



## Poetic Language and the Computer: A Corpus Stylistic Study of *The Waste Land*

Eman Adil Jaafar

Department of English  
College of Education for Women  
University of Baghdad

[eman\\_jafer@coeduw.uobaghdad.edu.iq](mailto:eman_jafer@coeduw.uobaghdad.edu.iq)

DOI: <https://doi.org/10.36892/ijlls.v3i1.525>

**Received:**  
25/02/2021

**Accepted:**  
10/03/2021

**Keywords:**

Corpus Stylistics,  
poetry, WebCorp  
live, Wmatrix

**Abstract**

This study aims at analyzing one of the significant poems of the twentieth century, *The Waste Land* by T.S. Eliot. By means of applying the tools of the computer, namely; Wmatrix (Rayson 2009) and WebCorp Live (Birmingham City University). This paper seeks 1. To examine whether corpus stylistics can be helpful in analyzing a single poem 2. To verify the importance of corpus tools in interpreting poetic language. Moreover, this study attempts to examine key semantic domains, keywords, and concordances in the poem. This study proves that corpus tools are crucial in matters of saving time, reaching to accurate results and achieving much more objectivity than applying only the qualitative method in analyzing the data. Thus, it is recommended to integrate both methodologies (qualitative and quantitative) in the study of poetic language.

### 1. INTRODUCTION

This paper presents a *corpus-stylistic analysis* of T.S.Eliot's poem *The Waste Land*, henceforth (TWL). The poem will be analyzed according to the corpus stylistic approach. Baker, Hardie, and McEnery (2006, p.48-49) define the word *corpus* (plural "corpora") is a Latin word that refers to a combination of written and spoken texts. The word corpus also has the meaning of "body". The texts are generally stored and analyzed electronically. Corpus stylistic starting to appear in the discipline of linguistics has become increasingly popular in recent years. It is a combined process to the study of the language of literature (for more details about corpus stylistics see Ho, 2011; Mahlberg, 2014).

Hunston (2002, p. 2-3) pinpoints that "Although a corpus does not contain new information about language by using software packages which process data, we can obtain a new perspective on the familiar". To put it differently, corpora reflect the real authentic language and by using quantitative as well as qualitative methodology in terms of integrating manual and corpus analysis one can get interesting facts about patterns of language.

Bujanova (2013) sees corpus stylistics as "a way of bringing the study of language and literature close together"(p.3). Furthermore, Mahlberg (2014) maintains that corpus stylistics joins principles from corpus linguistics and literary stylistics. Though, subsequent a corpus stylistic process be able to help to mark the uses of the new linguistic of the language of texts by

quantitative methods (p.378-380). Moreover, Wynne (2005) pointed out some underlying similarities of approach: "Traditional and computational forms of stylistics have more in common than is obvious at first sight. Both rely on the close analysis of texts, and both benefit from opportunities for comparison" (p.677-679).

Mahlberg & McIntyre (2011, p.205) ascertain that corpus stylistics does not substitute the qualitative analyses, but rather foster the computational analyses. Furthermore, they argued rightly, that "corpus methods can aid the identification of elements of a text worthy of further qualitative analysis" (p.223). In this regard, computer programs can aid researchers immensely as Rayson (2003, p.4) pinpoints that "Although the computer saves us time with its processing of the texts into frequency lists, it presents us with so much information that we need a filtering mechanism to pick out significant items before the analysis can proceed".

## 2. METHODOLOGY

The following methodological steps are followed to conduct this study:

- 1-Preparing the corpus of the selected text and converting it into a plain text form (txt).
- 2-Using WebCorp Live as the tool to examine Keywords by focusing on content words. This tool enables researchers from isolating function words like (articles, prepositions, pronouns ...etc) in order to give the needed attention to the words that carry meaning.
- 3-Using Wmatrix software to benefit from its unique advantage of identifying key semantic domains.

## 3. RELATED WORK ON CORPUS STYLISTICS AND POETRY

Brooke, Hammond, & Hirst (2015) conducted research seeking distinguished voices in *The Waste Land*. Their major aim is to find out whether the computer can locate different voices as in human analysis (p.4). Similar to the current study, Jaafar (2019) analyzed Seamus Heaney's poem "A Herbal" by means of computational tools. It is concluded that the integration of both quantitative and qualitative approaches can be fruitful in terms of obtained objective results. In contrast, Jaafar (2014) conducted a manual qualitative study to analyse selected poems that focuses on deviation and other stylistic tools.

## 4. TYPES OF CORPUS TOOLS

In fact, there are a variety of corpus tools or software programs that can be useful in conducting corpus stylistic research. Researchers should have an idea about such tools and then they can select the suitable one for their analysis. Some of these toolkits, for example, **WordSmith** tools which is a word list tool, a concordance tool, and a keyword list tool. Another web-based corpus software **Wmatrix** ( Rayson 2009) which is distinguished from other tools by its unique feature of identifying key semantic domains as well as having the function to automatic POS tagging. **AntConc** (Anthony 2004) is free software that can be used freely by anyone who has an interest in finding n-grams, keywords, concordances, and collocations. Moreover, Mahlberg & Smith's (2012) tool **CLiC** (Corpus Linguistics in Cheshire). One more beneficial tool is **WebCorp Live** is a search engine for linguists, teachers, and learners that is useful to retrieve corpus data online.

## 5. THE DATA

*The Wasteland* is considered T.S.Eliot's masterpiece written in 1922. It is one of the complicated poems of the twentieth century. It depicts the destructed life after the First World War. Not one this, but also it shows the decline of moral values of the society after the war. It is an obscure poem due to the inclusion of many allusions and references from other texts. It was written he underwent personal difficulties in his first marriage. Its complexity made it one of the crucial and significant works of the English literature and in particular of the modern age, which contain some lines from a variety of languages such as German, French, Spanish and Hindi. ( see, North , 2001; McHale & Stevenson 2006).

*The Waste Land* is a long poem consists of a total 3,028 words and divided into five sections. The main ideas and themes of the poem include; death, corruption and many other pessimistic views of the life postwar. It is considered the mouthpiece of that time. However, the last line of the poems shows a different shift. The Sanskrit word 'Shanith' means 'peace' is repeated three times.

The first section "The Burial of the Dead" contains four different characters each one has a story that is not completed. The language of the poem contains some verses written in both German and French which add difficulty in understanding and interpreting the text.

Section two "A Game of Chess" narrates stories of women of different social classes the high and the lower class. These women have their own dilemma. This section consists of 261 words. Section Three *The Fire Sermon* is the longest section of the poem. Section four *Death by Water*. This the shortest section of the poem. Section five "What the Thunder Said" is pessimistic and shows a harsh reality in terms of the imagery of destruction, there is a temporal renewal but again the destruction is near and final.

## 6. The Analysis

Roughly speaking, under the umbrella of qualitative analysis, the poem depicts boredom and monotony of the modern life, which illustrated by means of continuous repetitions of words, phrases, and sentences throughout the text, examples include:

“My nerves are bad tonight. Yes, bad. Stay with me. “Speak to me. Why do you never speak. Speak.

“What are you thinking of? What thinking? What?“I never know what you are thinking. Think.”

I think we are in rats' alley

Where the dead men lost their bones.

“What is that noise?“  
The wind under the door.

“What is that noise now? What is the wind doing?”

Nothing again nothing.

“Do

“You know nothing? Do you see nothing? Do you remember

“Nothing?” ( *Section II A Game of Chess*)

<https://www.poetryfoundation.org/poems/47311/the-waste-land>

These lines give a clear idea about the nothingness and the hollowness of the people of the modern life.

In addition to that, T.S. Eliot uses allusions and includes many images and references to famous works and personalities in the text. He creates ideas and leaves them unfinished which in turns gives the complexity of the poem. Another important aspect of the poem is the inclusion of many voices or characters by using personal pronouns like ( she, he, I , you). This can also be the reason behind the difficulty of the poem.

The poem consists of a total 3,028 words. According to WebCorp Live calculation, the number of content words is 1034 and the total number with repetitions throughout the text is 1552.

Count: 1034 Total: 1552

---

**Generate Wordlist**

Filter Out Stopwords:  ⓘ

Case Insensitive:  ⓘ

Ngram Size: 1 word ⓘ

Minimum Frequency: 1 ⓘ

**Reset** **Submit**

Screenshot (1) Wordlist in *The Wasteland*

Screenshot (1) illustrates that by copying and pasting the text in the specific area the generator of the wordlist count the words according to the choices made by the researcher, for example, the n-gram size ( single words) and the option filter out stop words or high-frequency words( skipping function words from the retrieval results).

In the first section of the text *The Burial of the Dead*, contains (252) words repeated once one and excluding high-frequency words like (articles and pronouns and other function words). The most repeated word is 'Dead' which is mentioned five times only in this part.

10 occurrences.		Ext
MIGLIOR FABBRIO I. The Burial of the	Dead	1
April is the cruellest month , breed	land , mixing Memory and desire , st	2
month , breeding Lilacs out of the	dead tree gives no shelter , the cricket	3
ges , where the sun beats , And the	dead , and I knew nothing , Looking into	4
s failed , I was neither Living nor	dead sound on the final stroke of nine .	5
Mary Woolnoth kept the hours With a	dead men lost their bones . What is that	6
hink we are in rats alley Where the	dead . ) Bestows one final patronising ki	7
And walked among the lowest of the	dead , Forgot the cry of gulls , and the	8
hlebas the Phoenician , a fortnight	dead We who were living are now dying Wit	9
mountains He who was living is now	Dead mountain mouth of carious teeth that	10
re were only water amongst the rock		

Screenshot (2) the Retrieval Results of the Word 'Dead'

The word 'dead' for example, occurs (10) times throughout the text. Here are some examples:

1. Dead April.
2. Dead land.
3. Dead tree.
4. Dead sound.
5. Dead mountain.

These words mentioned with the word 'dead' to give them a kind of pessimistic personification, related to the pessimistic atmosphere of the poem.

In addition to that, part one contains some verses written in both German and French.

**6.1. Keywords Analysis**

Wmatrix is employed to identify the main keywords in the *TWL*. The selected reference corpora is BNC which contain 100 million words, 90 % percent consists of written data the rest 10% percent of spoken data. Here, is this paper The BNC sampler (1 million words) consists of many categories. The BNC Written Imaginative sampler is used as a reference corpus. This sampler contains literary works and creative writing. It is necessary to use a reference corpus that suits the content of the text to be analyzed and examined (cf Culpeper, 2009, McIntyre 2010).

	Item	O1	#1	O2	#2	LL	%DIFF
1	Concordance c	11	0.38	10	0.00 +	€6.94	3329.58
2	Concordance rock	11	0.38	13	0.01 +	€2.98	5384.29
3	Concordance mountains	8	0.28	4	0.00 +	€4.46	13226.52
4	Concordance jug	8	0.28	5	0.00 +	€2.44	12161.21
5	Concordance water	15	0.52	73	0.03 +	€2.09	1474.64
6	Concordance hurry_up	5	0.17	0	0.00 +	43.52	172176288184795136.00
7	Concordance goonight	5	0.17	0	0.00 +	43.52	172176288184795136.00
8	Concordance violet	4	0.14	0	0.00 +	34.82	137741035701794864.00
9	Concordance nt	4	0.14	0	0.00 +	34.82	137741035701794864.00
10	Concordance co	4	0.14	0	0.00 +	34.82	137741035701794864.00
11	Concordance cant	4	0.14	0	0.00 +	34.82	137741035701794864.00
12	Concordance Albert	4	0.14	1	0.00 +	29.84	30553.03
13	Concordance burning	5	0.17	8	0.00 +	26.40	4689.54
14	Concordance under	10	0.34	30	0.04 +	26.32	857.91
15	Concordance bones	5	0.17	9	0.00 +	25.50	4157.37
16	Concordance la	4	0.14	3	0.00 +	25.33	10117.68
17	Concordance departed	4	0.14	3	0.00 +	25.33	10117.68
18	Concordance was	11	0.38	2823	1.27 -	24.85	-70.14
19	Concordance drop	5	0.17	11	0.00 +	23.93	3383.30
20	Concordance shadow	4	0.14	4	0.00 +	23.83	7563.26
21	Concordance dead	10	0.34	105	0.05 +	21.81	629.83
22	Concordance unreal	3	0.10	1	0.00 +	21.64	22889.77
23	Concordance shantih	3	0.10	1	0.00 +	21.64	22889.77
24	Concordance falling_down	3	0.10	1	0.00 +	21.64	22889.77
25	Concordance Tiresias	3	0.10	1	0.00 +	21.64	22889.77
26	Concordance thunder	4	0.14	6	0.00 +	21.51	3008.84
27	Concordance dry	6	0.21	31	0.01 +	20.23	1383.21
28	Concordance it	15	0.52	3031	1.36 -	19.82	-62.08
29	Concordance to	31	1.07	4802	2.16 -	19.48	-50.53
30	Concordance towers	3	0.10	2	0.00 +	19.43	11394.89
31	Concordance spring	4	0.14	9	0.00 +	19.00	3305.89
32	Concordance weialala	2	0.07	0	0.00 +	17.41	68870517850898432.00

33	Concordance	throbbing	2	0.07	0	0.00 +	17.41	68870517850898432.00
34	Concordance	thats	2	0.07	0	0.00 +	17.41	68870517850898432.00
35	Concordance	rudely	2	0.07	0	0.00 +	17.41	68870517850898432.00
36	Concordance	rico	2	0.07	0	0.00 +	17.41	68870517850898432.00
37	Concordance	pluckest	2	0.07	0	0.00 +	17.41	68870517850898432.00
38	Concordance	nymphs	2	0.07	0	0.00 +	17.41	68870517850898432.00
39	Concordance	mon	2	0.07	0	0.00 +	17.41	68870517850898432.00
40	Concordance	leialala	2	0.07	0	0.00 +	17.41	68870517850898432.00
41	Concordance	leia	2	0.07	0	0.00 +	17.41	68870517850898432.00
42	Concordance	hyacinth	2	0.07	0	0.00 +	17.41	68870517850898432.00
43	Concordance	clutch	2	0.07	0	0.00 +	17.41	68870517850898432.00
44	Concordance	barges	2	0.07	0	0.00 +	17.41	68870517850898432.00
45	Concordance	Wallala	2	0.07	0	0.00 +	17.41	68870517850898432.00
46	Concordance	Sweet_Thames	2	0.07	0	0.00 +	17.41	68870517850898432.00
47	Concordance	Mrs._Porter	2	0.07	0	0.00 +	17.41	68870517850898432.00
48	Concordance	London_bridge	2	0.07	0	0.00 +	17.41	68870517850898432.00
49	Concordance	broken	5	0.17	25	0.01 +	17.13	1432.65
50	Concordance	only	14	0.48	295	0.13 +	15.51	263.68
51	Concordance	upon	7	0.24	72	0.03 +	15.50	645.04
52	Concordance	red	6	0.21	52	0.02 +	14.99	784.22
53	Concordance	twit	3	0.10	6	0.00 +	14.81	3731.63
54	Concordance	wrinkled	2	0.07	1	0.00 +	13.61	15226.52
55	Concordance	stony	2	0.07	1	0.00 +	13.61	15226.52
56	Concordance	stirring	2	0.07	1	0.00 +	13.61	15226.52
57	Concordance	phoenician	2	0.07	1	0.00 +	13.61	15226.52
58	Concordance	phlebas	2	0.07	1	0.00 +	13.61	15226.52
59	Concordance	of_a_winter	2	0.07	1	0.00 +	13.61	15226.52
60	Concordance	hooded	2	0.07	1	0.00 +	13.61	15226.52
61	Concordance	gaily	2	0.07	1	0.00 +	13.61	15226.52
62	Concordance	divan	2	0.07	1	0.00 +	13.61	15226.52
63	Concordance	Richmond	2	0.07	1	0.00 +	13.61	15226.52

### Screenshot (3) Keywords by Wmatrix

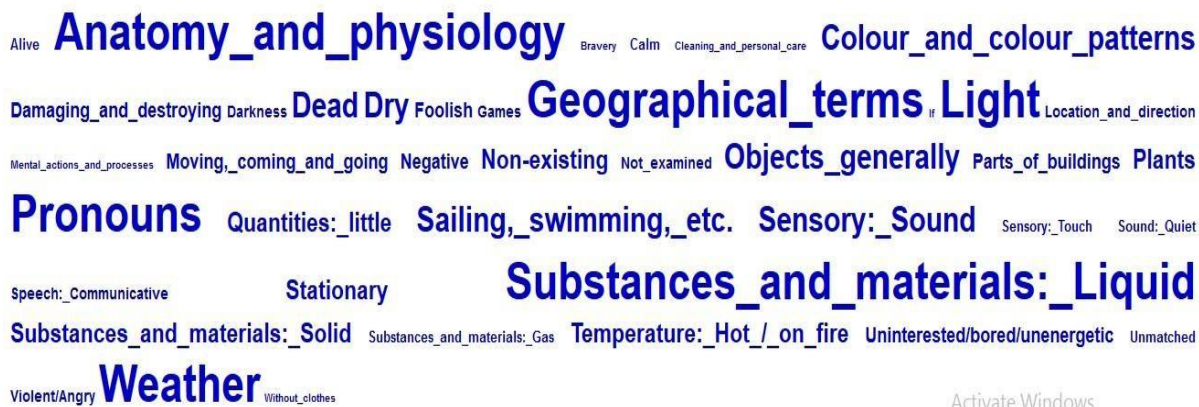
Screenshot 3 shows keyness results retrieved by Wmatrix.

#### 6.2.Key Semantic Domains

Semantic domains enable readers/researchers locate semantic categories of words. Wmatrix has this unique tool of gathering words under suitable domains. In this way, it facilitates examining the dominant group of words to understand thematic ideas.

Key domain cloud

Larger items are more significant.  
Underused items are shown in italics.  
Move your mouse over each item to show extra information in a tooltip.  
Click on a word to show the concordance.



Screenshot (4) Key semantic domains in *The Waste Land*

Screenshot (4) shows the most significant key semantic domains in the poem. Words with larger font have more importance than words with small fonts. Key domains help to give readers an idea of the main points and ideas proposed by the poet.

The most prominent semantic domains in the poem are:

**Table (1)** *Key semantic domains in The Waste Land*

No	Semantic domains	Words within these domains from <i>The Wasteland</i>
.1	Anatomy and psychology	Eyes, ears, finger, heart, feet, breasts, hands, bones, faces, belly, hair...
.2	Colour and colour patterns	Shadow, white, black, golden, glitter, red, brown, violet
.3	Dead	Burial, drowned, death, corpus, departed, obituaries, dying
.4	Geographical terms	Rivers, land, canal, sea, hill, desert, earth, mountain, shore, sandy, plains
.5	Light	Sunlight, shone, moonlight, rays, candle-flames, light, torchlight
.6	Substances, materials and liquids	Water, drop, drip, oil, tar, damp
.7	Weather	Rain, thunder, fog, wind, snow, clouds
.8	Pronouns	He, she, they, you, I, his, her, their, my, we, your, it, yourself, its

65 occurrences.			Extend context
from the Hyacinth garden , Your	arms	full , and your hair wet , I coul	1 More   Full
rden , Your arms full , and your	hair	wet , I could not Speak , and my	2 More   Full
wet , I could not Speak , and my	eyes	failed , I was neither Living nor	3 More   Full
knew nothing , Looking into the	heart	of light , the silence . Oed und	4 More   Full
( Those are pearls that were his	eyes	. Look ! ) Here is Belladonna , t	5 More   Full
here the Wheel , And here is the	one-eyed	merchant , and this card , Which	6 More   Full
, is something he carries on his	back	, Which I am forbidden to see . I	7 More   Full
hs , short and infrequent , were	exhaled	, And each man fixed his eyes bef	8 More   Full
exhaled , And each man fixed his	eyes	before his feet . Flowed up the h	9 More   Full
ch man fixed his eyes before his	feet	. Flowed up the hill and down Kin	10 More   Full
don peeped out ( Another hid his	eyes	behind his wing ) Doubled the fla	11 More   Full
world pursues , Jug Jug to dirty	ears	. And other withered stumps of ti	12 More   Full
irelight , under the brush , her	hair	Spread out in fiery points Glowed	13 More   Full
hen would be savagely still . My	nerves	are bad tonight . Yes , bad . Sta	14 More   Full
ey Where the dead men lost their	bones	. What is that noise ? The wind u	15 More   Full
r Those are pearls that were his	eyes	. Are you alive , or not ? Is the	16 More   Full
am , and walk the street With my	hair	down , so . What shall we do tomo	17 More   Full
game of chess , Pressing lidless	eyes	and waiting for a knock upon the	18 More   Full
he gave you To get yourself some	teeth	. He did , I was there . You have	19 More   Full
rivers tent is broken : the last	fingers	of leaf Clutch and sink into the	20 More   Full
eak not loud or long . But at my	back	in a cold blast I hear The rattle	21 More   Full
d blast I hear The rattle of the	bones	, and chuckle spread from ear to	22 More   Full
bones , and chuckle spread from	ear	to ear . A rat crept softly throu	23 More   Full
, and chuckle spread from ear to	ear	. A rat crept softly through the	24 More   Full
he vegetation Dragging its slimy	belly	on the bank While I was fishing i	25 More   Full
fathers death before him . White	bodies	naked on the low damp ground And	26 More   Full
naked on the low damp ground And	bones	cast in a little low dry garret ,	27 More   Full
dry garret , Rattled by the rats	foot	only , year to year . But at my b	28 More   Full
only , year to year . But at my	back	from time to time I hear The soun	29 More   Full

### Screenshot (5) Retrieval Results of the Words under the Domain of Anatomy and Psychology

## 7. Conclusion

Analyzing literary text manually (qualitatively) has a noticeable significance, but analyzing text by integrating both qualitative and quantitative approaches lead to much more significant and accurate results. The numbers and calculations retrieved by the tools of the computer enable researchers to focus on details that can be missed by manual analysis. It is, thus, crucial to merge both methodologies to reach an accurate and objective interpretation of any literary text.



## REFERENCES

- Anthony, L. (2004). AntConc: A learner and classroom-friendly, multi-platform corpus analysis toolkit. *proceedings of IWLeL*, 7-13.
- Baker, P., Haredie, A., & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburgh, Edinburgh University Press.
- Brooke, J., Hammond, A., & Hirst, G. (2015). Distinguishing Voices in The Waste Land using Computational Statistics. *LiLT (Linguistic Issues in Language Technology)*, 12.
- Bujanova, K. (2013). *A corpus-stylistic analysis of Mitchell's gone with the wind and Hemingway's farewell to arms*. (Master's thesis). Retrieved from <http://www.duo.no/handle/10852/37048.pdf>
- Burke, M. (2014). *The Routledge handbook of stylistics*. Abingdon, Utrecht University.
- Culpeper, J. (2009). Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet. *International Journal of Corpus Linguistics*, 14(1), 29-59.
- Ho, Y. (2011). *Corpus stylistics in principle and practice: A stylistic exploration of John Fowles' The Magus*. A & C Black.
- Hunston, S. (2002) *Corpora in applied linguistics*, Cambridge: Cambridge University Press.
- Jaafar, E. (2014). "A Stylistic Analysis of Two Selected Poems". *Journal of the College of Education for Women*, Vol. 25, no. 1, <https://jcoeduw.uobaghdad.edu.iq/index.php/journal/article/view/722>.
- Jaafar, E. (2019). Examining Seamus Heaney's Poem "": A Corpus Stylistic Approach. *Al Ustath Journal for Human and Social Sciences*, 58(1), 17-26. Retrieved from <https://alustath.uobaghdad.edu.iq/index.php/UJIRCO/article/view/826>
- Mahlberg, M., & Smith, C. (2012). Dickens, the suspended quotation and the corpus. *Language and Literature*, 21(1), 51-65.
- Mahlberg, M. (2014). Corpus stylistic. In M. Burke (Ed), *The Routledge handbook of stylistics* (pp.378-380). Abingdon: Routledge.
- Mahlberg, M., & McIntyre, D. (2011). *A case for corpus stylistics: Ian Fleming's Casino Royale*. *English Text Construction*, 4(2), 204-227.
- McIntyre, D., & Archer, D. (2010). A corpus-based approach to mind style. *Journal of Literary Semantics*, 39(2), 167-182.
- North, M. (2001). *The Waste Land: A Norton Critical Edition*. Ed. Michael North. NY: Norton.
- McHale B. & Stevenson R. (Eds.) (2006). *Edinburgh Companion to Twentieth Century Literatures in English*. Edinburgh: Edinburgh University Press.

**Poetic Language and the Computer: A Corpus Stylistic Study of *The WasteLand***

- Rayson, P. (2003). *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison* (Doctoral dissertation, Lancaster University).
- Rayson, P. (2008). From keywords to key semantic domains, *International Journal of Corpus Linguistics*, 13 (4): 519–549.
- Rayson, P. (2009). *Wmatrix: A web-Based Corpus Processing Environment*. School of computing and communications, Lancaster University  
<http://ucrel.lancs.ac.uk/wmatrix/>
- Wynne, M. (2005). *Stylistics: Corpus approaches*. (Master's thesis). Retrieved from <http://www.pala.an.uk/uploads/2/5/1/0/25105678.pdf>

**AUTHOR'S BIO**

---

*Eman Adil Jaafar is an assistant professor of linguistics at the Department of English/ College of Education for Women / University of Baghdad. Her research interests include but not restricted to stylistics, corpus and cognitive stylistics and applied linguistics.*

---