Digital Literary Stylistics

Anne BANDRY-SCUBBI

Women's Novels 1750s-1830s and the Company They Keep:

A Computational Stylistic Approach

'what **corpus stylistics** can do beyond the obvious provision of quantitative data, is help with the analysis of an individual text by providing various options for the **comparison of one text with groups of other texts to identify tendencies, intertextual relationships, or reflections of social and cultural contexts**'.

Michaela Mahlberg

'Corpus Stylistics: Bridging the Gap between Linguistic and Literary Studies', in *Text, Discourse and Corpora: Theory and Analysis*, by Michael Hoey, Michaela Mahlberg, Michael Stubbs, Wolfgang Teubert (London: Continuum, 2007), p.221.

ROUTLEDGE ADVANCES IN CORPUS LINGUISTICS

Corpus Stylistics and Dickens's Fiction

Michaela Mahlberg

Routledge

CORPUS LINGUISTICS IN LITERARY ANALYSIS

BETTINA FISCHER-STARCKE

Jane Austen and her Contemporaries

STUDIES IN CORPUS AND DISCOURSE
SERIES EDITORS: WOLFGANG TEUBERT AND MICHAELA MAHLBERG

STYLE IN FICTION
A LINGUISTIC INTRODUCTION TO ENGLISH FICTIONAL PROSE

Geoffrey N. Leech & Michael H. Short

ENGLISH·LANGUAGE·SERIES

J. F. BURROWS

COMPUTATION INTO CRITICISM
A Study of Jane Austen's Novels and an Experiment in Method

CLARENDON PRESS
OXFORD

Chawton Novels on line

- 34 "**domestic dramas** where heroines blush, swoon, or face unbearable social ostracism because of minor breaches of decorum"

  The English Novel 1770-1829, Peter Garside, James Raven and Rainer Schöwerling eds (OUP, 2000), I p. 28

- "the '**feminine' novel**—domestic comedy, centring on a heroine, in which the critical action is an inward progress towards judgment"

  Marilyn Butler, *Austen and the War of Ideas* p.145

→**CHAWTN34**: 3.9 million words

38% pre 1800, 62% post1800

- CHAWTN34 $\subset$ W42

- Haywood, Burney, Edgeworth, Austen, Ferrier

$\rightarrow$ W42:   5.8 million words

41% pre 1800, 51% post1800

$\rightarrow$ CHAWTN34: 3.9 million words

38% pre 1800, 62% post1800

- # CHAWTN34 $\subset$ W42
  1769-1830   1752-1834
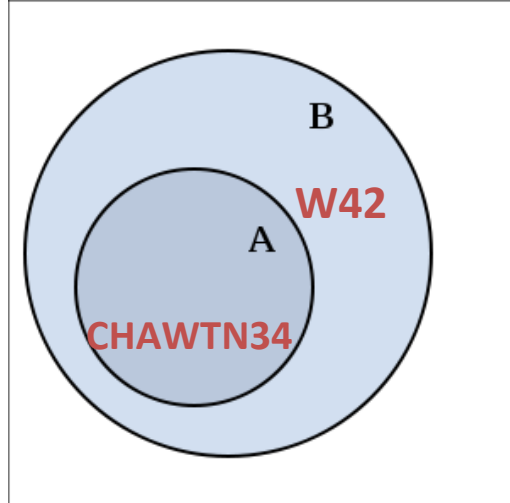
- # **CTROL34** : 5.4 million words
  - 1748-1834
  - 41% pre1800
  - 44% male
  - Clive Probyn's *English Fiction of the 18th Century, 1700-1789* and Gary Kelly's *English Fiction of the Romantic Period 1789-1830* **(Longman 1987, 1990)**
  - **http://www.gutenberg.org/**
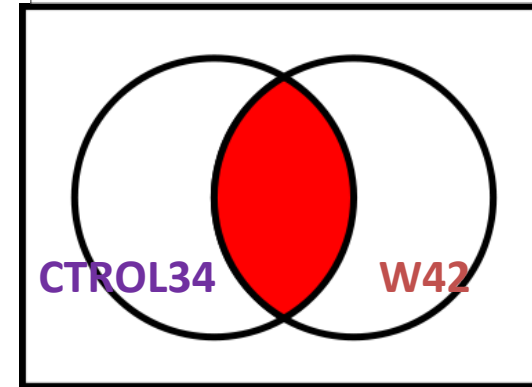
http://fr.wikipedia.org/wiki/Alg%C3%A8bre_des_parties_d%27un_ensemble#mediaviewer/

Figure 1 – Lexical connection on types for CNTROL34 (Principal component analysis, *Hyperbase*)
Titles in red by women, in black by men
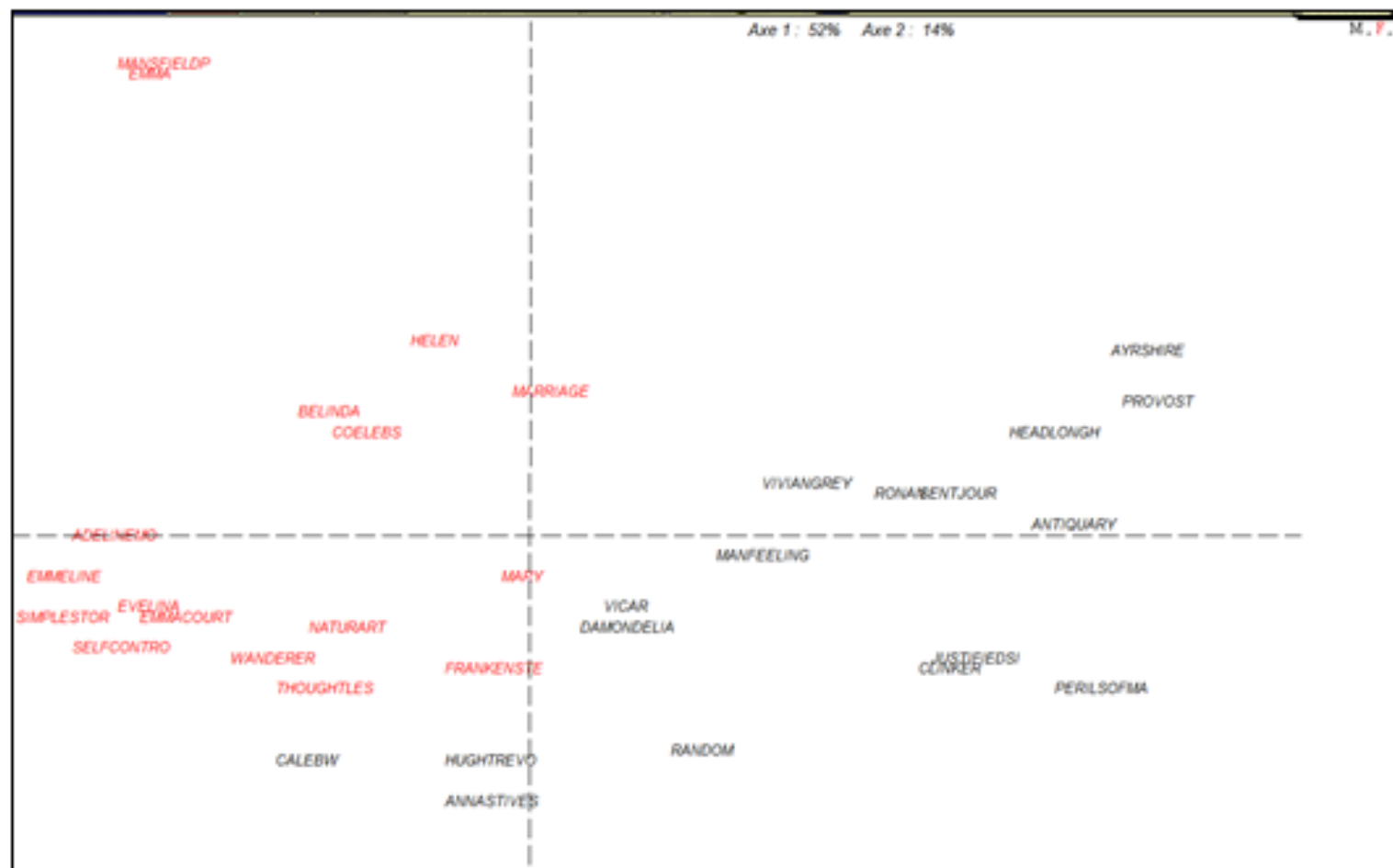
Figure 2– lexical connection on tokens for *CNTROL34* (Principal component analysis, *Hyperbase*)
Titles in red by women, in black by men

M . F.

HELEN

EMMA

RONAN   ANTIQUARY

BELINDA

VIVIANGREY

MANSFIELDP   MARRIAGE

PERILSOFMA

EVELINA   DAIMONDELIA   AYRSHIRE

SIMPLESTOR   HEADLONGH

COELEBS   JUSTIFIEDSI

MANFEELING   PROVOST

NATURART   SENTJOUR

CLINKER

ANNASTIVES

THOUGHTLES   MARY

VICAR   CALEBW   RANDOM

WANDERER   FRANKENSTE

ADELINEMO

EMMELINE   HUGHTREVO

SELFCONTRO   EMMACOURT

Axe 1 : 44%   Axe 2 : 19%

Figure 3 – Words whose relative frequency is higher
in texts written by female authors than by male
authors (CNTROL34): top of the 'positive specificity'
list established by *Hyperbase* (left) exported into
Excel (right)

(326 words in all, 266 without names)

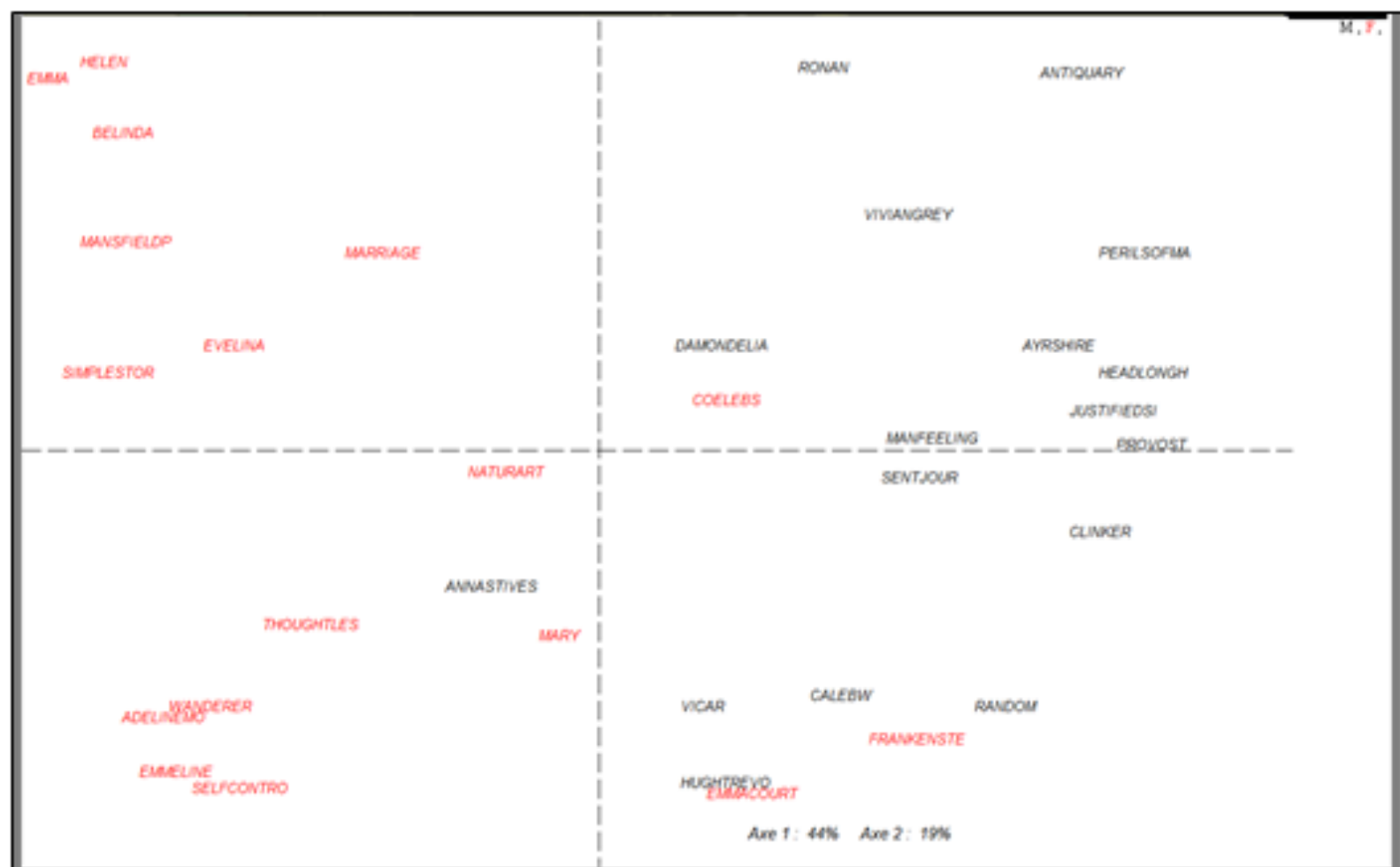| écart	corpus	texte	mot	HIERAR POSITIF
37.58	1529	1523	helen	
37.58	1497	1496	emmeline	
37.58	1252	1252	adeline	
37.58	1246	1246	delacour	
37.58	1219	1218	cecilia	
37.58	1153	1128	ellis	
37.58	1118	1117	juliet	
37.58	1016	1016	delamere	
37.58	977	976	emma	
37.58	971	968	fanny	
37.58	949	949	belinda	
37.58	934	934	montreville	
37.58	924	924	betsy	
37.57	56675	45923	her	
37.57	43315	36227	she	
37.57	9615	7876	lady	
37.57	8196	6819	miss	
37.57	6937	5803	mrs	
37.57	4146	3588	herself	
37.57	2217	2121	laura	
37.39	41516	26211	you	
36.03	3792	2999	cried	
35.57	773	773	harleigh	
35.41	110230	66123	-	
32.70	667	667	davenant	
32.61	664	664	hervey	

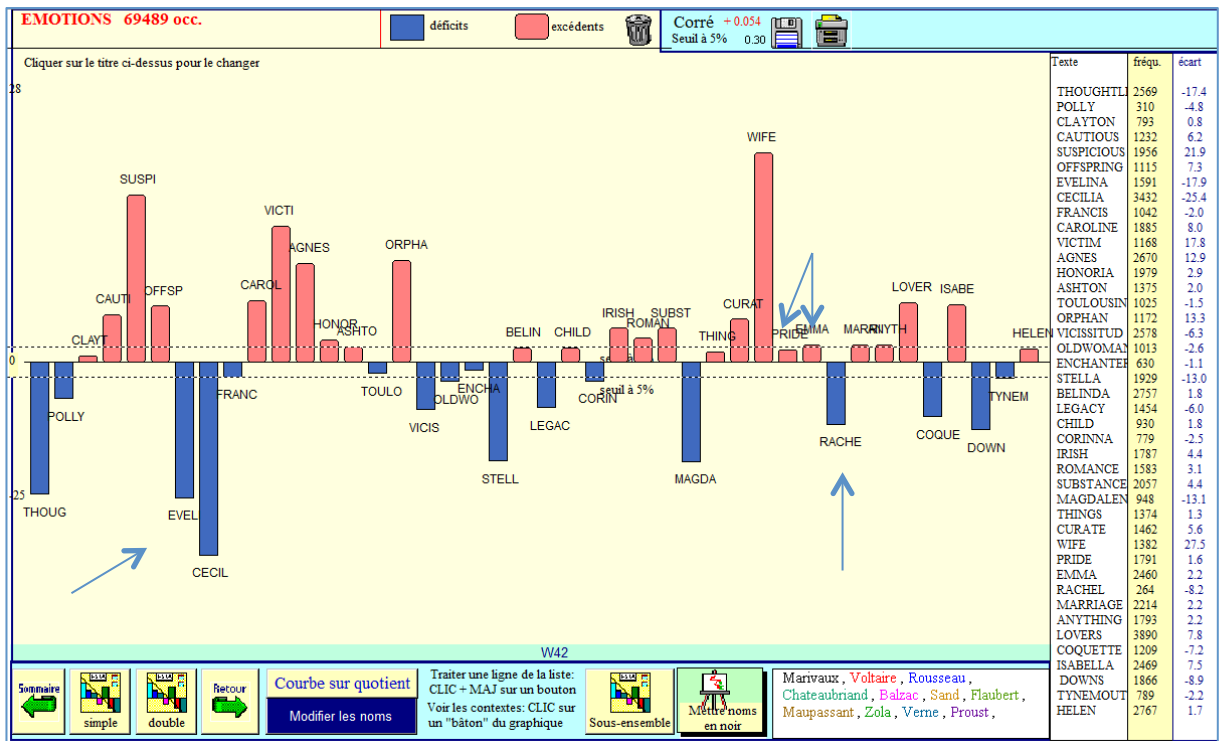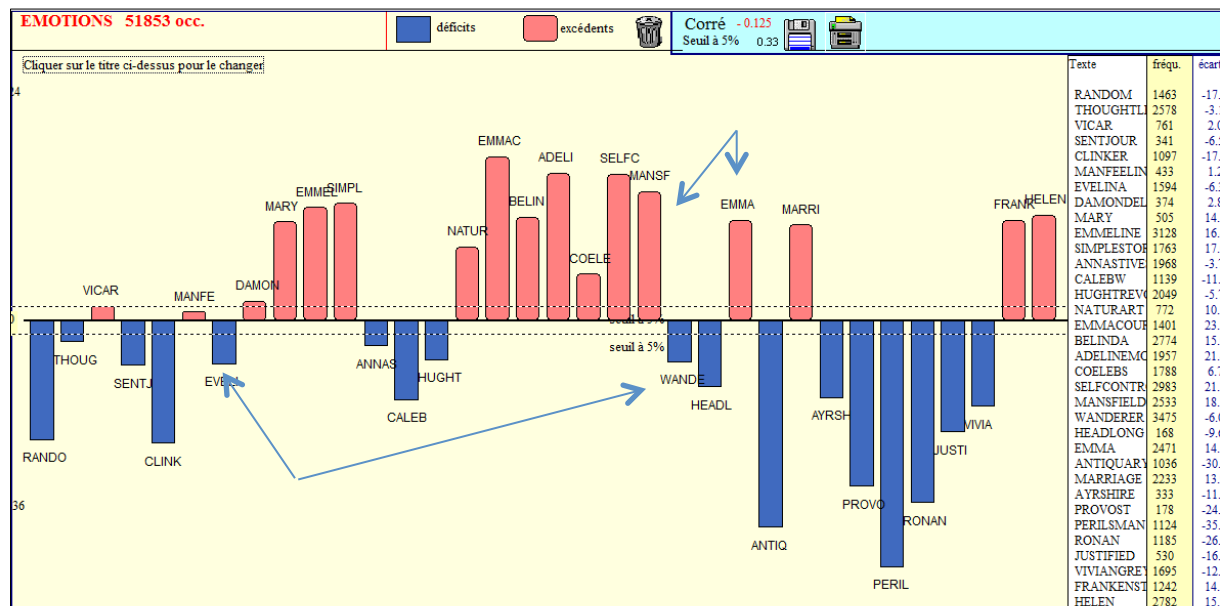| CNTROL34	Female authors	56%	
56 675	45 923	81%	her
43 315	36 227	84%	she
9 615	7 876	82%	lady
8 196	6 819	83%	miss
6 937	5 803	84%	mrs
4 146	3 588	87%	herself
41 516	26 211	63%	you
3 792	2 999	79%	cried

Figure 4 – Words whose relative frequency is lower in texts written by female authors than by male authors (CNTROL34): top of the 'negative specificity' list established by *Hyperbase* (67 words in all)

écart	corpus	texte	mot HIERAR NEGATIF
-37.55	217241	104872	the
-37.55	89135	44774	a
-36.49	46468	22547	his
-35.67	141241	72959	of
-35.39	5841	2111	our
-33.69	138822	71933	and
-29.40	10104	4347	we
-28.53	1627	414	master
-24.75	7134	3050	man
-21.78	2080	716	men
-19.45	3013	1195	old
-18.98	1125	340	fellow
-18.89	69171	36236	in
-17.94	25873	13082	me
-17.77	4157	1791	us
-16.30	4539	2021	like
-12.78	1751	723	three
-12.16	1621	671	country
-12.11	15300	7807	an
-11.81	1872	800	gentleman
-11.80	13971	7115	they
-11.22	1339	550	captain
-10.97	27750	14573	which
-10.61	53644	28678	he
-10.52	9424	4758	there
-10.26	19927	10398	on
-10.16	1065	434	small
-10.11	991	399	case
-9.96	1119	463	god

CTROL34

EMOTIONS 51853 occ. — déficits — excédents — Corré −0.125 — Seuil à 5% 0.33

Cliquer sur le titre ci-dessus pour le changer

seuil à 5%

Texte	fréqu.	écart
RANDOM	1463	-17.
THOUGHTLI	2578	-3.
VICAR	761	2.0
SENTJOUR	341	-6.
CLINKER	1097	-17.
MANFEELIN	433	1.2
EVELINA	1594	-6.
DAMONDEL	374	2.8
MARY	505	14.
EMMELINE	3128	16.
SIMPLESTOP	1763	17.
ANNASTIVE	1968	-3.
CALEBW	1139	-11.
HUGHTREV	2049	-5.
NATURART	772	10.
EMMACOUR	1401	23.
BELINDA	2774	15.
ADELINEMO	1957	21.
COELEBS	1788	6.
SELFCONTR	2983	21.
MANSFIELD	2533	18.
WANDERER	3475	-6.
HEADLONG	168	-9.
EMMA	2471	14.
ANTIQUARY	1036	-30.
MARRIAGE	2233	13.
AYRSHIRE	333	-11.
PROVOST	178	-24.
PERILSMAN	1124	-35.
RONAN	1185	-26.
JUSTIFIED	530	-16.
VIVIANGRE	1695	-12.
FRANKENST	1242	14.
HELEN	2782	15.

W42

EMOTIONS 69489 occ. — déficits — excédents — Corré +0.054 — Seuil à 5% 0.30

Cliquer sur le titre ci-dessus pour le changer

seuil à 5%

Texte	fréqu.	écart
THOUGHTLI	2569	-17.4
POLLY	310	-4.8
CLAYTON	793	0.8
CAUTIOUS	1232	6.2
SUSPICIOUS	1956	21.9
OFFSPRING	1115	7.3
EVELINA	1591	-17.9
CECILIA	3432	-25.4
FRANCIS	1042	-2.0
CAROLINE	1885	8.0
VICTIM	1168	17.8
AGNES	2670	12.9
HONORIA	1979	2.9
ASHTON	1375	2.0
TOULOUSIN	1025	-1.5
ORPHAN	1172	13.3
VICISSITUD	2578	-6.3
OLDWOMAN	1013	-2.6
ENCHANTER	630	-1.1
STELLA	1929	-13.0
BELINDA	2757	1.8
LEGACY	1454	-6.0
CHILD	930	1.8
CORINNA	779	-2.5
IRISH	1787	4.4
ROMANCE	1583	3.1
SUBSTANCE	2057	4.4
MAGDALEN	948	-13.1
THINGS	1374	1.3
CURATE	1462	5.6
WIFE	1382	27.5
PRIDE	1791	1.6
EMMA	2460	2.2
RACHEL	264	-8.2
MARRIAGE	2214	2.2
ANYTHING	1793	2.2
LOVERS	3890	7.8
COQUETTE	1209	-7.2
ISABELLA	2469	7.5
DOWNS	1866	-8.9
TYNEMOUT	789	-2.2
HELEN	2767	1.7

Sommaire — simple — double — Retour — Courbe sur quotient / Modifier les noms — W42 — Traiter une ligne de la liste: CLIC + MAJ sur un bouton. Voir les contextes: CLIC sur un "bâton" du graphique — Sous-ensemble — Mettre noms en noir — Marivaux , Voltaire , Rousseau , Chateaubriand , Balzac , Sand , Flaubert , Maupassant , Zola , Verne , Proust ,

Lexical connection on types in W42 (*Hyperbase*)
Canonical novels in black, Minerva novels in blue,
non-Minerva Chawton novels in red

Canon , Chawton , Lane ,

MAGDALEN   STELLA   CHILD

TYNEMOUTH   CURATE ANYTHING   THINGS SAFE
                         ISABELLA

IRISH   AGNES ASHTON   SUSPICIOUS
                         FRANCIS VICTIM   HONORIA
                         LEGACY   ORPHAN

            SUBSTANCE   OLDWOMAN   VICISSITUDE
                                           CAROLINE

        LOVERS   RACHEL   OFFSPRING   CAUTIOUS

                       ROMANCE
    COQUETTE   CORINNA   ENCHANTER POLLY   THOUGHTLES
                                          CLAYTON

                                      TOULOUSIN
                              CECILIA

                              EVELINA

        MARRIAGE   PRIDE

                        BELINDA

DOW

            HELEN   EMMA

Axe 1 : 58%   Axe 2 : 13%

Lexical connection on tokens in W42 (*Hyperbase*)
Canonical novels in black, Minerva novels in blue,
non-Minerva Chawton novels in red

Canon , Chawton , Lane

HELEN

MARRIAGE

CORINNA

SUBSTANCE

ISABELLA

PRIDE   EMMA

ROMANCE   RACHEL   THINGS
              LOVERS   ENCHANTER
                                            BELINDA
COQUETTE
                              CECILIA

ANYTHING

CURATE
DOWNS

CAROLINE

MAGDALEN

                    THOUGHTLES

HONORIA
TYNEMOUTH   ORPHAN
            CHILD   EVELINA
STELLA IRISH   ASHTON

POLLY   LEGACY

FRANCIS
                            TOULOUSIN
            AGNES   OFFSPRING   CAUTIOUS

                    VICTIM
              CLAYTON   OLDWOMAN   WIFE
                                        SUSPICIOUS
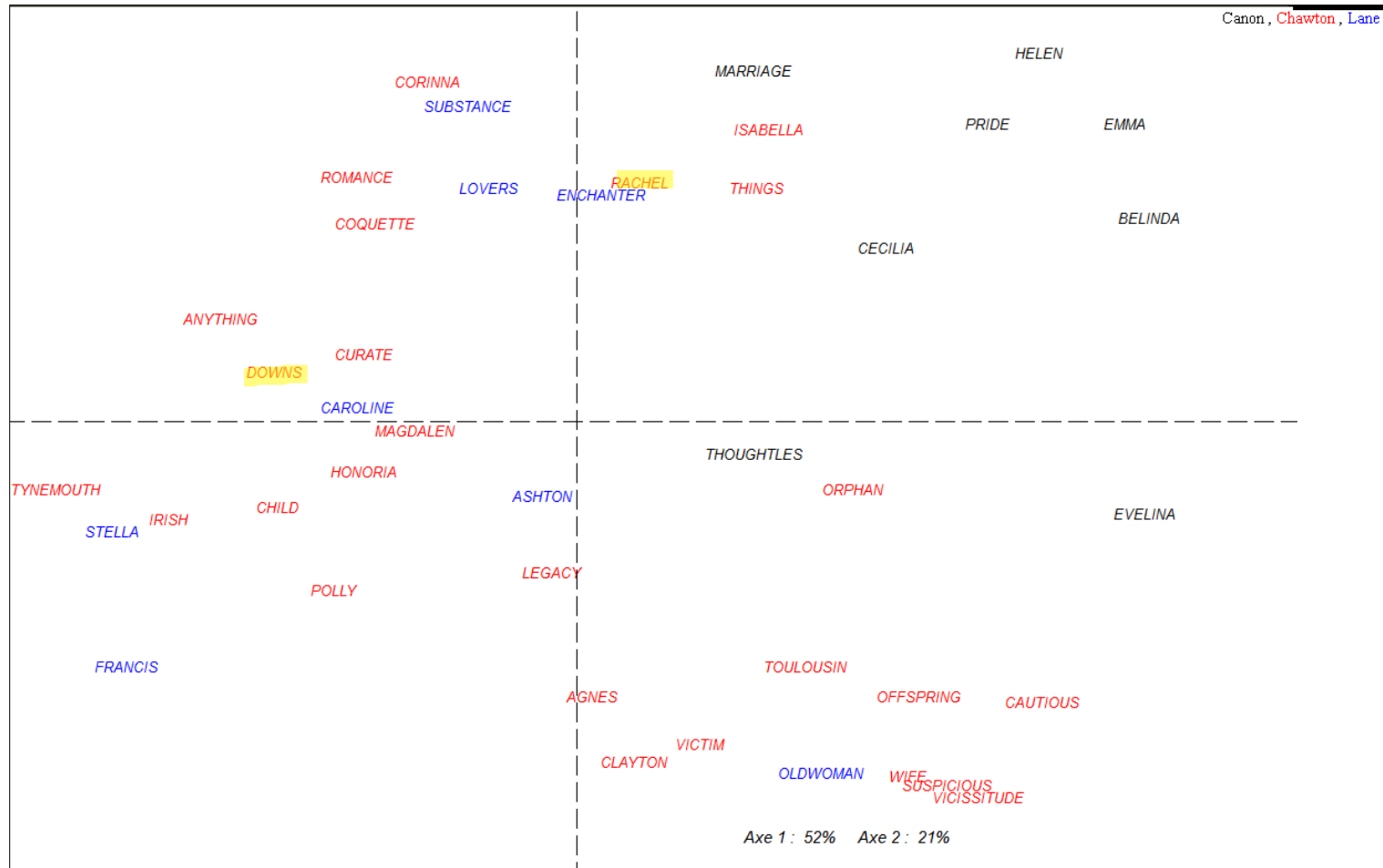   VICISSITUDE

Axe 1 : 52%   Axe 2 : 21%

Lexical connection on types in W42 (*Hyperbase*)
Canonical novels in black, Minerva novels in blue,
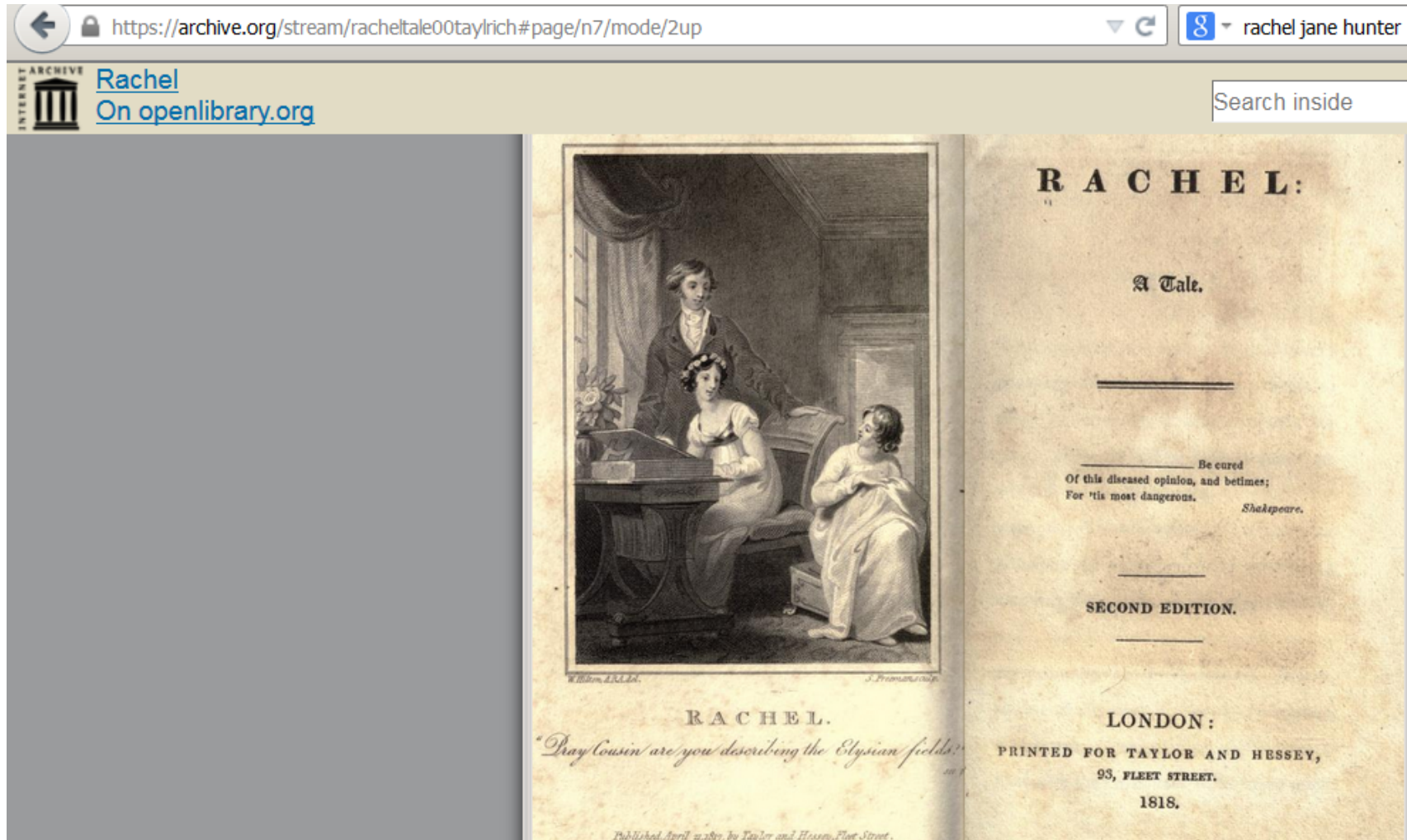## non-Minerva Chawton novels in red

Canon , Chawton , Lane ,

MAGDALEN   STELLA   CHILD

TYNEMOUTH   CURATE ANYTHING   THINGS   SAFE
                        ISABELLA

IRISH   AGNES ASHTON   SUSPICIOUS
                          FRANCIS VICTIM   HONORIA
                          LEGACY   ORPHAN

        SUBSTANCE   OLDWOMAN   VICISSITUDE
                                        CAROLINE

    LOVERS   RACHEL   OFFSPRING   CAUTIOUS

            ROMANCE
    COQUETTE   CORINNA   ENCHANTER POLLY   THOUGHTLES
                                CLAYTON

                            TOULOUSIN
DOW   CECILIA

                        EVELINA

        MARRIAGE   PRIDE

                    BELINDA

        HELEN   EMMA

            Axe 1 : 58%   Axe 2 : 13%

Lexical connection on tokens in W42 (*Hyperbase*)
Canonical novels in black, Minerva novels in blue,
## non-Minerva Chawton novels in red

Canon , Chawton , Lane

HELEN
MARRIAGE
CORINNA
SUBSTANCE
ISABELLA   PRIDE   EMMA
ROMANCE   LOVERS   RACHEL   THINGS
ENCHANTER
BELINDA
COQUETTE
CECILIA
ANYTHING
CURATE
DOWNS
CAROLINE
MAGDALEN
THOUGHTLES
HONORIA   ORPHAN
TYNEMOUTH   EVELINA
CHILD   ASHTON
IRISH
STELLA
LEGACY
POLLY
FRANCIS
TOULOUSIN
AGNES   OFFSPRING   CAUTIOUS
VICTIM
CLAYTON   OLDWOMAN   WIFE
SUSPICIOUS
VICISSITUDE

Axe 1 : 52%   Axe 2 : 21%

Rachel, Jane Hunter (1817):
A 'highly original' tale told in unoriginal terms

W42+Display: Lexical connection types
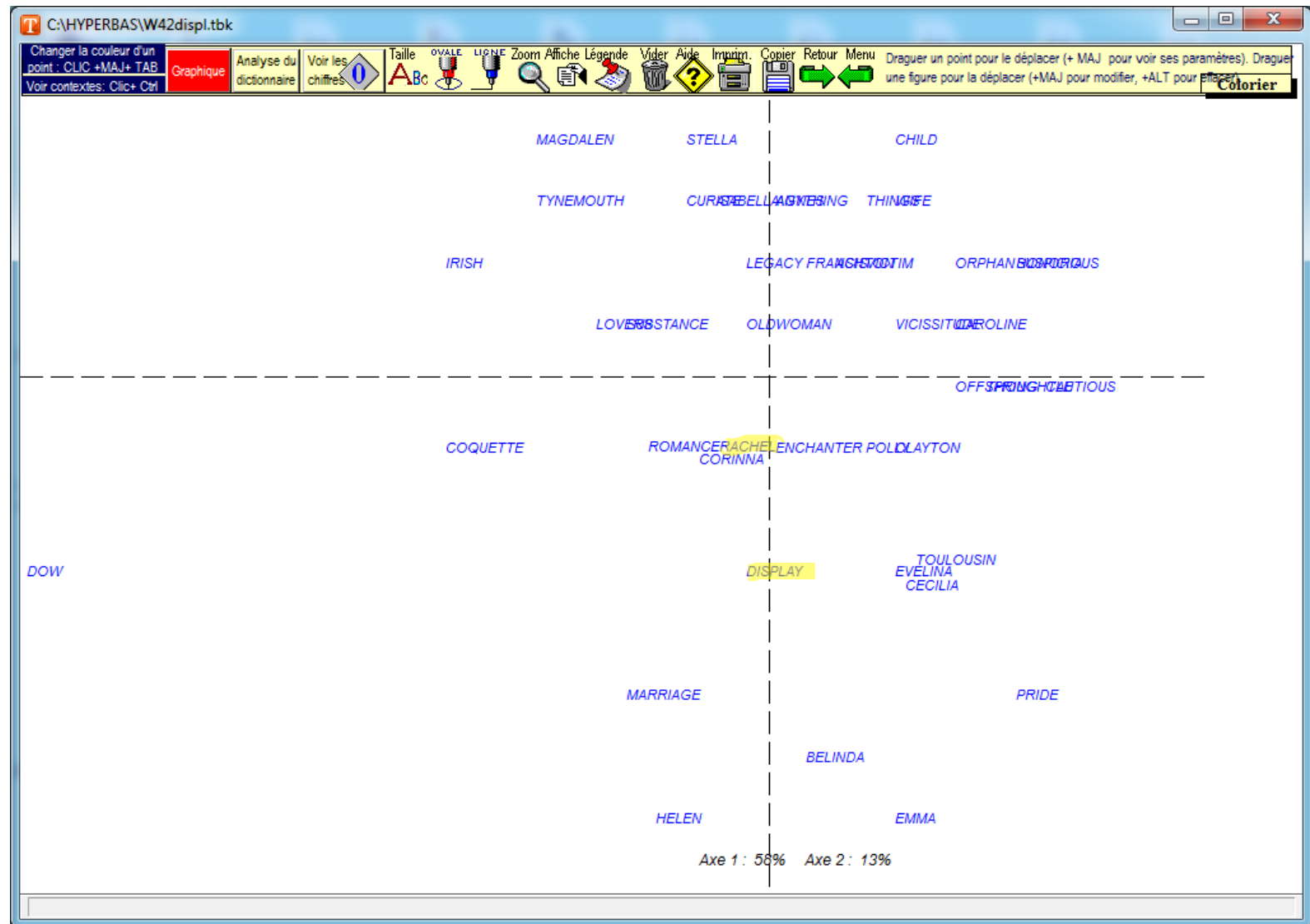
W42+Display: Lexical connection tokens
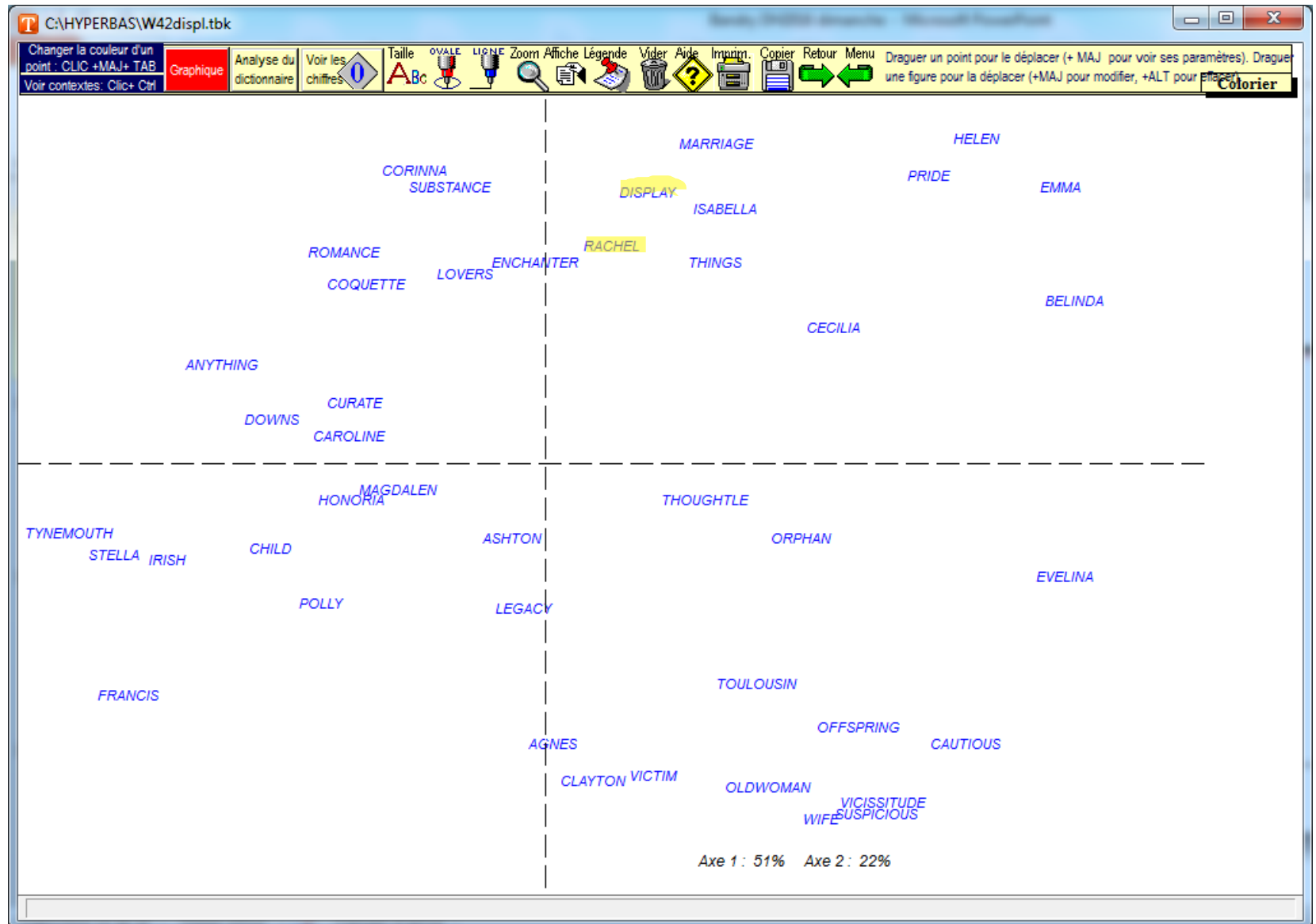
What is style?

- Oh, well….

- Contrastive stylistics
  - *XVII-XVIII:* Sterne & Swift, fiction & sermons
  - JADT 2010: Smollett's *Roderick Random*

- ***ABO: Interactive Journal for Women in the Arts, 1640-1830 :***
  [Chawton Novels Online, Women's Writing 1751-1834 and Computer-Aided Textual Analysis](http://scholarcommons.usf.edu/abo/vol5/iss2/1/)
  http://scholarcommons.usf.edu/abo/vol5/iss2/1/

- ***Georgian Cities: 18th-century Cities***
  [The Use of Toponyms in *Pride and Prejudice*](http://www.18thc-cities.paris-sorbonne.fr/Space-and-Emotions)
  http://www.18thc-cities.paris-sorbonne.fr/Space-and-Emotions

bandry@unistra.fr