

Computing Latinate Word Usage in Jane Austen's Novels

This article sponsored by



Mary Margolies
DeForest and Eric Johnson
believed that Austen varied
the density of Latinate words
in her characters' speeches

to stamp them as ignorant or educated, emotional or self-possessed, friendly or formal, honest or pretentious, lower class or upper class, female or male, and devised a mechanism to analyse this aspect of her work.

Over the years, as we have taught and reread the novels of Jane Austen, her characters never fail to come alive - the dialogue of each character always seems to be composed of appropriate diction. We wanted to measure one aspect of this diction: the density of Latinate words in collected speeches, letters and internal monologues of all Austen's characters.

JALATIN is a computer program we designed to identify and count the words of Latin origin in the texts of novels by Jane Austen. We used slightly modified versions of the files of the Oxford Electronic Text Library Edition of *The Complete Works of Jane Austen*. When JALATIN is executed, it first fills a table with the words and codes in JACODES.DAT, and it then asks the user for the name of a text file (an ASCII file with line ends) to be analyzed.

Distinguishing between Latinate and non-Latinate words proved difficult. Essentially, we included the criterion that a word must look Latinate to be classified as Latinate. We included all words from Greek, which mostly entered the English language from Latin authors, and excluded two types of words derived from classical languages: monosyllables—because Germanic words tend to be monosyllabic—and words that no longer look Latin:

```
art (O)
artist (L)
deceive (O)
deception (L)
polish (O)
polite (L)
```

Fig. 1. Some choices for Latinate (L) and non-Latinate (O) words.

Output of JALATIN for the text of Austen's *Emma* is presented in Figure 2. It can be seen that the code file (of all words in Austen novels) contains a little more than 36 percent Latinate words. By contrast, the total running words for *Emma* is only slightly more than 8 percent Latinate words. The difference is great since the most commonly repeated words in *Emma* (or in any novel) are not Latinate words.

```
Number of coded words: 13775

Word
Code  Number  Percent
O     8778   63.7241
L     4997   36.2759
Total running words in EM.ASC: 160861

Word
Type  Number  Percent
O    147583  91.7457
L    13278  8.2543
```

Fig. 2. The output of JALATIN showing the number and percentage of Latinate words in the code file and those in Austen's *Emma*.

Since the most commonly repeated words are function words and proper nouns, it may interest the researcher to exclude such words from the calculations. JALATIN, in fact, asks the user if a file of words should be excluded. We constructed a file (called JACOMMON.WDS) that contains 945 function words and proper nouns. When the file for *Emma* is processed and the words in JACOMMON.WDS are excluded from the counts, the percentage of Latinate words increases to over 21 percent (see Figure 3).

```
Number of coded words: 13775

Word
Code  Number  Percent
O     8778   63.7241
L     4997   36.2759

Total running words in EM.ASC: 62024
(Excluding 98837 occurrences of 945
 words in JACOMMON.WDS.)

Word
Type  Number  Percent
O    48831  78.7292
L    13193  21.2708
```

Fig. 3. Counts and percentages of types of words in *Emma* with 945 function words and proper nouns excluded.

JALATIN can process a series of files simultaneously, and it can sort the results based on the percentage of Latinate words. Figure 3 shows the output for the processing of seven files—each file contains the words of dialogue of a member of the Bennet family (in *Pride and Prejudice*). A list of words may or may not be excluded from the processing of multiple files (function words and proper nouns—found in JACOMMON.WDS—are excluded from the results in Figure 3). When all of the speakers in Austen's novels are ranked according to the percentage of Latinate diction, the full range of characters in her world can be seen (see Figure 4).

The Bennet family spans the range of female densities. It is no surprise that Mary, bookish, pretentious, and ostentatious, speaks with a significantly higher percentage of Latinate words than the Bennets, or, indeed, any character in Austen. Mary Bennet, like all the characters with higher densities than the narrators' (marked with an asterisk in Figure 4), is flawed. She and Mr. Collins are stupid

and pretentious; Sir Thomas Bertram and Lady Russell are pompous; Mr. Darcy is frigidly formal until he is reformed by Elizabeth; Mr. Elliot is a hypocrite. Mary Bennet's sister Elizabeth uses almost the same density of Latinate words as the narrator of *Pride and Prejudice*. In fact, the heroines we identify with are those whose levels are closest to the narrators, Elinor Dashwood, Emma, and Anne Elliot. Elizabeth, Jane, and Mr. Bennet are well educated (and thus use the second, third, and fourth most Latin words), but they do not use the artificial Latinate diction of Mary. A place at the lowest end of the Latinate spectrum indicates low intelligence, low birth, or both. Mrs. Bennet, Lydia, and Kitty do not speak like well-educated characters, and their dialogue shows significantly less Latinate use. The Steeles,

Mrs. Jennings, and the Thorpes are lower class. Sailors tend to have a low density—Admiral Croft has a lower density than Lydia Bennet—because they are presented as coming from the lower class and advancing through intelligence, courage, and energy. Sir Walter Elliot resented naval officers because they owed their high status to character not birth (PE, 19). Men have a higher average than women: a difference of 1.4%. That difference does not seem very great, but if ten Mary Bennets were added to each novel, like ten lead balloons, they would only bring the average density of women to 20.6% - still under the men's average density. Emotional stress brings down the Latinate levels. Elizabeth Bennet drops from 25% to 9% when she learns about Lydia's elopement. When she is in love with Willoughby, Marianne Dashwood slips almost seven percentage points. Letters often show an increase in density when the author is concerned with making a favourable impression. Frank Churchill's letter explaining why he deceived his father and everyone else uses a higher density than normal. Marianne Dashwood writing to Willoughby uses a much higher density, perhaps because she knows she shouldn't be writing to him. Lady Bertram, who 'rather shone in the epistolary line' (MP,425), writes about her son's accident in Latinate words, while he is away. When he returns and she sees him with her own eyes, her Latinate density drops dramatically from 28.9% to 7.1%.

JALATIN can be used to find passages of high Latinate density, from which we culled the following:

'The elegance, propriety, regularity, harmony - and perhaps, above all, the peace and tranquillity of Mansfield, were brought to her remembrance every hour of the day, by the prevalence of everything opposite to them here.' (MP 391)

Fanny Price is exiled to Portsmouth, to live with her parents in a squalid, noisy house. There she pines for her uncle's manor with its 'elegance, propriety, regularity, harmony.' The Latinate words perfectly convey the stateliness of the house (no doubt a fine, neo-classical structure) and the civilized behaviour of its inhabitants. When her cousins Julia and Maria Bertram competed bitterly for Henry Crawford, neither ever raised her voice or said anything rude. In Portsmouth,

'[E]very body was noisy, every voice was loud. Whatever was wanted, was halloo'd for, and the servants halloo'd out their excuses from the kitchen. The doors were in constant banging, the stairs were never at rest, nothing was done without a clatter.' (MP 392).

The Germanic words suggest chaos and squalor. The same antithesis can be found in the diction of Austen's characters, with high densities implying civilized values or the perversion of civilized values; and low densities indicating sincerity or stupidity.

Thus, JALATIN when applied to Austen's texts reveals her rhetorical goals and strategies. It could readily be modified to reveal the goals and strategies of other authors.

Mary Margolies DeForest

crypto@pop.ecentral.com

Eric Johnson

JohnsonE@jupiter.dsu.edu

College of Liberal Arts

Dakota State University

Character	Novel of	Number of Words	Latinate %
Mary Bennet	(PP)	160	33.8%
Mr. Collins	(PP)	1561	29.4%
Letters (62-3, 296-7)		442	33.7%
Sir Th. Bertram	(MP)	2828	29.0%
Mr. Darcy (to 195)	(PP)	954	28.8%
Letter (196-203)	(PP)	1019	34.8%
(from 251)	(PP)	624	22.1%
Mr. Elliot	(PE)	382	27.7%
Letter (203-4)	(PE)	71	12.7%
Elinor Dashwood	(SS)	4311	26.9%
*Narrator	(SS)	25942	26.4%
Lady Russell	(PE)	522	26.1%
*Narrator	(NA)	17491	25.5%
*Narrator	(PP)	23322	25.4%
Marianne D'wood (to 42)	(SS)	470	25.3%
In love (from 42)	(SS)	1373	18.4%
Letters (186-8)	(SS)	170	27.1%
Recovery (from 343)	(SS)	492	22.4%
Colonel Brandon	(SS)	1418	25.0%
Elizabeth Bennet	(PP)	6812	25.0%
Letters (150, 320, 382-3)180		23.9%	
(In shock (276-8)	(PP)	112	9.8%
Anne Elliot	(PE)	4118	24.4%
*Narrator	(PE)	18180	24.3%
Lady Catherine	(PP)	952	24.1%
George Wickham	(PP)	795	24.0%
*Narrator	(MP)	28708	24.0%
Mr. Gardiner	(PP)	318	23.9%
Letter (312-3)	(PP)	142	24.6%
*Narrator	(EM)	21992	23.8%
Mrs. Dashwood	(SS)	1384	23.7%
Henry Tilney	(NA)	2550	23.5%
Emma	(EM)	16117	23.4%
[97 lines omitted for reasons of space]			
Mrs. Ferrars	(SS)	32	6.3%
Mrs. Allen	(NA)	662	6.2%
Lydia Bennet	(PP)	699	6.3%
Letter (291-2, 380)	(PP)	115	5.2%
Dr. Grant	(MP)	66	4.5%
Kitty Bennet	(PP)	46	4.3%
Admiral Croft	(PE)	780	4.1%
Mr. Price	(MP)	178	3.4%
Proportion of L words for men:			
		10837/52390	20.7%
Proportion of L words for women:			
		18147/93918	19.3%

Fig.4. Extract from JALATIN analysis of character densities.