



Title	Corpus and Literary Criticism : An Analysis of Tess of the d'Urbervilles Using Concordancers
Author(s)	Matsuda, Masako
Citation	長崎大学総合環境研究. 2007, 10(1), p. 35-44
Issue Date	2007-11-01
URL	http://hdl.handle.net/10069/21489
Right	

This document is downloaded at: 2019-04-26T00:28:06Z

Corpus and Literary Criticism

An Analysis of *Tess of the d'Urbervilles* Using Concordancers

Masako Matsuda*

Abstract: Close reading is considered important in the criticism of English literature; however, in analyzing novels researchers should make a precise memo to record how and where certain words are used in the work. If we could use both corpus made up from electronic texts on the internet and a concordancer, it would make the above efforts much easier.

In this essay Thomas Hardy's *Tess of the d'Urbervilles* was examined by using three tools of WordSmith, a concordance software, in comparison with George Eliot's *Middlemarch* and Somerset Maugham's *Of Human Bondage*. WordSmith is a set of programs, which has three main tools called Wordlist, Keywords, and Concord. Wordlist as a tool makes a list of all the words in a text, set out in frequency order. With Keywords, the key words can be observed by comparing its wordlist with that of other reference text(s). Concord, a concordancer, can be used to plot any keyword or phrase in context to check what sort of wording it accompanies.

With these tools we could get a lot of linguistic information about the text, which leads us deeper understanding of the novel. Especially, Concord gives useful information on wording about keywords of a novel. How to select keywords is up to each researcher; however, concordance lines can help us to gain a clear view about a theme. In this way a concordancer is a useful tool for literary criticism.

Key Words: *literary criticism, concordancer, Tess of the d'Urbervilles, WordSmith, corpus*

1. Introduction

There has been a long tradition of close reading in the criticism of English literature since the 1920s, when 'practical criticism' was put into practice by I. A. Richards. Perhaps, it is true that literary criticism cannot have its valid standpoint without close reading of text, although its social and cultural context is important as well.

In poetry it is fairly easy to analyse words and

phrases precisely line by line; however, in analysing novels some important words are usually picked up according to a reader's choice and interest and make up a specific reading. Referring to this tendency, Eagleton argues that "we always interpret literary works to some extent in the light of our own concerns...." (Eagleton 1983)

A huge variety of words are used in each work; therefore, it is difficult to judge which selection of keywords is valid as representatives to organize its main theme in the novel. However, if corpus is consulted with a concordancer, would it be possible to avoid being arbitrary and get more objectivity in

* Faculty of Environmental Studies, Nagasaki University

Received 17 July 2007

Accepted 18 September 2007

analysing the language of novels? Could it lead to more or less impartial reading?

Nowadays, electronic texts of literary works are gaining popularity owing to massive efforts of such universities as University of Virginia and Illinois Benedictine College as well as Oxford University, so that it is possible to compile corpus easily making the most of these texts. In addition, concordancing softwares with their multi-functions are available to utilize corpus in different ways. By using some electronic texts and a software, this essay will explore possibilities to make use of corpus effectively in the analysis of novels.

2. Method and Theory

As an example of novels, Thomas Hardy's *Tess of the d'Urbervilles* (1891) (abbreviated as *Tess* in this essay) was chosen and its electronic text was downloaded from the site of Project Gutenberg in Illinois Benedictine College. To compare with Hardy's vocabulary in *Tess*, *Middlemarch* (1872) by George Eliot was collected from the University of Virginia's Ebook Library. *Middlemarch* was chosen because it is considered that there might be some kind of similarity between Hardy and Eliot. Hardy's earlier novel, *Far From the Madding Crowd*, had been presumed as Eliot's work when it was first published in 1874 anonymously.

In addition, Somerset Maugham's *Of Human Bondage* (1915) from Project Gutenberg was used as an example of the 20th century novels. It might show slight differences of the language and interests from the Victorian Age. These works were published more than 70 years ago; therefore, they can be used without infringing any copyright laws.

A concordance software, WordSmith, was available at the Oxford University Press. It is a set of programs, which has three main tools called Wordlist, Keywords, and Concord, for investigating how words are used in texts. According to its manual, the Wordlist tool is to make "a list of all the

words or word-clusters in a text, set out in alphabetical or frequency order." With Keywords the key words in a text can be observed by comparing its wordlist with that of other reference text(s). Finally, Concord, a concordancer, is to plot any keyword or phrase in context to check what sort of wording it accompanies.

Using Wordlist and Keywords, frequent keywords of *Tess* will be searched and after that, concordance lines for each word will be consulted to find how keywords are used throughout the novel for representing messages. WordSmith programmes are, Scott explains, to spot patterns and trends "through changing the shape of data." "By transforming a text into a list, or by plotting keywords in terms of where they crop up in their source texts, the human user will tend to see a pattern." (Scott 1998)

3. Results

3-1. Wordlist

The statistical wordlist (S) of *Tess* shows that its electronic text has 151,085 tokens, which means that it consists of the same number of words in total. However, many words are used repeatedly; therefore the wordlist also says that there are 12,152 types, which means that the text has the same number of different words. The average word length is 4.30, while there are 5,475 sentences, of which average length is 4.64.

Among the wordlist according to frequency (F) there are so many function words such as 'the,' 'and,' and 'of' on the top. Since these function words are not so important as keywords, content words i.e. nouns and verbs should be selected from the list between the first word and the 250th word.

3-1-1. Noun

Proper nouns in this section are ‘Tess,’ ‘Clare,’ ‘Angel,’ ‘d’Urberville,’ ‘Durbeyfield,’ ‘Marian,’ and ‘Izz.’

Tess	791 times
Clare	298
Angel	210

d’Urberville	162
Durbeyfield	121
Marian	102
Izz	99

Common nouns from the list of 250 frequent words in *Tess*, *Middlemarch*, and *Of Human Bondage* are as follows:

Figure 1

common nouns among frequent 250 words and frequencies

Tess		Middlemarch		Of Human Bondage	
tokens	151,085		1,948,705		260,980
types	12,152		321,908		12,313
time	276	man	646	time	395
day	226	time	394	room	378
man	225	way	392	day	377
mother	185	life	357	man	344
way	169	things	314	life	337
woman	150	money	264	eyes	299
eyes	148	men	252	things	298
house	133	day	238	heart	273
father	132	eyes	238	way	261
family	127	thing	235	people	239
life	126	people	233	thing	231
home	122	husband	231	money	198
door	109	room	215	bed	195
wife	107	world	187	night	186
morning	105	house	185	house	180
night	104	chapter	174	men	175
place	102	woman	169	boy	164
moment	97	part	162	morning	161
room	91	family	161	years	157
husband	89	morning	159	hands	155
light	89	father	158	days	152
heart	86	moment	158	door	147
words	78	marriage	157	uncle	146
girl	77			school	144
thing	76			moment	143
men	73			woman	135
things	73			tea	132
days	70			home	130
dairy	69			hair	129
				week	129

				doctor	128
				table	122

Comparing with the other two novels, there are more nouns related to family in *Tess*; ‘mother,’ ‘father,’ ‘family,’ ‘home,’ ‘wife,’ and ‘husband.’ Such words as ‘life,’ ‘money,’ and ‘people’ are common in *Middlemarch* and *Of Human Bondage* but they don’t appear in the list of *Tess*. This is due to their respective themes; a possibility of misalliance in *Tess*, while to pursue self-realization in life in the other novels.

3-1-2. Verb

Frequent verbs in *Tess* are ‘was,’ ‘be,’ ‘were,’ ‘said,’ ‘is,’ ‘been,’ ‘are,’ ‘come,’ ‘know,’ ‘being,’ and they are common with *Middlemarch* and *Of Human Bondage*. Verbs which are particularly

observed in *Tess* are ‘done,’ ‘stood,’ ‘heard,’ and ‘found.’

3-2. Keywords

The ‘Keywords’ program compares the words in a given text with a reference text, which usually consists of a larger corpus. Any word found to be outstanding in its frequency in the text is considered ‘key’ and these key words are listed in order of outstandingness.

When the text of *Tess* is compared with *Middlemarch* and *Of Human Bondage*, following keywords of common nouns and verbs are obtained:

Figure 2

Common nouns according to keyness		Verbs	
Tess vs Middlemarch	Tess vs Of Human Bondage	Tess vs Middlemarch	Tess vs Of Human Bondage
dairyman	dairy	murmured	being
dairy	dairyman	milking	is
cows	mother	stood	am
vale	vale	reached	milking
mother	cows	cried	continued
milk	miles	asked	having
village	family	grew	perceived
miles	spot	crept	beheld
road	village	passed	
lane	horse	beheld	
farm	farm		
day	father		
waggon	milk		
night	lane		
farmer	husband		
field	direction		
phase	gate		
valley	phase		

cottage	aspect		
colour	field		
folk	hedge		
door	brothers		
sun	waggon		
cart			
butter			
maids			
inn			
mead			

3-3. Concord

So far frequent words in *Tess* were searched out through Wordlist and Keyword Tools. The search results indicate that the frequently used common nouns are somewhat related to family and marriage, and that most of the key words are in connection with dairy farming. Considering the theme of *Tess* is the quest of the meaning of women's purity as a wife and the possibility of misalliance on the basis of romantic love with the background of dairy farming, the results are easily expected as a matter of course. This means that meaningful patterns in the text cannot be spotted automatically. Perhaps the possibilities of various

readings remain in existence here. Therefore, it is necessary to make clear standpoints of reading this novel or hypotheses, which guide to a specific interpretation before making a further investigation with Concord.

3-3-1. 'Wife' and 'husband'

Since the possibility of marriage between the middle class and the working class on the basis of romantic love is one of the themes in *Tess*, 'wife' and 'husband' among the frequent words are searched by the concordancer. Some examples with adjectives are as follows:

Figure 3

wife: 109 entries		husband: 98 entries	
a farmer's wife	3	my first husband	1
a deserted wife	2	her idolized husband	1
Angel's wife	2	The unbelieving husband	1
a practical farmer's wife	1	a too fond husband	1
a gentleman's wife	1	my ruined husband	1
his pretty wife	1	her now wedded husband	1
an unworthy wife	1	her fond husband	1
a virtuous wife	1	your wonderful husband's	1
his young wife	1		
this deserted wife of his	1		
a loving wife	1		
a properer wife	1		
the unbelieving wife	1		
the only sensible kind of wife	1		
his recently married wife	1		

To compare with 'husband,' 'wife' is described with more varieties of adjectives. To be 'pretty,' 'virtuous,' 'young,' and 'loving' are considered desirable qualities for a wife. Tess, a milkmaid in a dairy farm, is commended as a good wife to a farmer, whom Angel is going to be. Angel is trying to trespass the boundary of class; therefore, he is called as a "gentleman-farmer" twice in the novel.

As for Tess, she is represented more as 'd'Urbervilles' than 'Durbeyfield,' although this is only historical and not true any more in a practical sense. As is shown in the section of 3-1-1, 'D'Urbervilles' is used 162 times while 'Durbeyfield' 121 times, although the original d'Urberville family is extinct already. This gives the reader an impression that the marriage of Angel and Tess may not be so anti-social. Hardy's claim is that historically speaking, the present distinction between classes may be changeable in a long period of time. The novel insists that the framework of the class system might not hold validity from historical backgrounds, and its values are being

challenged by the protagonists' choice of spouses.

3-3-2. 'Pure' and 'Mercy Chant'

The Victorian family became the keystones of society. It is said that men "retreated to the bosom of their family as if to escape from the harsh world of moneymaking and the cash nexus." Home means "comfort, rest, peace, love, holiness" and religious metaphors were used in connection with family and home. As a result, housewives came to be worshipped as the 'Angel in the house,' and the most important attribute of a wife was purity in the 19th century (Himmelfarb 1996).

The subtitle of this novel, "A Pure Woman," is said to imply Hardy's protest against the Victorian morals and values (Lefebure 1993). 'Pure' is one of the keywords related to the important themes and the concordance lines are cited for 'pure. The characters of 'Mercy Chant,' Tess's rival for marriage with Angel, will be checked with the concordancer to find her profile.

PURE (26 entries)

1 were alike among them. Some approached pure blanching; some had a bluish pallor;
 2 who are true, and honest, and just, and pure, and lovely, and of good report--as
 3 and virtuous goes without question?" " pure and virtuous, of course, she is."
 4 stood on a basis approximating to one of pure reason, even if initiated by impulse
 5 its bitterness." And that she is pure and virtuous goes without
 8 country-house built for enjoyment pure and simple, with not an acre of
 10 and who had believed in her as pure! With an instinct as to
 11 to add," he said, "that for a pure and saintly woman you will not
 12 I have seen her, Angel. Since she is pure and chaste she would have been
 13 trustworthy as his song is breezy and pure, gets his authority for speaking of
 14 mouth thin. She looked absolutely pure. Nature, in her fantastic
 15 of the landscape; a field woman pure and simple, in winter guise;
 16 the d'Urbervilles A pure Woman Faithfully Presented
 17 "She is so good and simple and pure. O, Angel--I wish you would marry
 18 ou think that a young woman equally pure and virtuous as Miss Chant, but
 21 pectedness of things. Nothing so pure, so sweet, so virginal as Tess
 23 the village had to be kept pure. So on this, the first Lady-Day on

24 The Froom waters were clear as the pure River of Life shown to the Evangelist
(There are 26 citations in total; however, lines with the name of the village pub, The Pure
Drop, are omitted)

CHANT (11 entries)

1 least exhibit a face such as Mercy Chant could not show. But it was done with
2 oposed to and been accepted by Mercy Chant. Clare crumpled up the paper,
3 way back he encountered Miss Mercy Chant by the church, from whose walls she
4 excite our sympathies," said Miss Chant. "Yes, it must have been, for they
5 -in feeling and nature." "Mercy Chant is of a very good family."
6 in. It is true that my neighbour Chant's daughter had lately caught up the
7 -minded friend and neighbour, Dr Chant--" "But ought she not
8 equally pure and virtuous as Miss Chant, but one who, in place of that lady's
10 of them remarked, "There is Mercy Chant. Let us overtake her." Tess knew
11 The young lady was Miss Mercy Chant, the only daughter of his father's

The phrase of 'pure and virtuous' in 3, 5, 18 makes up a collocation and 'pure and chaste' in 12 are almost same in meaning. Although 'pure and simple' in 8 and 15 are examples of another collocation, it refers to people and things in the countryside. From the concordance lines, it is recognisable that there is a group of synonyms for 'pure' in the context of the novel; 'true,' 'honest,' 'just,' 'lovely,' 'of good report,' 'virtuous,' 'saintly,' 'chaste,' 'sweet,' and 'virginal.' Some of them, 'true,' 'honest,' 'just' denote moral values; 'lovely' and 'sweet' means desirable features and characters for women. 'Of good report,' 'virtuous,' 'saintly,' 'chaste,' and 'virginal' have something to do with female sexual behaviours.

Mercy Chant is regarded as 'pure and virtuous,' and in addition, since she is of a very

good family, she is supposed a most desirable candidate of a good wife from the standpoint of those days. Nevertheless, Tess is more beautiful and sexually attractive as a partner of romantic love.

3-3-3. 'Pleasure,' 'Sex,' and 'Mouth'

Today most of the readers assume that Tess is "a symbol of female sexuality" (Lefebure 1993). Her sexual attractiveness is elaborately represented with the narrator's keen interest and gaze in such scenes as her visit to the d'Urbervilles, the seduction in the Chase wood, and Tess and Angel's romance at the Crick's dairy farm. A search of words will be made to spot characteristics in one of the earliest examples of female sexual representation.

PLEASURE (25 entries)

3 automatic tendency to find sweet pleasure somewhere
4 to snatch ripe pleasure before the iron teeth of pain
8 pleasure girdled about with pain
17 To the aesthetic, sensuous, pagan pleasure in natural life and lush womanhood
19 positive pleasure and positive pain.
25 take your fill of pleasure on earth by making the life

Here 'pleasure' is modified with 'sweet,' 'ripe,' 'aesthetic,' 'sensuous,' and associated with 'pain,' 'pagan,' and 'on earth.' Although it is not compatible with Christian ethics, still it is claimed

that people have '(t)he irresistible, universal, automatic tendency' to find pleasure.

SEX (14 entries)

1	most spiritual beauty bespeaks itself flesh; and	sex takes the outside place in the presentation
4	a visionary essence of woman—a whole	sex condensed into one typical form
6	faith to him the ordinary sensibilities of	sex seemed a flippancy
12	each was but portion of one organism called	sex. There was so much frankness
13	the elusive quality of her	sex which attracts men in general

According to the narrator, sex consists of a huge organism using each person as its portion. S/he tries to heighten the value of 'sex,' claiming that it is the presentation of the most spiritual beauty in flesh. At the same time, Tess is idealized as 'a visionary essence of woman,' and 'a whole sex condensed into one typical form' in her.

His look and aesthetic judgements become dominant over the story because moral values and female beauty are bound together.

Whenever the narrator mentions 'sex,' s/he automatically tries to indicate female sexuality. Perhaps, this implies that the narrator should be male. He does not care about men's sexuality at all, while his gaze is always fixed on Tess's face, body, and movements. He asserts that the spiritual beauty of women, which means 'purity' in those days, takes another form in sexual attractiveness.

Mulvey argues that "(i)n a world ordered by sexual imbalance, pleasure in looking has been split between active/male and passive/female. The determining male gaze projects its fantasy onto the female figure which is styled accordingly" (1975). Although she mentions this regarding the cinema, it can be applied to Tess, where the male narrator's gaze plays an important role in aesthetic and moral evaluation. Therefore, it is possible to develop literary criticism in this line, "woman as image, man as bearer of the look." (Mulvey 1975)

MOUTH (43 entries)

3	the usually ripe red	mouth
5	her ripe	mouth
7	it was in her	mouth that this culminated.
8	a beautiful	mouth and figure:
17	at the unpracticed	mouth and lips,
20	a maddening	mouth
21	The maiden's	mouth
22	too tempting	mouth,
23	the curves of her	mouth
24	peony	mouth
25	that pretty red	mouth
35	red interior of her	mouth as if it had been a snake's.
37	pouted-up deep red	mouth

38 her flower-like mouth and large tender eyes
 39 her mouth he had seen nothing to equal
 42 thy cherry mouth!

'Mouth' is an important word along with 'figure' (41 entries) to represent Tess's sexual beauty. Her 'red,' 'ripe,' 'beautiful,' and 'maiden's,' mouth is compared with 'peony,' 'flower-like,' and 'cherry,' and consequently it is 'maddening' and 'tempting.' As it is compared with a snake's mouth, it is possible to find the imagery of female sexuality in relation to the Biblical metaphors.

4. Discussion

The general idea of frequently used words in the novel can be grasped with Wordlist and Keyword programmes. Although the results are generally useful, some meaningful patterns in the text cannot be spotted automatically. Perhaps the possibilities of various interpretations remain in existence here. A concordancer is convenient to investigate what vocabulary, language patterns, symbols, similes, and metaphors are used in the whole text, but it is necessary to make clear standpoints of reading before a search.

For example, there are some possibilities of reading *Tess*; (1) admiring the narrator's description of female sexuality, (2) a criticism of the narrator's view of "(w)oman as image, man as bearer of the look," (3) an analysis of the narrative structure with a legend and ballads, and (4) a historical approach analysing the making of the modern family and changing attitudes toward marriage in the Victorian Age. Among these potentialities, perhaps the reader will choose one in the light of their own concerns as Eagleton says. From the study with WordSmith, it is difficult to find a single reading which is surmised objectively correct.

If this is the case, how are certain works of novels chosen to be excellent as a canon? For *Tess*, the modern reader is generally attracted by the

representation of female sexuality. G. Eliot's *Adam Bede* (1859) deals with the same themes of seduction, murder, and execution as *Tess*'s, but Eliot's narrator supports the family and moral values while Hardy's is sympathetic to the heroine and greatly fascinated with her physical beauty.

Around the end of the 19th century it is a revolutionary attitude to choose one's wife on the basis of romantic love and sexual attraction. The lower classes tended to value marriage of romantic love since they were "free to pursue individual rather than family objectives." On the contrary, the middle class clung to "the whole array of anti-erotic, communitarian personal values of traditional society," because the bourgeois had property to transmit to following generations.

Shorter argued that "(s)elf-realization — accomplished through sexual gratification — has taken command of courtship," and that "(w)ith time, the romanticization of courtship survived farther and farther into married life, until finally the couple would expect to behave affectionately and to be sexually active until the end of its days." Thus, romantic love with a strong component of eroticism became the most important factor in marriage in the 20th century. As a consequence, *Tess* might attract more attention of the readers than *Adam Bede*. Reading and evaluation of each work is based on the concern of individuals affected by the trend of the day, although Bakhtin's argument is also true that to force a single interpretation to the reader is impossible, since it is generally understood that there are various discourses even within a novel (Lodge 1990).

5. Conclusion

WordSmith is a good device for helping the reader to analyse and evaluate literary works.

Usually novels are fairly long and consist of a large number of words; therefore, to see language patterns and behaviours of particular linguistic forms in the whole text gives the reader insights and evidences in developing her/his own interpretation.

In teaching English as a second language, literary works are important as they could sometimes grip the learners' imagination so tightly that they would accelerate the speed of reading. They are good examples of authentic language material, and investigation with a concordance software is a good chance for learners to know collocations, authors' favourite phrases, and some expressions, which are comprised of particular themes of each work, for further study, analysis, and appreciation.

6. References

- Eagleton, T., 1983 (1st ed.), 1996, *Literary Theory: An Introduction*, 2nd ed. The University of Minnesota Press, Minneapolis.
- Eliot, G., 1859, *Adam Bede*, in S. Gill (eds), 1980, Penguin, London.
- Hardy, T., 1891, *Tess of the d'Urbervilles: A Pure Woman*, in D. Skilton (eds), 1978, Penguin, Harmondsworth.
- Himmelfarb, G., 1994, *The De-moralization of Society: From Victorian Virtues to Modern Values*, Vintage Books, New York.
- Lefebure, M., 1996, *Thomas Hardy's World*, Carlton, London.
- Lodge, D., 1990, *After Bakhtin: Essays of Fiction and Criticism*, Routledge, London.
- Mulvey, L., 1975, 'Visual Pleasure and Narrative Cinema,' *Screen*, vol. 16, no.3, in *Visual Culture: the Reader*, J. Evans and S. Hall (eds), 1999, Sage Publications, London.
- Shorter, E. 1975, *The Making of the Modern Family*, Basic Books, New York.

(Homepages)

Project Gutenberg in Illinois Benedictine College:

<http://www.gutenberg.net/>

Scott, M., 1998, *Manual to WordSmith*, Oxford University Press, Oxford:

<http://www.oup.com/elt/global/catalogue/multimedia/wordsmithtools3/>

The Oxford Text Archive: <http://ota.ahds.ac.uk/>

University of Virginia's Ebook Library:
<http://etext.lib.virginia.edu/ebooks/ebooklist.html/>