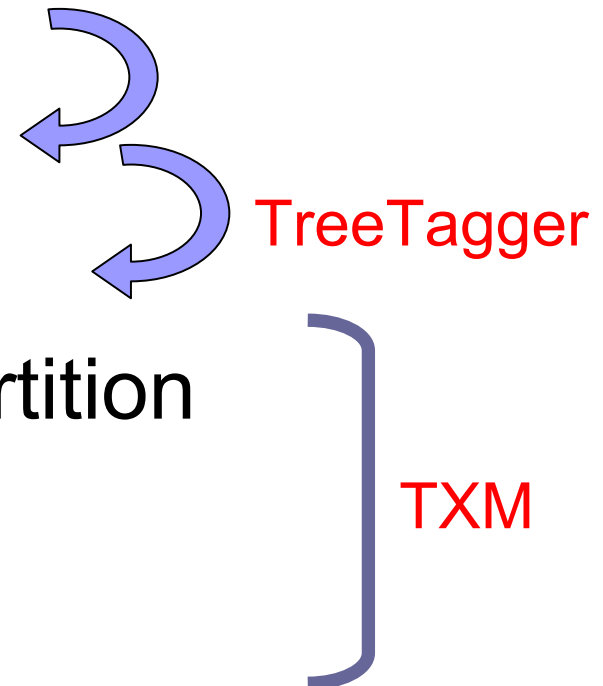


TXM import environment



Basic import & analysis workflow

- TXT texts directory
- XML texts + metadata
- Tagged texts
- Contrasts : sub-corpus & partition
- Contexts
- Lexical facets



TXM corpus data model

- Text Units (book, article...)
 - Metadata (author, date, domain, genre...)
 - Textual Planes
 - Comments, Direct Speech, Speaker Turns
 - Main language (french...), Secondary language (latin...)
 - Out-of-Text (comments...)
 - Internal Structure (sentence, paragraph, sections...)
- Lexical Units (graphical form, lemma, part of speech...)
- NLP tools involved (taggers...)
- Reading
 - Pagination (page breaks)
 - Rendering properties (styles)
- Bibliographic references
- Alignment (aligned corpora)

TXM import charter

TXM import levels charter

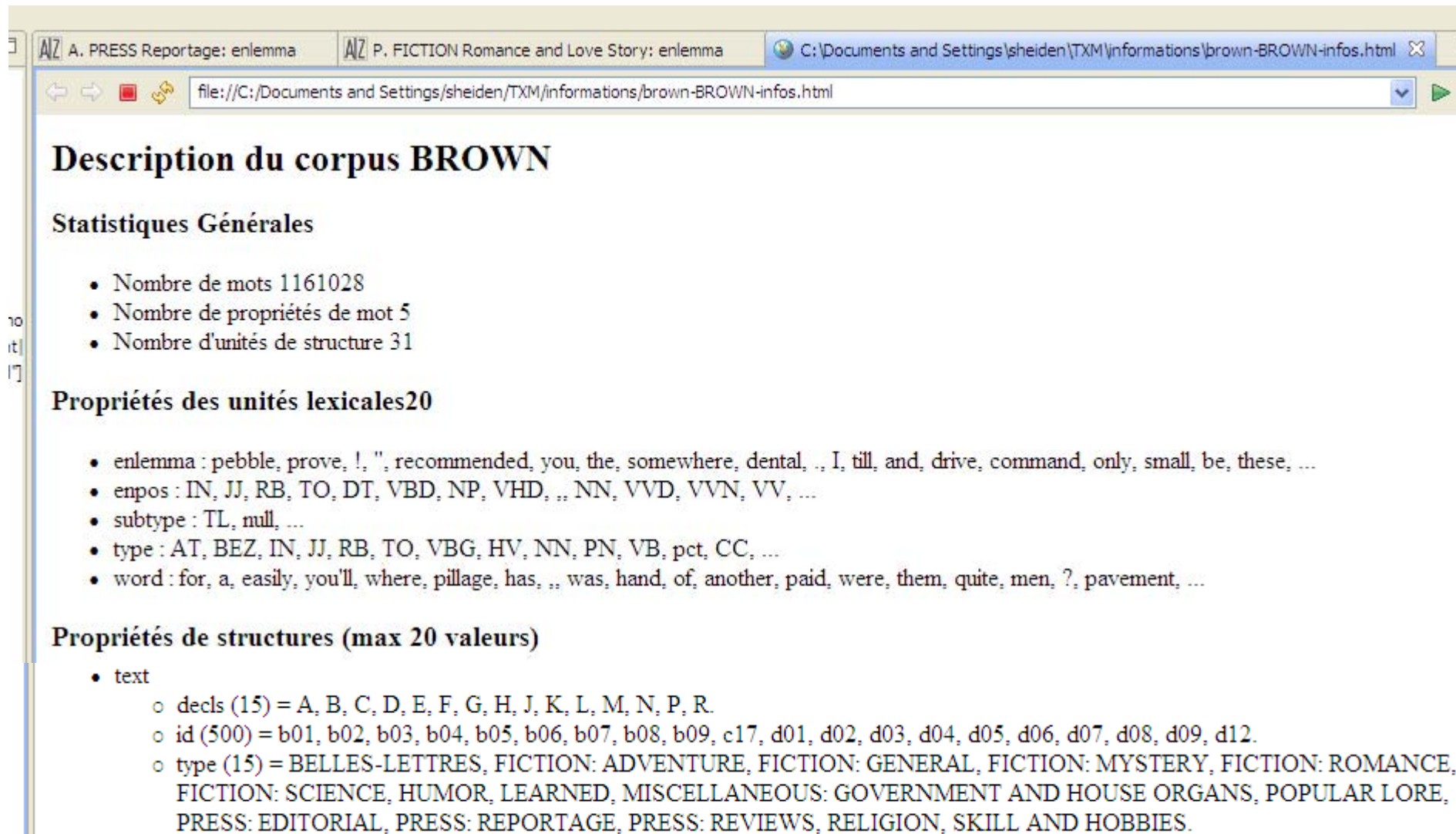
	TXT	XML/w	XML-TEI
<i>Text units</i>	files	files	files
<i>Metadata</i>	CSV	CSV	teiHeader
<i>Words</i>	raw	<w>?	<w>?
<i>Structures</i>	-	any	specific
<i>Planes</i>	-	front XSL	specific

TXM import modules

corpus input formats

- Various proprietary formats : Hyperbase, Alceste, CNR (Cordial)
- *Calibre* – open ebook digital library (ePub)
- Copy/Paste
- TXT Unicode+CSV (metadata) : raw texts directory
- XML/w+CSV : XML texts directory
- XML-TEI P5 **BFM** : TEI standard compatible XML
- XML-TEI P5 **BVH**
- XML-TEI P5 **FRANTEXT texts**
- XML-TEI P5 **FRANTEXT search results**
- **XML-TEI-TXM** : TEI compatible XML+NLP (**pivot**)
- XML-Transcriber+CSV – audio aligned transcriptions
- *XML-TMX* – multilingual aligned corpora
- *XML-PPS-Factiva* – press portal

Brown corpus: W. N. Francis & H. Kucera, Brown University, 1964 – XML-TEI NLTK version



Description du corpus BROWN

Statistiques Générales

- Nombre de mots 1161028
- Nombre de propriétés de mot 5
- Nombre d'unités de structure 31

Propriétés des unités lexicales

- enlemma : pebble, prove, !, ", recommended, you, the, somewhere, dental, ., I, till, and, drive, command, only, small, be, these, ...
- enpos : IN, JJ, RB, TO, DT, VBD, NP, VHD, ,, NN, VVD, VVN, VV, ...
- subtype : TL, null, ...
- type : AT, BEZ, IN, JJ, RB, TO, VBG, HV, NN, PN, VB, pct, CC, ...
- word : for, a, easily, you'll, where, pillage, has, ,, was, hand, of, another, paid, were, them, quite, men, ?, pavement, ...

Propriétés de structures (max 20 valeurs)

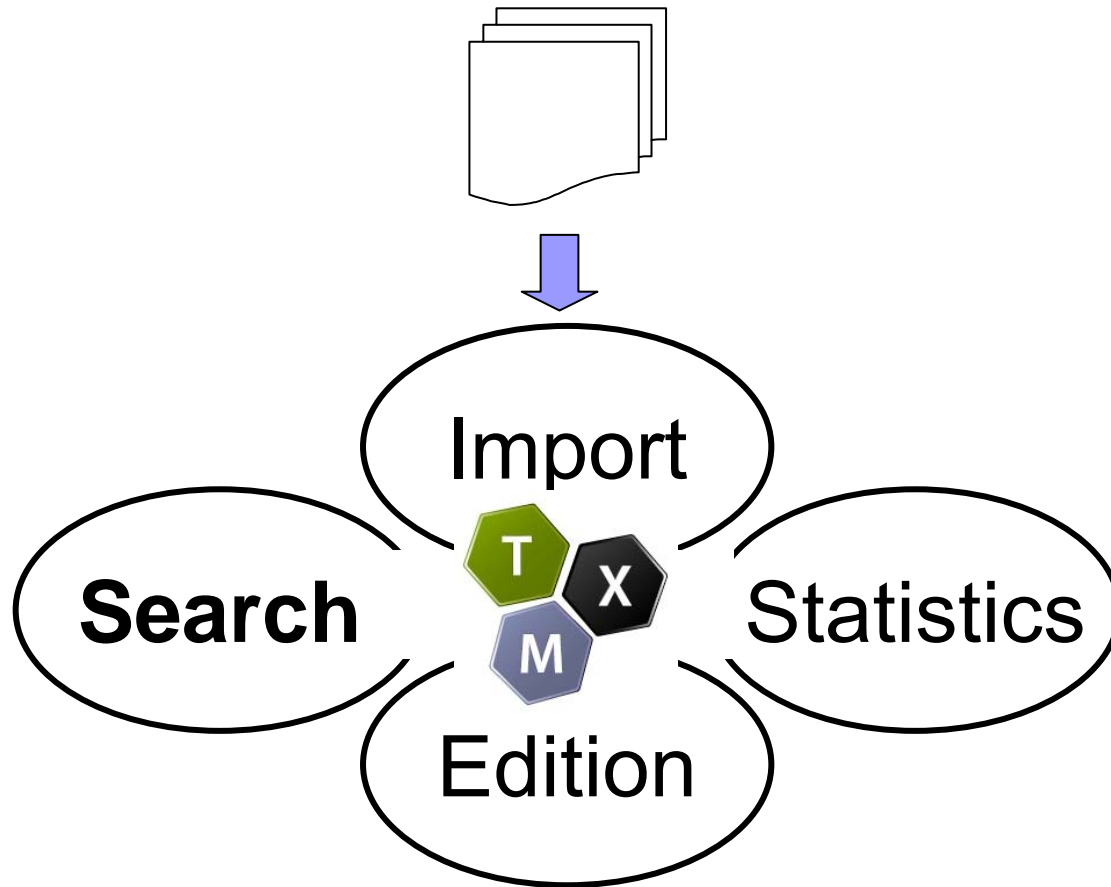
- text
 - decls (15) = A, B, C, D, E, F, G, H, J, K, L, M, N, P, R.
 - id (500) = b01, b02, b03, b04, b05, b06, b07, b08, b09, c17, d01, d02, d03, d04, d05, d06, d07, d08, d09, d12.
 - type (15) = BELLES-LETTRES, FICTION: ADVENTURE, FICTION: GENERAL, FICTION: MYSTERY, FICTION: ROMANCE, FICTION: SCIENCE, HUMOR, LEARNED, MISCELLANEOUS: GOVERNMENT AND HOUSE ORGANS, POPULAR LORE, PRESS: EDITORIAL, PRESS: REPORTAGE, PRESS: REVIEWS, RELIGION, SKILL AND HOBBIES.

Brown – tagset

http://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP\$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative

TXM search





TXM Search tools

- Selecting **Focus** = Constructing **Events**
to count for
for example the word « woman »

Versus

- **Unfocused** or **Panoramic** tool : every
event participates
*for example all the vocabulary of a text (all
words)*

TXM Search tools – focus: CQL query language



- Simple word form :
« **woman** »
- Pattern of word form :
« **wom.n** »
- Word property :
« **[enpos="NN.*"]** »
- Combination of word properties :
« **[enlemma="woman" & enpos="NNS"]** »
- Sequence of words :
« **[enpos="NN.*"] [enpos="VBD"]** »
- Variable Sequence of words :
« **[enpos="NN.*"] []* [enpos="VBD"]** »
- Context constraints : « **[enpos="NN.*"] []* [enpos="VBD"] within s** »

– “.” = any character

– “[]” = any word

– “*” = 0 to N times (character or word)

TXM Search tools – focused Index of Nouns lemma in Press & Fiction

A. PRESS Reportage: [enpos="N,*"]:enlemma

Query: [enpos="N,*"] Properties:

Thresholds: Fmin: 1 Fmax: 9999999 Vmax: 9999999

1-100 / 7954 t 31494

enlemma	Frequency T=100541
Mrs.	253
year	244
Mr.	170
home	134
man	119
state	115
time	115
week	112
member	109
school	102
day	101
New	88
President	88
program	78
government	74
month	74
game	73
car	68
House	68

P. FICTION Romance and Love Story: [enpos="N,*"]...

Query: [enpos="N,*"] Properties:

Thresholds: Fmin: 1 Fmax: 9999999 Vmax: 9999

1-100 / 4024 t 135

enlemma	Frequency T=70012
man	121
time	107
day	96
eye	87
way	87
thing	83
I'm	77
something	67
hand	65
night	58
nothing	56
woman	56
life	53
boy	51
room	51
head	50
people	50
place	49
year	49

TXM Search tools – focused Concordance of lemma “woman”

P. FICTION Romance and Love Story: [enpos="N.*"]:enlemma

P. FICTION Romance and Love Story:[enlemma="woman"]


Query: [enlemma="woman"] Keyword: word Edit

sort keys: #1 None #2 None #3 None #4 None Sort

1-56 / 56

text_id	Left context	Keyword	Right context
p01	. `` But she shouldn't have come here in the first place ", the	women	had said. `` No, no. Not that one. She thought she
p01	at one time he had looked down the street at the fleshy suppleness of a	woman	he had consumed - watching her become thinner and thinner
p01	voice had not commanded her to stop . It had not questioned why . The	women	said they had seen him wave an exhausted farewell;; but he
p01	him and the children , had he guessed that all he would remember of his	woman	was the memory of her climbing into that streetcar?? There s
p02	You might set up a dress shop in the living room " . `` Every	woman	in the block has tried that " . `` What about a tea room,
p03	his lips : `` Under the window in stormy weather I marry this man and	woman	together. Let none but Him who rules the thunder Put this man
p03	woman together . Let none but Him who rules the thunder Put this man and	woman	asunder " . Absolution for his lie?? He questioned God's taking
p04	enough of that ; ; Doaty would never have left her house to a godless	woman	. She found herself wishing an old wish, that she had told Do
p04	so well's I remember you " , he said . `` Y're the young Blackwell	woman	. Ran away on a black night with a lawful wedded man. I kno
p04	only a kind of shadow . She started to move away , just as a	woman	came out of the cottage, a big-boned, drab-haired figure with
p04	" ? ? Henrietta nodded . How much they knew about her !! The	woman	(she must have been a tiny baby when Hetty and Delia had s
p05	and blue eyes ; ; all their men were bearded , and many of their	women	were beautiful. A pitiful few of them are left now, to subsist r
p05	and - back in the blessed United States - they were regarded simply as pretty	women	. Now, having been sent halfway around the world on a job s
p06	saw me in the lobby . He was disappointed to find a nervous , scrawny	woman	with a big hat standing at the door. She frowned at his green
p06	She said , semi-professionally . `` Our church is sponsoring a group of very courageous	women	up in Alaska. We call them lay-sisters and they go among the
p06	understand the difficulties and I want to help these lay-sisters . Do you know these	women	go all through Alaska, and they don't have the proper facilities
p06	stomach . The Golden Calf was dimly lit with shaded neon . There were more	women	than men in the place, but he couldn't find a flowerpot. They a
p06	settlers in covered wagons , who appeared much more animated than the gamblers . The	women	had a bright shining expectancy as they leaned out from the v
p08	country , in the kibbutzim . She loved the children . She was a strange	woman	, your mother. When she loved, it was with a passion that dro
p08	danger as much as she sought out helping other people . She was most strange	woman	. Ready to follow her impulse. It was an impulse when she w
p09	sees , and at night you feel when he is asleep or with the other	woman	the one that never could love him the way you do. the one

TXM Search tools – focused Concordance sorted+ref.

Query:  [enlemma="woman"] Keyword: word Edit Search

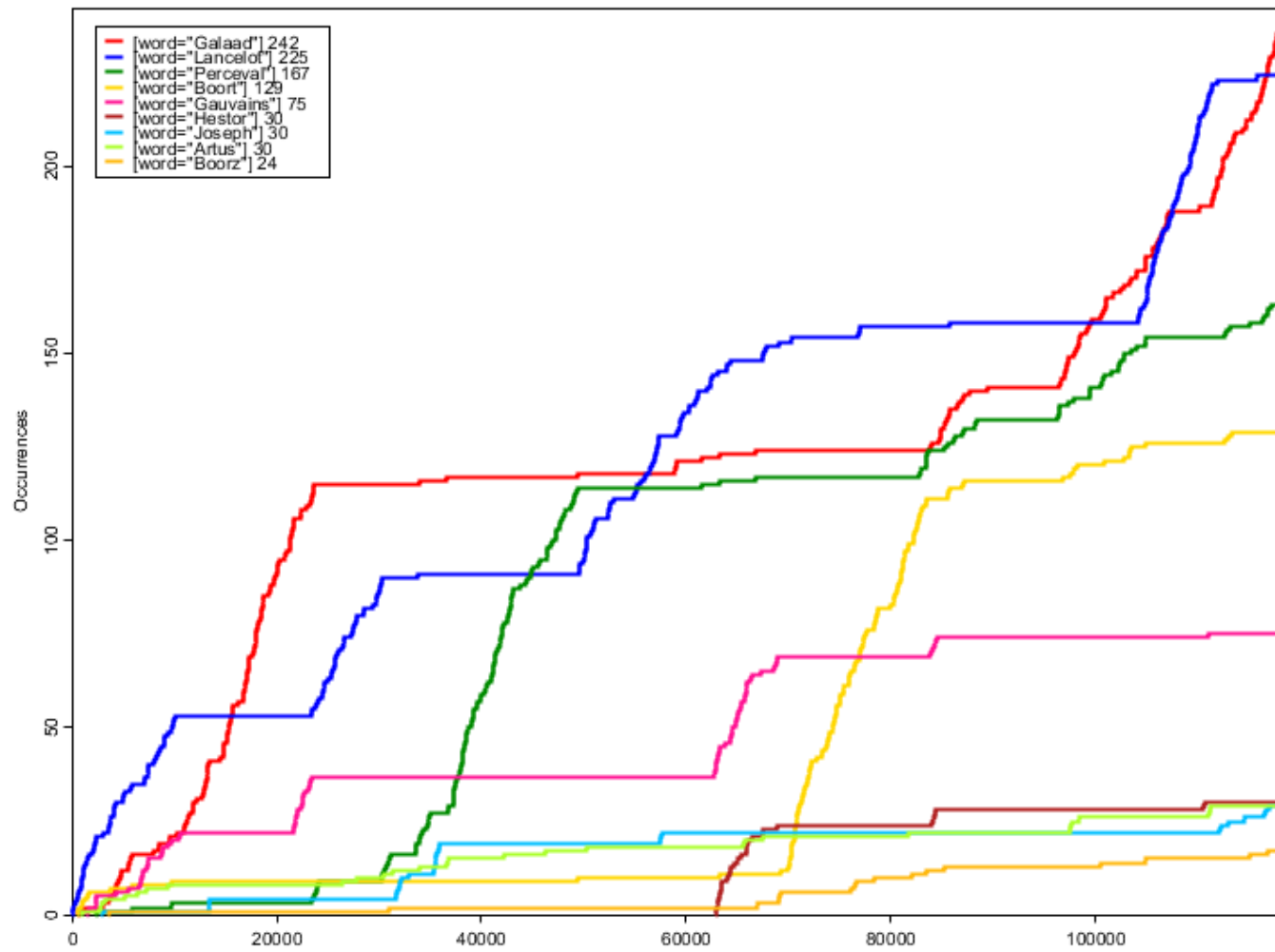
sort keys: #1 Right context #2 None #3 None #4 None Sort

1-56 / 56

text_id	Left context	Keyword	Right context
p19	walked in , the whole miserable feeling of the day seemed to focus on the	woman	in the bed. They'd cleaned her up some, of course, and she'd
p02	You might set up a dress shop in the living room ". `` Every	woman	in the block has tried that ". `` What about a tea room,
p11	But suppose she had not taken Mrs. Mathias' advice and lived on like thousands of	women	in towns, dispossessed of love, hanging on to makeshifts, and alto
p22	love for Johnnie was young and clean - how could I possibly compete with a	woman	like that, who didn't hesitate to use her sex. Johnnie was a trucker
p13	the end of the camp . Elsewhere others heard and stopped and waited , the	women	peering from their lodges then gathering in small, curious clusters. E
p11	us . `` She's nowhere , nowhere " !! She screamed , and both	women	ran up to the house, and I followed. The search began, in
p12	explained that he had no experience with boxcars . Just the same , the old	woman	said, she would write to her nephew in his boxcar and tell him she
p13	Go take helsq'iyokom , your evil spirit , to the young boys " , the	woman	said. `` They do not have to face battle. I will not let
p01	voice had not commanded her to stop . It had not questioned why . The	women	said they had seen him wave an exhausted farewell;; but he might
p13	like cavalry . At times they would ride frenziedly through the camp , letting the	women	see their courage, how handsome they were in their regalia. Then e
p06	stomach . The Golden Calf was dimly lit with shaded neon . There were more	women	than men in the place, but he couldn't find a flowerpot. They all
p12	for Carla to have those old prophets gripping her imagination now ? ? Being a	woman	though, she would take only what she needed from church. It was
p03	his lips : `` Under the window in stormy weather I marry this man and	woman	together. Let none but Him who rules the thunder Put this man and v
p11	paid attention to that half-murmured remark , but it seemed one of those extreme statements	women	under stress indulge in. I love you, I hate you, I feel
p06	She said , semi-professionally . `` Our church is sponsoring a group of very courageous	women	up in Alaska. We call them lay-sisters and they go among the Eskimi
p01	him and the children , had he guessed that all he would remember of his	woman	was the memory of her climbing into that streetcar?? There seemed
p20	in a fresh uniform . I felt like a Hun in Rome . All the	women	were beautiful, and the men were equal to them;; everything was g
p05	and blue eyes ; ; all their men were bearded , and many of their	women	were beautiful. A pitiful few of them are left now, to subsist mainly
p12	said solemnly , trying to get Carla's eye . Surely she could see that these	women	were her Italians, too, he thought. Devout, orthodox and plain like
p17	that summer of 1936 when she was by all odds the prettiest and brightest young	woman	west of the Allegheny Mountains, and John was surely one of the t
p22	it . It wasn't Johnnie's fault that he was hopelessly tied down to that frightful	woman	who did her best to make his life unbearable. Just because he was

TXM Search tools – focused progression graph

Progression graph of in corpus QGRAAL



TXM Search tools – unfocused vocabulary Press & Fiction r.

The image displays two side-by-side screenshots of the TXM software interface, showing search results for the term 'enlemma' in two different corpora.

Left Window: A. PRESS Reportage: enlemma

Properties: enlemma

Thresholds: Fmin: 1 Fmax: 9999999 Vmax: 9999999 Page size: 1

1-100 / 11120 t 100541 , v 11120

enlemma	Frequency T=100541
the	6386
,	5188
.	4027
of	2861
be	2838
and	2220
a	2124
to	2116
in	2020
@card@	1152
for	969
have	861
that	829
''	732
"	700
on	691
he	642
at	636
with	567
as	517
by	504
it	478
say	465
his	428
will	389

Right Window: P. FICTION Romance and Love Story: enlemma

Properties: enlemma

Thresholds: Fmin: 1 Fmax: 9999999 Vmax: 9999999 Page size: 1

1-100 / 6672 t 70012 , v 6672

enlemma	Frequency T=70012
,	3899
.	3736
the	2988
be	1987
and	1907
to	1502
a	1383
of	1202
he	1068
''	1045
"	1036
have	1009
I	951
in	930
she	728
it	717
?	690
her	680
that	612
his	603
you	558
with	473
for	441
at	431
say	417

TXM paradigmatic contrast:

Partition building: text@domain

The screenshot shows the 'Create Partition' dialog box in the TXM software. The window has a blue title bar with the TXM logo and the text 'Create Partition'. Inside, there is a 'Name:' field containing 'domains'. Below this are three tabs: 'Simple', 'Assisted' (which is selected), and 'Advanced'. Under the 'Assisted' tab, there are two dropdown menus: 'Structure:' set to 'text' and 'Property:' set to 'deds'. Below these is a list box labeled 'Select values to assign:' containing the letters A through K. At the bottom left, there are two buttons: 'New part' and 'Rmv all the parts'. Below these is a section for a specific part, with a 'Title:' field containing 'Part 1', a close button (X), and 'Assign' and 'Remove' buttons. At the bottom right of the dialog are 'OK' and 'Cancel' buttons.

Create Partition

Name: domains

Simple Assisted Advanced

Structure: text Property: deds

Select values to assign:

- A
- B
- C
- D
- E
- F
- G
- H
- I
- J
- K

New part Rmv all the parts

Title: Part 1

Assign Remove

OK Cancel

TXM paradigmatic contrast: verbs lemma Index on text@domain

Query: [enpos="v.*"] Properties: enlemma Edit

Thresholds: Fmin: 1 Fmax: 9999999 Vmax: 9999999 Page size: 100

1-100 / 4765 t 164214 , v 4765 , fmin 1 , fmax 9999999

enlemma	Frequency T=1161028	J t=181872	G t=173074	F t=110288	A t=100541	E t=82321	H t=70104	P t=70012	N t=69328	K t=68483	B t=61592	L t=58483
be	37334	7053	5863	3681	2838	2553	2207	1987	1669	1996	2106	1848
have	11793	1350	1908	1114	861	714	431	1009	755	937	687	584
do	3370	382	526	289	185	239	95	310	228	270	233	199
say	2779	115	300	156	465	32	33	417	332	251	106	88
make	2322	314	315	231	184	206	196	134	145	129	120	100
go	1794	63	211	143	113	68	29	254	210	228	79	66
take	1571	141	217	151	150	102	65	146	126	124	76	64
come	1558	100	237	147	100	70	41	171	170	176	75	63
see	1511	151	189	126	68	83	38	160	186	159	61	50
get	1472	53	122	116	142	65	10	211	200	160	59	48
know	1462	89	223	126	59	57	33	181	133	156	71	58
give	1260	219	205	110	103	71	57	88	64	74	55	45
find	1032	186	175	103	50	80	34	66	58	71	41	33
use	1015	281	90	141	58	173	71	36	33	29	30	24
think	969	59	109	55	48	21	21	174	122	121	32	26
look	890	40	121	42	32	49	17	156	107	124	26	21
seem	829	134	138	78	36	27	8	64	58	73	50	40
become	766	148	160	108	49	56	21	16	24	34	42	33
tell	764	32	90	70	74	12	6	94	92	88	24	19
leave	692	74	85	63	42	43	6	74	57	63	43	33
feel	638	53	92	38	38	16	9	109	55	86	30	24
want	632	35	80	48	43	36	9	91	67	62	30	24
show	624	200	77	40	35	62	35	16	36	18	28	21
call	614	58	109	72	57	25	15	52	28	37	37	28

TXM paradigmatic contrast: verbs lemma Lexical Table on text@domain

C:\Documents and ... Dimensions: BROWN... BROWN_text_decls:... BROWN_text_decls:... BROWN_text_decls:...

t 1161028, v 4766, Fmin 1, Fmax 996814

Keep Number of lines: 31 Fmin: 509

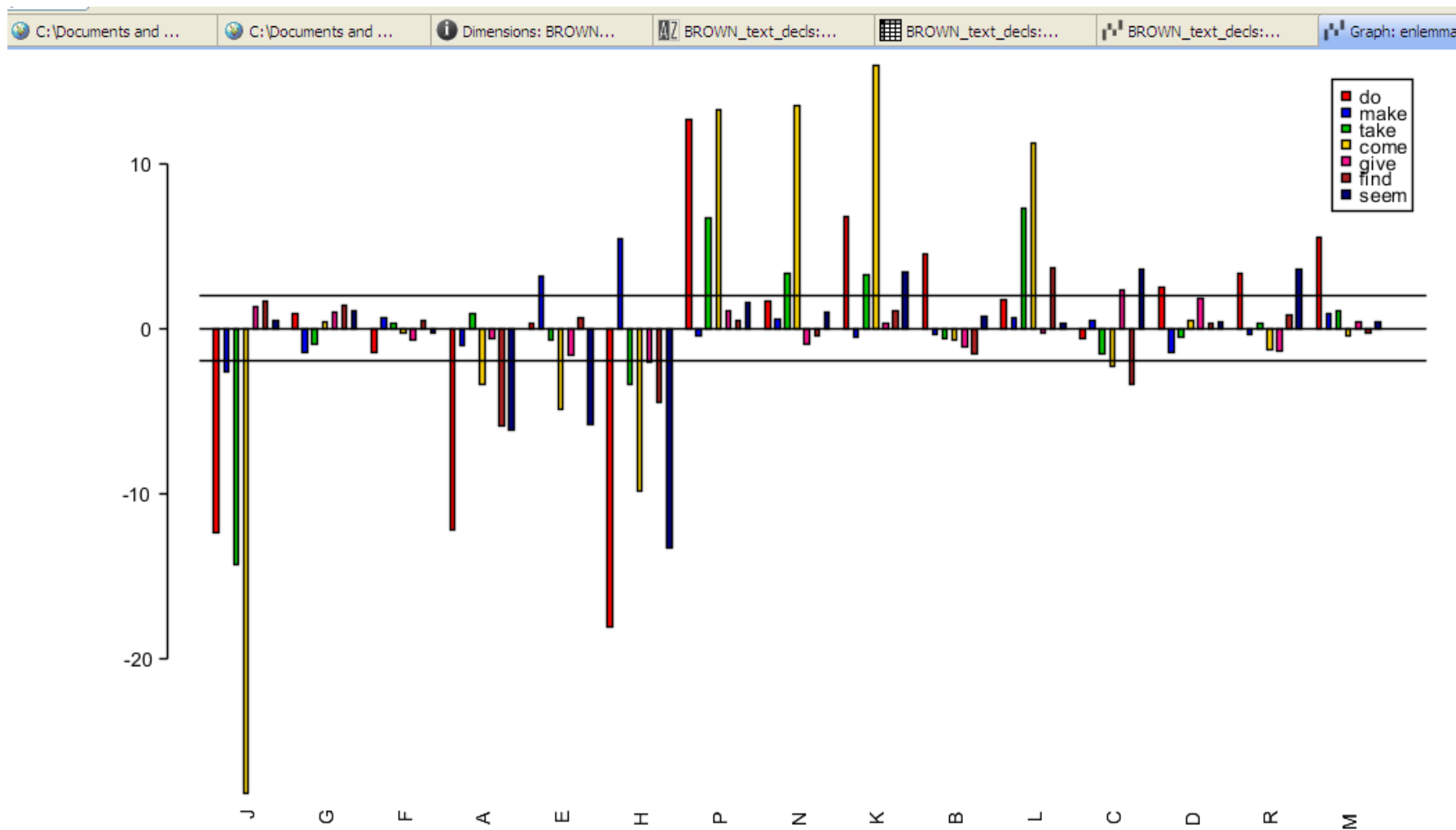
Merge or Delete columns Merge or Delete rows

enlemma	Frequency	J t=181872	G t=173074	F t=110288	A t=100541	E t=82321	H t=70104	P t=70012	N t=69328	K t=68483	B t=61592	L t=57159
#RESTE#	996814	157567	149098	94900	87619	71223	61751	58494	58052	57208	52932	47705
do	3370	382	526	289	185	239	95	310	228	270	233	192
say	2779	115	300	156	465	32	33	417	332	251	106	272
make	2322	314	315	231	184	206	196	134	145	129	120	122
go	1794	63	211	143	113	68	29	254	210	228	79	255
take	1571	141	217	151	150	102	65	146	126	124	76	127
come	1558	100	237	147	100	70	41	171	170	176	75	141
see	1511	151	189	126	68	83	38	160	186	159	61	141
get	1472	53	122	116	142	65	10	211	200	160	59	200
know	1462	89	223	126	59	57	33	181	133	156	71	176
give	1260	219	205	110	103	71	57	88	64	74	55	61
find	1032	186	175	103	50	80	34	66	58	71	41	77
use	1015	281	90	141	58	173	71	36	33	29	30	19
think	969	59	109	55	48	21	21	174	122	121	32	113
look	890	40	121	42	32	49	17	156	107	124	26	103
seem	829	134	138	78	36	27	8	64	58	73	50	41
become	766	148	160	108	49	56	21	16	24	34	42	15
tell	764	32	90	70	74	12	6	94	92	88	24	104
leave	692	74	85	63	42	43	6	74	57	63	43	83
feel	638	53	92	38	38	16	9	109	55	86	30	63
want	632	35	80	48	43	36	9	91	67	62	30	90
show	624	200	77	40	35	62	35	16	36	18	28	27
call	614	58	109	72	57	25	15	52	28	37	37	51
ask	608	28	94	49	54	15	9	60	46	77	33	69
begin	579	68	98	66	53	28	25	39	48	48	25	29
turn	563	47	64	48	31	21	4	74	78	55	16	69
follow	545	150	75	45	43	48	35	18	35	15	16	18
keep	526	40	54	66	34	57	20	62	34	41	28	46

TXM Statistics – focused specificity: verbs/text@domains

Units	Frequency T=1029742	J t=160993	score	G t=153675	score	F t=97783	score	A t=90068	score	E t=73079	score	H t=62741	score	P t=61
#RESTE#	996814	157567	1000	149098	7	94900	5	87619	1000	71223	1000	61751	1000	58
do	3370	382	-12	526	0	289	-1	185	-12	239	0	95	-18	
say	2779	115	-82	300	-9	156	-13	465	1000	32	-49	33	-38	
make	2322	314	-2	315	-1	231	0	184	-1	206	3	196	5	
go	1794	63	-61	211	-4	143	-1	113	-4	68	-8	29	-20	
take	1571	141	-14	217	0	151	0	150	0	102	0	65	-3	
come	1558	100	-28	237	0	147	0	100	-3	70	-4	41	-9	
see	1511	151	-9	189	-2	126	-1	68	-9	83	-2	38	-10	
get	1472	53	-49	122	-14	116	-1	142	0	65	-4	10	-26	
know	1462	89	-28	223	0	126	0	59	-11	57	-6	33	-11	
give	1260	219	1	205	1	110	0	103	0	71	-1	57	-2	
find	1032	186	1	175	1	103	0	50	-5	80	0	34	-4	
use	1015	281	1000	90	-8	141	5	58	-3	173	1000	71	0	
think	969	59	-19	109	-3	55	-5	48	-5	21	-11	21	-8	
look	890	40	-25	121	0	42	-7	32	-9	49	-1	17	-8	
seem	829	134	0	138	1	78	0	36	-6	27	-5	8	-13	
become	766	148	2	160	5	108	4	49	-1	56	0	21	-4	
tell	764	32	-23	90	-2	70	0	74	0	12	-11	6	-13	
leave	692	74	-3	85	-1	63	0	42	-2	43	0	6	-11	
feel	638	53	-7	92	0	38	-3	38	-2	16	-6	9	-8	
want	632	35	-14	80	-1	48	-1	43	-1	36	-1	9	-8	
show	624	200	1000	77	-1	40	-2	35	-2	62	2	35	0	
call	614	58	-5	109	1	72	1	57	0	25	-2	15	-4	
ask	608	28	-17	94	0	49	0	54	0	15	-6	9	-7	
begin	579	68	-2	98	1	66	1	53	0	28	-1	25	-1	
turn	563	47	-6	64	-2	48	0	31	-2	21	-3	4	-10	
follow	545	150	11	75	0	45	0	43	0	48	1	35	0	
keep	526	40	-7	54	-2	66	1	34	-1	57	2	20	-1	
write	524	63	-1	197	1000	65	1	15	-7	17	-3	8	-6	
put	510	45	-5	69	0	41	0	30	-1	41	0	13	-3	
hold	509	58	-2	55	-2	50	0	58	1	36	0	27	0	

TXM Statistics – focused specificity graphics: verbs/text@domains

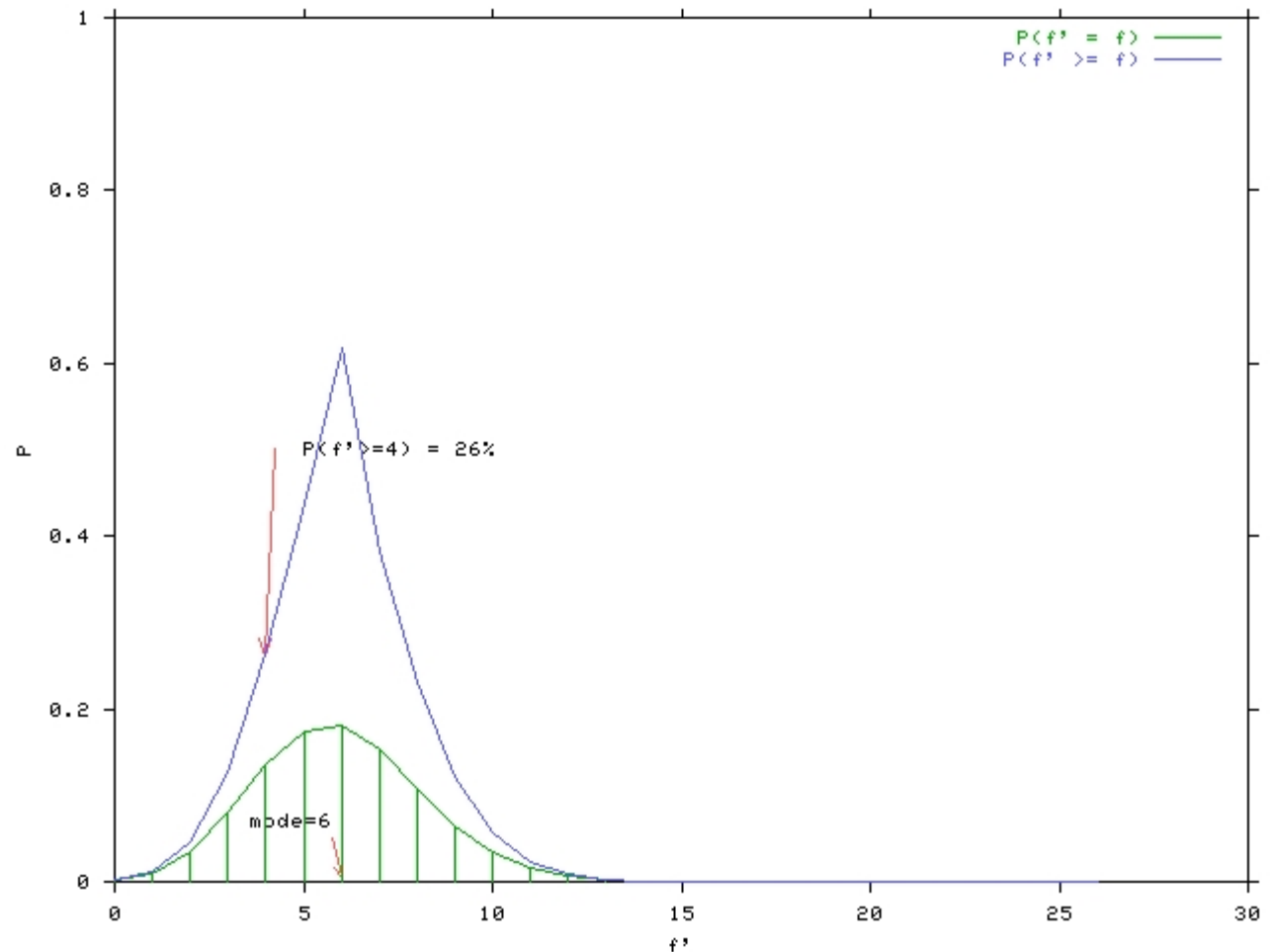


TXM Statistics – specificity model

f 4, F 29, t 11891, T 58023

$P(f' \geq 4) = 26\%$

mode = 6



TXM syntagmatic contrast:

Cooccurrences of lemma “do”

N. FICTION Adventure and Western: [enlemma="do"] (5, 5)

Query: [enlemma="do"]

Cooccurrences properties: word Thresholds: Freq ≥ 2 Co-freq ≥ 2 Score ≥ 2.0

Context: ☒ word ☐ structure ab from - 5 to - 0 and from 0 to 5

Cooccurent	Frequency	Cofrequency	Score	Mean distance
not	179	49	1000	83.4
?	518	75	14	111.8
What	39	16	10	1.9
what	110	21	6	1.9
you	362	44	6	1.6
mean	19	7	4	2.1
I	652	57	3	2.2
listen	8	4	3	1.5
did	84	13	3	315.8
know	58	10	3	1.6
job	16	5	2	2.4
something	50	9	2	1.6
shooting	6	3	2	2.3
fail	2	2	2	2.5
favor	2	2	2	2.5
process	2	2	2	3.0
Mr.	22	5	2	2.6
wear	3	2	2	3.5
filmed	3	2	2	1.0
hall	35	6	2	3.7
perhaps	9	3	2	3.3
had	591	19	-2	2.2
their	156	2	-2	4.0
her	444	12	-2	341.7
The	410	10	-2	409.0
!	314	6	-2	3.2
from	260	4	-2	2.0
on	460	10	-3	3.7

A. PRESS Reportage: [enlemma="do"] (5, 5)

Query: [enlemma="do"]

Cooccurrences properties: word Thresholds: Freq ≥ 2 Co-freq ≥ 2 Score ≥ 2.0

Context: ☒ word ☐ structure ab from - 5 to - 0 and from 0 to 5

Cooccurent	Frequency	Cofrequency	Score	Mean distance
not	254	52	1000	.3
what	76	21	12	2.0
can	93	14	5	.9
they	205	20	4	.9
something	14	5	4	2.0
you	55	9	3	.3
Why	5	3	3	.0
What	19	5	3	.2
mention	6	3	3	1.0
contracted	2	2	2	1.0
allowed	8	3	2	1.3
work	66	8	2	1.5
know	28	5	2	2.4
investment	9	3	2	3.0
could	86	9	2	.8
anything	10	3	2	1.3
chores	4	2	2	2.0
we're	4	2	2	1.0
stores	4	2	2	3.5
is	732	15	-2	2.4
at	598	11	-2	3.1
was	717	14	-2	2.9
by	497	8	-2	2.4
Mrs.	253	2	-2	4.0
from	344	4	-2	3.2
are	328	2	-3	.0
will	389	3	-3	2.7
for	943	16	-3	2.6

TXM Statistics – Cooccurrent Search: “do” lemma & “something” word

A. PRESS Reportage: [enlemma="do"] (5, 5)

Query: [enlemma="do"]

Cooccurents properties: word Thresholds: Freq ≥ 2 Co-freq ≥ 2 Score ≥ 2.0

Context: ☒ word ☐ structure ab from - 5 to 0 and from 0 to 5 ☐ include the keyword stru

Cooccurrent	Frequency	Cofrequency	Score	Mean distance
not	254	52	1000	.3
what	76	21	12	2.0
can	93	14	5	.9
they	205	20	4	.9
something	14	5	4	2.0
you	55	9	3	.3
Why	5	3	3	.0
What	10	5	3	.2

A. PRESS Reportage: ([enlemma="do"] [* [word="something"]]) | ([word="something"] [* [enlemma="do"]]) within 6

Query: [mma="do"] [* [word="something"]]) | ([word="something"] [* [enlemma="do"]]) within 6 Keyword: word

sort keys: #1 None #2 None #3 None #4 None

1-5 / 5

text_id	Left context	Keyword	Right c
a02	the confidence of bank customers . `` If you destroy confidence in banks , you	do something	to the e
a08	most politicos would care to acknowledge publicly . They would like to convey the notion	something is being done	, even t
a08	care to acknowledge publicly . They would like to convey the notion something is being	done , even though it is something	they kn
a31	Crusade school here about six months ago , Jim became convinced that an individual can	do something	constru
a37	Rayburn , after huddling in Palm Beach with President-elect Kennedy , decided that this year	something had to be done	about th

TXM Statistics – Cooccurents

Concordance to fulltext: Edition reading

A. PRESS Reportage: [enlemma="do"] (5, 5) | A. PRESS Reportage:([enlemma="do"] [* [word="something"]] | ([word="something"] [* [enlemma="do"]] within 6

Query: Keyword: word

sort keys: #1 #2 #3 #4

1-5 / 5

text_id	Left context	Keyword	Right context
a02	the confidence of bank customers . `` If you destroy confidence in banks , you	do something	to the economy " , he said . `` You take out of
a08	most politicians would care to acknowledge publicly . They would like to convey the notion	something is being done	, even though it is something they know to b
a08	care to acknowledge publicly . They would like to convey the notion something is being	done , even though it is something	they know to be ineffectual . Underlying con
a31	Crusade school here about six months ago , Jim became convinced that an individual can	do something	constructive in the ideological battle and set
a37	Rayburn , after huddling in Palm Beach with President-elect Kennedy , decided that this year	something had to be done	about the Rules Committee - and that he was

BROWN Edition - Page 5

file:///C:/Documents and Settings/sheiden/TXM/corpora/brown/HTML/BROWN/default/a37_5.html

There was sufficient pretext to demand Colmer's ouster: he had given his lukewarm support to the anti-Kennedy electors in Mississippi. Reprisals are not unheard of in such situations, but the recent tendency has been for the Congress to forgive its prodigal sons. In 1949 the Dixiecrats escaped unscathed after their 1948 rebellion against Harry Truman, and in 1957, after Congressman Adam Clayton Powell campaigned for Dwight Eisenhower in 1956, his fellow Democrats did not touch his committee assignments, although they did strip him temporarily of his patronage. (In the heat of the anti-Colmer drive last week, Judge Smith threatened reprisal against Powell. Said he: `` We will see whether whites and Negroes are treated the same around here .") But Speaker Sam Rayburn, after huddling in Palm Beach with President-elect Kennedy, decided that this year something had to be done about the Rules Committee- and that he was the only man who could do anything effective.

In a tense, closed-door session with Judge Smith, Rayburn attempted to work out a compromise: to add three new members to the Rules Committee (two Democrats, including one Southerner, and one Republican). Smith flatly rejected the offer, and Mister Sam thereupon decided to join the rebels. The next morning he summoned a group of top Democrats to his private office and broke the news: he would lead the fight to oust

<< < 5 / 8 > >>

TXM demo portal:

<http://txm.bfm-corpus.org/bfm/?locale=en>

The screenshot displays the TXM demo portal interface, which is a web-based tool for corpus analysis. The interface is divided into several sections:

- Header:** Includes the TXM logo, navigation links (Fichier, Édition, Affichage, Historique, Delicious, Marque-pages, Outils, Aide), and user information (Alexei Lavrentiev, Mon profil, Se déconnecter, Aide, Contact, fr).
- Left Panel:** Contains a sidebar with a search bar (Query: [pos="ADJ.*"] [wo]) and a list of corpora (BFM1_09, GRAAL). Below this is a table showing keyword frequency for the selected corpus.
- Table:** A table with columns 'Keyword' and 'Frequency'. It lists various forms of 'chevalier' and their occurrences.
- Main Content Area:** Displays a search result for the query 'bon chevalier'. It shows a list of references (1-6) with their corresponding text snippets. The text is in Old French and includes words like 'chevalier', 'Galaad', and 'Jhesucrist'.
- Right Panel:** Contains a sidebar with a search bar (Query: GRAAL) and a list of corpora (BFM1_09, GRAAL). Below this is a table showing keyword frequency for the selected corpus.
- Bottom Panel:** Displays a detailed view of the search results, including a table with columns 'Reference', 'Left context', 'Keyword', and 'Right context'. It shows the context of the keyword 'bon chevalier' in the selected corpus.

The interface is designed to allow users to search for specific keywords or phrases within a corpus and analyze the results in detail.

TXM portal : Frequency list of <adjectives followed by « chevalier » sequences>

CQP Query : [pos="ADJ.*"] [word="chevaliers?"%c]

The screenshot shows the TXM portal interface in Mozilla Firefox. The browser window title is "TXM - Mozilla Firefox". The address bar shows the URL "txm.risc.cnrs.fr/bfm". The page has a navigation bar with links: "Welcome", "Subscribe", "Login", "Help", "Contact". The main content area displays a frequency list for the query "[pos="ADJ.*"] [word="chevaliers?"%c]". The list is titled "Index /GRAAL:[pos="ADJ.*"] [word="chevaliers?"%c]". The table shows the following data:

	Keyword	Frequency
1	bons chevaliers	24
2	bon chevalier	15
3	mieldres chevaliers	11
4	verai chevalier	9
5	dui chevalier	8
6	verais chevaliers	8
7	autres chevaliers	7
8	.ii. chevaliers	5
9	meillor chevalier	5
10	meillors chevaliers	5
11	biaux chevaliers	3

At the bottom of the table, it says "41 forms (136 occurrences)". The pagination bar shows "1-41/41" and "Export".

TXM portal : Kwic Concordance of « bon » / « chevalier » and Critical edition

The screenshot displays the TXM portal interface, which is used for linguistic analysis of medieval texts. The interface is divided into several sections:

- Top Bar:** Includes navigation links (Welcome, Subscribe, Login, Help, Contact) and a language selector (en).
- Corpus Selection:** On the left, a sidebar shows the selected corpus (GRAA) and other available corpora (BFM1).
- Search Results:** The main area displays a Kwic Concordance for the query "bon" "chevalier". It shows 15 occurrences of the query in the corpus. The results are presented in a table with columns for Reference, Left context, Keyword, and Right context.
- Critical Edition:** On the right, a critical edition of the text is displayed, showing the original manuscript text in a medieval script.
- Navigation and Settings:** At the bottom, there are controls for page navigation (1-15/15), a search bar, and a settings menu.

The Kwic Concordance results are as follows:

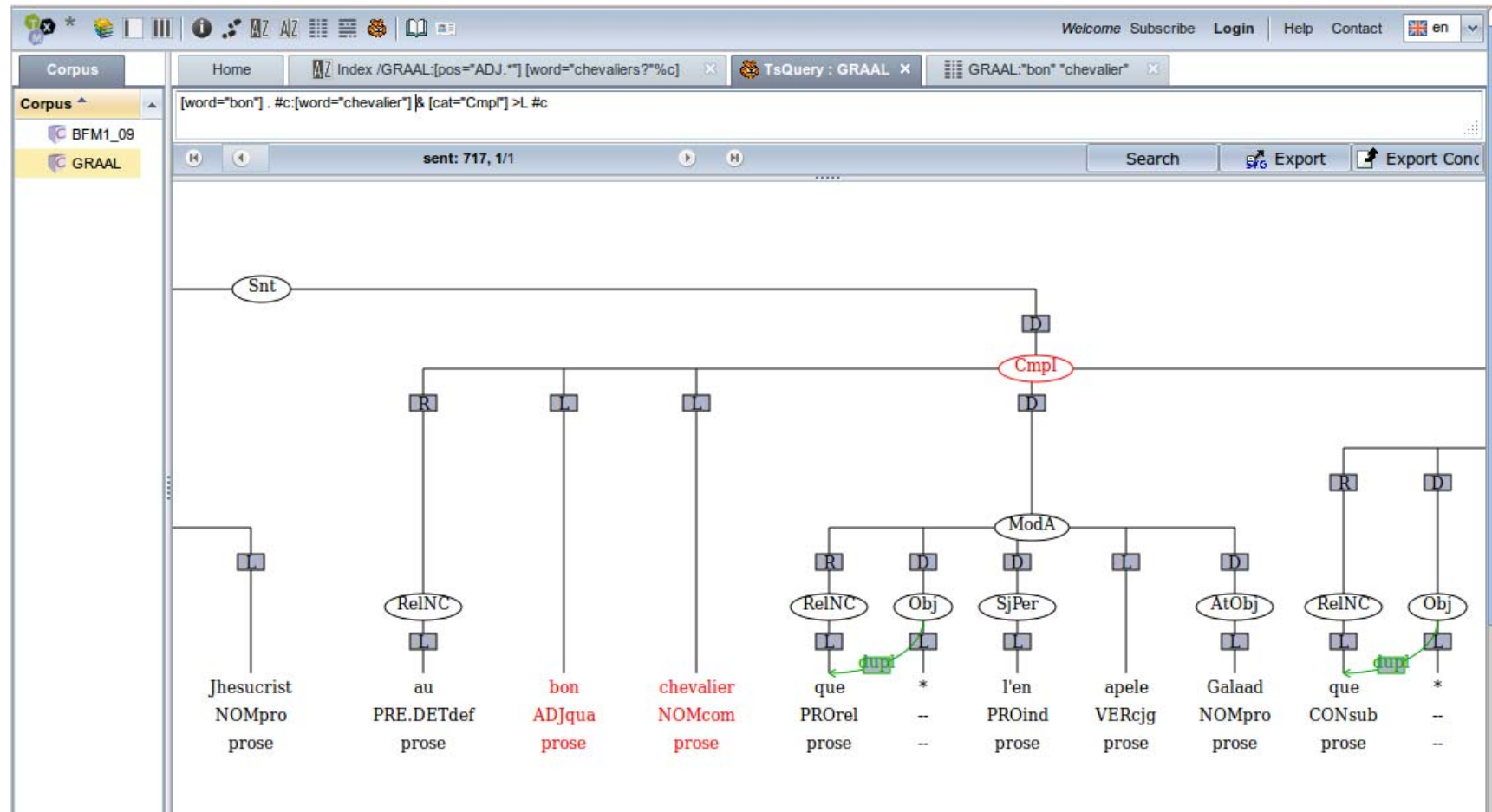
Reference	Left context	Keyword	Right context
1	QGRAALFRO	mon seignor Gauvain et li dist : « Biaux niés or avons nos Galaad le bon chevalier parfet que nos et cil de la table reonde avons tant desirré a avoir,	
2	QGRAALFRO	Tien , va t'en , et porte cest escu au serjant Jhesucrist , au bon chevalier que l'en apele Galaad que tu lessas ore en l'abeie, et li di	
3	QGRAALFRO	ne desirrai onques mes autant chose que je veïsse come je fesoie a conoistre le bon chevalier qui de cest escu porteroit la seignorie. » Et Galaad respont qu'il le	
4	QGRAALFRO	meslee lor avoit celui jor Galaad rendu , et cil qui mout estoient preudome et bon chevalier les meinent si mal qu'il les ocient en poi d'ore. Si les	
5	QGRAALFRO	qu' il estoit las , se ne fust ce qu' il ne pot oublier le bon chevalier qui le blanc escu emporte, et quant il s'est grant piece esperiz si	
6	QGRAALFRO	le commença a besier et dist que ce avoit il fet por l' amor dou bon chevalier qui s'i reposeroit, et il li distrent maintenant : " Mellin que porra	

The critical edition on the right shows the original text in a medieval script, with the same query highlighted in red.

TXM portal : syntactic tree containing « bon » / « chevalier » as a complement

Tiger Search Query :

[word="bon"] . #c:[word="chevalier"]&[cat="Cmpl"] >L #c



TXM portal : Corpus selection

TXM portal : Corpus selection

Home Sélection BFM1_09 x

Remise à zéro Tout Stats Valider

Corpus

BFM1_0

GRAAL

sigles (0/29)

auteur (0/9)

titre (0/23)

siecle (0/6)

date oeuvre

date ms

forme (0/3)

domaine (3/5)

valeur n N t T

historique 5 5 321 321

juridique 0 1 0 129

litteraire 11 11 937 937

genre (0/11)

dialecte (0/6)

morphosynt (0/2)

restrictions (0/3)

Fiches	T	auteur	titre
<input type="checkbox"/>	129	anonyme	Serments de Stra
<input type="checkbox"/>	215	anonyme	Séquence de sain
<input type="checkbox"/>	5806	anonyme	Vie de saint Alexi
<input checked="" type="checkbox"/>	16854	Philippe de Thaon	Comput
<input checked="" type="checkbox"/>	1509	anonyme	Description d'Eng
<input checked="" type="checkbox"/>	18037	Wace	Brut
<input checked="" type="checkbox"/>	1587	anonyme	Brut
<input checked="" type="checkbox"/>	66388	Jean Renart	Escoufle
<input type="checkbox"/>	19790	Gautier de Coinci	Miracles de Nostr
<input type="checkbox"/>	48658	Gautier de Coinci	Miracles de Nostr
<input type="checkbox"/>	82106	Gautier de Coinci	Miracles de Nostr
<input type="checkbox"/>	101603	Gautier de Coinci	Miracles de Nostr
<input checked="" type="checkbox"/>	118622	anonyme	Queste del saint c
<input checked="" type="checkbox"/>	98863	anonyme	Mort le Roi Artu
<input checked="" type="checkbox"/>	143687	anonyme	Suite du roman de

Compter en : Mots Textes

sigles

auteur

titre

siecle (6)

13

15

14

12

text_bfndate

text_bfmsdate

forme

domaine

genre

dialecte

morphosynt

restrictions



Thank you!

<http://textometrie.ens-lyon.fr>

textometrie@ens-lyon.fr