

Paul Brocklebank

Department of Liberal Arts
Tokyo University of Technology

Johnson and the Eighteenth–Century Periodical Essay: A Corpus–Based Approach

Summary

The style of Samuel Johnson's essays for the periodicals *The Rambler*, *The Adventurer* and *The Idler* is quite different from that of earlier eighteenth–century essayists such as Joseph Addison and Jonathan Swift. However, despite advances in recent years in corpus–based stylistic approaches to texts, a comparison of these three authors using current corpus–analytic techniques has yet to be attempted. This paper reports on the first stages of such a project. Johnson's essays are compared with Addison and Swift's essays using *WordSmith Tools 5*, and an analysis of keywords, semantic groupings of keywords, and key collocations of keywords in Johnson's essays are identified. It is argued that a keyword analysis brings to the fore grammatical aspects of Johnsonian sentence patterns and provides empirical support for what have hitherto been only intuitively–based statements regarding his style. Also, further patterns in the data will be identified through a phraseological analysis of the essays focusing on the most common four–word clusters (4–grams) that Johnson uses.

Key words: corpus stylistics, corpus linguistics, keywords, 4–grams, eighteenth century periodical essays, Samuel Johnson

Johnson in periodični eseji 18. stoletja: korpusni pristop

Povzetek

Slog v esejih Samuela Johnsona, ki so izhajali v časopisih *The Rambler*, *The Adventurer* in *The Idler* se precej razlikuje od sloga pisanja drugih esejistov zgodnjega 18. stoletja, kot sta na primer Joseph Addison in Jonathan Swift. V zadnjem času je sicer prišlo do napredka v stilističnih raziskavah besedil s pomočjo korpusnega pristopa, vendar bo besedila omenjenih treh avtorjev še potrebno primerjati s pomočjo korpusno–analitičnih tehnik. Članek predstavlja prve faze tega projekta. Johnsonove eseje primerjamo z esei Addisona in Swifta s pomočjo orodja *WordSmith Tools 5* in prepoznavamo ključne besede, semantične skupine ključnih besed ter ključne kolokacije ključnih besed v Johnsonovih esejih. Trdimo, da analiza ključnih besed potisne v ospredje slovnične vidike Johnsonovih stavčnih vzorcev in tako zagotovi tudi empirično podporo za sedanje intuitivne trditve o Johnsonovem slogu pisanja. Druge vzorce bomo pridobili s pomočjo frazeološke analize esejev, ki se bo omejila na štiri–besedne sklope (4–gram), ki so značilni za Johnsona.

Ključne besede: korpusna stilistika, korpusno jezikoslovje, ključne besede, 4–gram, periodični eseji iz 18. stoletja, Samuel Johnson

Johnson and the Eighteenth-Century Periodical Essay: A Corpus-Based Approach

1. Introduction

A recent trend in corpus stylistics has been to apply corpus-based approaches such as keyword analysis (see Scott 2002) and cluster analysis (as in Mahlberg 2007, 2009) to fictional texts, mainly novels. This paper reports on an attempt to extend the use of these techniques to the eighteenth-century periodical essay, focusing on an examination of Samuel Johnson's essays in *The Rambler*, *The Adventurer* and *The Idler*, as compared with essays written by Jonathan Swift and Joseph Addison in the earlier years of the same century.

Johnson's distinctive style has often been acknowledged, not only by scholars interested in the stylistics of eighteenth-century prose (see, for example, Wimsatt 1941, 1948, and McIntosh 1998), but also by his contemporaries. Indeed, Wimsatt (1941, 133) goes as far as to say, 'The *Rambler* style made a splash. Johnson is himself an event in the history of English prose. His style was recognized by contemporaries as "something extraordinary, a prodigy or monstrosity, a huge phenomenon."'

However, while certain idiosyncrasies of this prose style were identified by Wimsatt (1941), such as (syntactic and semantic) parallelism, antithesis and philosophic diction (meaning the use of scientific terminology derived from Greek and Latin sources), there has not yet been an attempt to employ more recent techniques from corpus linguistics to Johnson's periodical essays. This paper amounts to a first step towards doing this. Using Mike Scott's *WordSmith Tools* software (Scott 2007) I attempt to use keyword and cluster techniques in order to reveal what makes Johnson's prose style distinctive vis-à-vis the earlier stylistic models of Swift and Addison.

2. The data

This study focuses on the periodical essays that Johnson contributed to three publications from 1750 to 1760 – *The Rambler* (1750–52), contributions to which make up the bulk of the essays, *The Adventurer* (1752–54) and *The Idler* (1758–60). The text of *The Rambler* came from the Electronic Text Center at the University of Virginia (<http://www2.lib.virginia.edu/etext/index.html>). To enable analysis, the HTML pages for these essays were downloaded and converted into text format. Text files of *The Adventurer* and *The Idler* essays came straight from Project Gutenberg (<http://www.gutenberg.org/>). For all the essays I carried out some pre-editing of the text by removing the Latin and Greek mottos at the beginning of each contribution and deleted any lengthy quotation, whether poetry or prose, and whether in Latin, Greek, English or any other language.

The data with which Johnson's periodical output is compared comprises those essays by Addison and Swift that are readily available in electronic format at Project Gutenberg. These were all of Addison's contributions to *The Spectator*, and those periodical essays by Swift that were published in *The Tatler*, *The Examiner*, *The Spectator* and *The Intelligencer*. The Addison and Swift essays were pre-edited in the same way as the Johnson essays.

The composition of the three text files/corpora is summarized in Table 1:

Text file	Number of essays
<i>Johnson</i>	323 (<i>The Rambler</i> : 203, <i>The Adventurer</i> : 29, <i>The Idler</i> : 91)
<i>Addison</i>	255 – all from <i>The Spectator</i>
<i>Swift</i>	54 (<i>The Tatler</i> : 17, <i>The Examiner</i> : 33, <i>The Spectator</i> : 1, <i>The Intelligencer</i> : 3)

Table 1. *Composition of the three writers' text files.*

2.1 Keywords and their key collocates

This corpus-based analysis of Johnson's essays involves an examination of lexical differences between these essays and those of Addison and Swift. This section will look at the most statistically-significant keywords and the main collocates that they pattern with, Section 2.2 narrows the focus to key content words, and Section 2.3 deals with key four-word clusters (also known as four-word strings or '4-grams').

The notion of keyword is now a familiar one in corpus linguistics. A keyword is a word that appears in a particular corpus a statistically significant number of times more often than in another (usually larger) 'reference corpus.' Keywords, therefore, are lexical items that are prominent or foregrounded in [Text A] when contrasted with their use (or non-use) in [Text B]. The semantic content of keywords is seen as a good indicator of the foregrounded content of a text, reflecting what the text is 'about'. For examples of research in corpus stylistics where this idea of 'keyness' plays an important role, see Scott (2002), Rayson (2008), Culpeper (2009), and Fischer-Starcke (2009).

If we look at the number of tokens in each text file (see Table 2) one can observe that the Johnson corpus at 434,344 tokens is slightly larger than the combined size of the Addison and Swift reference corpus at 412,572 tokens. In addition, the Addison section of this reference corpus is over three and a half times larger than the section containing Swift's essays.

Section of text file	Number of tokens
<i>The Rambler</i>	295,625
<i>The Adventurer</i>	46,532
<i>The Idler</i>	92,187
JOHNSON TOTAL	434,344
<i>Addison (all)</i>	323,841
<i>Swift (all)</i>	88,731
ADDISON & SWIFT TOTAL	412,572

Table 2. *Number of tokens in each section of the text files.*

This lack of balance is potentially problematic. Merely combining the Swift and Addison text files and comparing this single corpus with Johnson's essays risks producing misleading results, as the comparison would lack balance and be heavily weighted towards Addison. Therefore, to give a more equitable comparison, when calculating the Johnsonian keywords I decided to run each comparison separately before merging the two sets of results.

Two lists of keywords were generated, one for Johnson versus Addison, the other for Johnson versus Swift. To do this I used the *WordSmith Tools KeyWords* program, which takes two wordlists and carries out a proportional statistical comparison by applying a log-likelihood test of significance to the frequency scores of each word in the lists. Application of this statistical test results in a 'keyness score' being obtained for each keyword, and the *KeyWords* program outputs an ordered list of keywords. Positive keywords are those words which appear in Text A proportionally more often than in Text B, whereas negative keywords are those which appear proportionally less often.

For this study, probability was set to $p < 0.00001$ and the minimum number of hits for inclusion in the list of keywords was 3. With these settings 525 positive keywords were generated for Johnson versus Addison and 124 positive keywords for Johnson versus Swift. These two lists were then reduced to a single 'key keyword list' of 92 'key keywords' by selecting only those words that were common to both lists. Finally, a combined ranking list of 'key keywords' was compiled by taking the keyness scores for each word and then calculating the average score.

Below are the top ten 'key keywords' for Johnson's essays:

Key keywords	Frequency in J	Frequency in A & S	Average keyness score
BY	5,291	2,660	450.78
WITHOUT	1,382	478	242.02
CAN	1,279	474	203.76
AND	16,002	12,269	198.64
OR	3,732	2,266	174.01
HAPPINESS	420	80	160.35
ALWAYS	755	240	151.39
LIFE	988	411	146.89
EVERY	1,564	823	137.59
NO	1,558	794	131.46

Table 3. Key keywords in Johnson's periodical essays.

Most of these most prominent keywords are functional (two prepositions BY and WITHOUT, a modal CAN, two conjunctions AND and OR, and two determiners NO and EVERY), and the only content words are HAPPINESS, ALWAYS and LIFE. The predominance of function words is somewhat surprising as it is often assumed that the main purpose of a keyword analysis is to identify the 'aboutness' of a text, and that therefore items such as content words and proper nouns will rise to the top of the list. In this case it is possible that the results reflect basic differences in sentence structure between Johnson and the earlier essayists. For example, it is likely that the presence of the two conjunctions points to a greater use of coordinate structures in the former, whether at the sentence or phrase level. Since Wimsatt (1941) it has been acknowledged that one signature of Johnson's style is the large amount of parallelism in the essays, where in many instances a conjunction operates as a 'hinge' between parallel elements (see below for further evidence of this parallelism in the data that *WordSmith* brings to our attention).

To investigate these conjectures it is important to examine how the keywords operate in context by consulting a concordance and seeing how other words collocate with the keyword. Using *WordSmith Concord* I generated a table of collocates (up to and including five places to the left and right of the headword) and a concordance for the top keyword, BY.

L1	Headword	R1
AND (121), BUT (106), ONLY (92), OR (56), IT (48), THAT (46), THEMSELVES (42), HIMSELF (34), THEM (34), HIM (31)	BY (5291)	THE (924), A (317), WHICH (291), HIS (166), THEIR (110), AN (95), SOME (86), THOSE (84), THIS (71), ANY (51)

Table 4. Top ten adjacent (L1 and R1) collocates for BY in Johnson's essays (with frequencies).

R1 (one place to the right of BY) was predominantly filled by articles or determiners, reflecting the fact that in most cases BY combines with a noun phrase (NP) to form a prepositional phrase. One place to the left of BY (L1) was more interesting. The large number of conjunctions in this position (AND, BUT, OR) reflects a tendency to coordinate the by-phrase; ONLY is often used as a modifier before the phrase; and a look at the concordance for BY reveals that THAT is in most cases the complement marker, indicating that the by-phrase often precedes the other elements in a complement clause.

Finally, the large number of object pronouns (IT, THEMSELVES, HIMSELF, THEM, HIM) in L1 point to sequences of 'verb + object pronoun + by-phrase' as common colligations in the essays. Here are some examples of this pattern as revealed by the concordance:

with seriousness, and improve it	by	meditation; and that,
the particles that impregnate it	by	their salutary or malignant
and to recommend themselves	by	minute industry
till they discovered themselves	by	some indubitable token
that he has enslaved himself	by	some foolish confidence
without time to prepare himself	by	previous studies.
things, because we measure them	by	some wrong standard.
when Truth ceased to awe them	by	her immediate presence
wrong when they are shewn him	by	another; but he that has
solicitudes, and divert him	by	cheerful conversation.

Table 5. Concordance examples of 'verb + object pronoun + by-phrase' from Johnson's essays.

The main problem with just looking at the concordances for each keyword is the lack of any comparative element. How are we to know whether the collocation patterns are unusual or not? It could well be the case that the patterns are also common in the reference corpus and therefore relatively insignificant.

One way to help overcome this is to extend the notion of keyness to the collocates themselves. To calculate these key collocates I first combined the Addison and Swift files into a single 'AS' file. Then text file concordances were generated for the top ten Johnsonian 'key keywords' for both of the files 'J' (Johnson) and 'AS', with a character span set at 60 characters around each keyword.

Wordlists were generated for each concordance, and finally the wordlists were compared by having *WordSmith* calculate the positive key collocates for the Johnson concordances. Maximum p value was adjusted to 0.0001 to allow more key collocates to be generated.

The ten most significant results ranked according to keyness scores are in Table 6 below. AND is the keyword that has the greatest number of the strongest collocates. BY is the strongest lexical collocate for OR and the second strongest collocate for AND; an examination of the concordance for these two pairs showed strong preferences for pre- ('and by NP') and post-coordination of by-phrases ('by NP and'), confirming that there is a marked tendency for coordination of by-phrases as hypothesized in the preliminary collocate analysis above. In addition, it appears that EVERY + MAN collocate so strongly because EVERY MAN is such a common phrase in Johnson's essays, occurring 272 times in all.

Rank	Key keyword	Key collocate	Keyness score
1	AND	WITHOUT	149.60
2	AND	BY	133.51
3	AND	TO	117.20
4	OR	BY	108.03
5	EVERY	MAN	100.59
6	AND	THEREFORE	88.76
7	AND	HOPES	80.24
8	AND	NO	78.84
9	AND	WHOSE	72.01
10	AND	CONFIDENCE	66.07

Table 6. Top ten key collocates of key keywords in Johnson's periodical essays.

Turning now to the strongest keyword-collocate pairing AND + WITHOUT, if we consult a concordance of examples for these two words when they occur in proximity to one another, we see that WITHOUT occurs most often (140 times) two places before AND, in other words a major configuration in the essays is [WITHOUT X AND]. The main category exponent is, unsurprisingly, a noun (N), and the main patterns with extended environments to the left and right of the configuration are as follows (with each extended pattern occurring ten or more times in the data):

- (1) VP [without N, and] VP
Ex... may wanton in cruelty without control, and trample the bounds of right (*The Rambler* 148)
- (2) VP [without N and] without NP
Ex... and [leaves us without importance, and without regard] (*The Rambler* 72)
- (3) NP [without N and] NP without NP
Ex... he has [birth without alliance, and influence without dignity] (*The Rambler* 142)
- (4) VP [without N, and] S
Ex... the malignity is continued without end, and it is common for old maids (*The Rambler* 46)

The coordination is, therefore, for the most part with a VP (verb phrase) (in 1), another prepositional phrase headed by 'without' (in 2), another NP that contains a prepositional phrase headed by 'without' (in 3), or with another clause (S) (in 4).

What is most striking is the large number of parallel examples involving ‘without’ that occur, either in [PP and PP] as in (2) above, or in [NP and NP] as in (3), or, to give some further patterns, in [VP and VP], (ex. he [roars without reply, and ravages without resistance] in *The Rambler* 72) or in [AP and AP], where AP is an adjectival phrase (ex. to be [inactive without ease, and drowsy without tranquillity] in *The Adventurer* 39). It would appear that parallelism is such a prominent aspect of Johnson’s periodical prose style that it surfaces even when stylistic analysis is at the lexical as well as the syntactic level.

2.2 Key content words and their behavior

As the top of the list of key keywords is dominated to such an extent by tokens belonging to functional categories, I decided to target my analysis on the content words (nouns, adjectives, verbs and adverbs) in my list of 92 so as to enable me to examine the thematic content of Johnson’s style more readily. In fact, of these 92 key keywords there were 72 content key keywords all together, even after omitting the two proper nouns RAMBLER and IDLER. It was, therefore, only at the very top of the list that the function words predominated.

Here are the ten content words with the highest average keyness scores:

Content keywords	Frequency in J	Frequency in A & S	Average keyness score
HAPPINESS	420	80	160.35
ALWAYS	755	240	151.39
LIFE	988	411	146.89
SCARCELY	156	0	115.92
KNOWLEDGE	399	97	111.89
ATTENTION	277	44	111.29
EASILY	278	47	99.86
THEREFORE	754	329	88.63
EQUALLY	245	41	87.42
HOPES	219	34	84.30

Table 7. Top content keywords in Johnson’s periodical essays.

The next step in the procedure was to group the content words according to semantic similarity. To do this I used the categories implemented in the *UCREL Semantic Analysis System* (or *USAS*; see Rayson 2008, 2009) for tagging text with the *Wmatrix* software tool. I checked the accuracy of the tagging manually and made additions when necessary. For example, I had to add a tag for *FELICITY*, which not appearing in the standard *Wmatrix* lexicon initially failed to receive a tag.

There were seven groups of three or more members. E4.1+, in other words, the group of ‘happiness and contentment’ lexical items (HAPPINESS, MERRIMENT, FELICITY, PLEASURE, GRATIFICATIONS, MISERIES and MISERY) was the largest. Clearly this is an important concept that Johnson investigates in his essays – indeed, it may be remembered that HAPPINESS was one of the ten main key keywords. Nine members belong to a broad ‘psychological’ grouping: five to the ‘expected’ group, X2.6+ (HOPES, HOPE, EXPECTED, EXPECTATION and

EXPECTATIONS), and four to the ‘interested/excited/energetic’ group, X5.2+ (ARDOUR, DILIGENCE, CURIOSITY and EAGERNESS).

‘Evaluation’ (A5: SUPERIORITY, EXCELLENCE and ERROUR) and ‘comparing’ (A6: EQUALLY, EQUAL, ACCUSTOMED) are also important things that Johnson is attempting, and although treated separately in the *USAS* categorization system can perhaps be placed together in one larger group of six members. The personality traits (S1.2) KINDNESS, ENVY and VANITY are also prominent. Finally, the group of three frequency adverbs (SOMETIMES, ALWAYS and SELDOM) is probably a reflection of Johnson’s many attempts at providing generalizations in his essays.

As HAPPINESS was not only a keyword but also evidently a key concept, I decided to look at how this important word patterned with other words in the essays, again using Addison and Swift’s essays as a standard of comparison. The methodology for obtaining the key collocates was identical to that described in Section 2.2, except that this time I restricted my examination to content collocates. The major key collocates are listed below:

Rank	Key content collocate	Frequency in J
1	LIFE	41
2	FOUND	14
3	DAY	12
4	MEANS	11
5	LONG	10
6	PLACE	10
7	MEN	10
8	RICHES	9
9	HOURL	8
10	KNOWLEDGE	8

Table 8. *Key content collocates of HAPPINESS.*

LIFE was by far the most significant collocate, appearing 41 times in the environment of HAPPINESS in Johnson, but only twice in Addison and Swift (none of the other collocates on this list appeared at all as collocates of HAPPINESS in the Addison and Swift corpus, so here I have just listed the frequencies in Johnson).

To identify how HAPPINESS collocates with these other words, I worked through the concordances and tried to identify common patterns in Johnson’s use of each pair of words in context. This procedure, therefore, signaled a shift from a quantitative approach to a more qualitative type of analysis. Below, I provide a snapshot of how the words are used, followed by some examples from Johnson’s essays to illustrate their use.

(a) HAPPINESS and LIFE

LIFE may be the locus of HAPPINESS, but more often it is the lack or provisional nature thereof which is foregrounded:

- (5) “Such is the condition of **life**, that something is always wanting to **happiness**.” (*The Rambler* 196)
- (6) “And so scanty is our present allowance of **happiness**, that in many situations **life** could scarcely be supported...” (*The Adventurer* 69)
- (7) “Thus every period of **life** is obliged to borrow its **happiness** from the time to come.” (*The Rambler* 203)
- (8) “What state of **life** admits most **happiness**, is uncertain...” (*The Adventurer* 111)

(b) HAPPINESS and FOUND

HAPPINESS is something to be FOUND, although finding it may be difficult or hardly worth the effort:

- (9) “(I) longed for the **happiness** which was to be **found** in the inseparable society of a good sort of woman.” (*The Idler* 100)
- (10) “But what is success to him that has none to enjoy it? **Happiness** is not **found** in self-contemplation...” (*The Idler* 41)
- (11) “But by him that examines life with a more close attention, the **happiness** of the world will be **found** still less than it appears.” (*The Adventurer* 120)

(c) HAPPINESS and LONG

HAPPINESS may well be a LONG time coming:

- (12) “... he that has been **long** accustomed to please himself with possibilities of fortuitous **happiness**...” (*The Adventurer* 69)
- (13) “... I wondered how it could happen that I had so **long** delayed my own **happiness**.” (*The Rambler* 165)
- (14) “... the **happiness** that I have been so **long** procuring is now at an end...” (*The Adventurer* 102)

(d) HAPPINESS and PLACE

HAPPINESS is a state in which one may PLACE oneself or one may PLACE in the hands of others:

- (15) “... (they) **place** themselves at will in varied situations of **happiness**...” (*The Rambler* 89)
- (16) “Among wretches that **place** their **happiness** in the favour of the great...” (*The Rambler* 189)

Change of PLACE, however, is not to be recommended:

- (17) “All my **happiness** has been destroyed by change of **place**...” (*The Idler* 53)
- (18) “... all expectations of **happiness** from change of **place** would cease.” (*The Idler* 50)

As for HAPPINESS and its other key content collocates, their main uses can be summarized as follows (because of space restrictions I omit the examples):

- Every DAY or every HOUR represents a step towards attaining the elusive state of HAPPINESS.
- HAPPINESS is something that MEN contemplate and wish to attain, and there are certain MEANS by which they may obtain it.
- But if they think RICHES will bring them HAPPINESS, they are sorely mistaken.
- Finally, KNOWLEDGE is seen as something worth attaining along with HAPPINESS (and virtue).

In this way one can use information from key collocate lists and concordances to provide sketches of how a particular word is used by an author. In the above, a single instance was chosen to illustrate the approach, but of course similar sketches can be obtained by following the same procedure for the other key content words.

2.3 Key four-word clusters

Finally, let us take a brief look at the most significant four-word clusters (also referred to as four-word strings or ‘4-grams’) used in Johnson’s essays. While keywords reflect textual content, clusters more often reflect structural features of a text. Clusters four words in length were chosen because clusters that are longer than this tend to occur in more restricted syntactic, semantic, pragmatic contexts (see Starcke 2008).

To obtain these clusters a list of four-member strings was generated from the Johnson text file using the *kfNgram* software tool (Fletcher 2002). This was repeated for the text file containing Addison and Swift essays, and then keyness scores for the 27 clusters occurring ten or more times in Johnson were calculated in the usual way. I then attempted to sort the 4-grams into semantic groups. Only four were clearly identifiable; these are shown in Table 9 together with their ranking which is indicated by the numbers in brackets.

Semantic group	Four-word clusters
Letter-writing formulas	to the rambler sir (1), I am sir c (4), to the idler sir (7), am sir your humble (12), sir your humble servant (13), the rambler sir I (22)
Extent	the greater part of (3), for the most part (5), greater part of mankind (9), the rest of mankind (11), it is common to (24)
Humankind	greater part of mankind (9), the rest of mankind (11), of the human mind (20), of the world will (25)
Ease/difficulty	it is not easy (6), is not easy to (23)

Table 9: Key groups of four-word clusters in Johnson’s essays.

The discourse structural features brought to the surface here are the (fragments of) letter writing formulas, which make up the largest identifiable group. In fact, ‘to the rambler sir,’ occurring 47 times in Johnson and, unsurprisingly, never in the Addison and Swift corpus, was the most prominent four-word cluster. There was also an ‘extent’ grouping, in which all of the clusters refer to majorities or parts, again possibly as a reflection of Johnson’s generalizing tendency touched on earlier. ‘Humankind’ was another group, another focus of Johnson’s enquiries – it may be remembered that MAN was a key collocate of the key keyword EVERY, and MEN a key collocate of HAPPINESS. The only other group was ‘ease/difficulty,’ which consists of just two strings, ‘it is not easy’ and ‘is not easy to.’

Not belonging to a group, but important because it was the cluster with the second highest keyness score, was ‘in a short time.’ This tended to be used in a particular way by Johnson. It hardly ever appeared clause-finally, and while semantically it has a temporal sense, pragmatically it is usually

used in the essays as a marker of change, with the change being non-neutral – in other words, from a non-negative to a negative state, as in these examples:

- (19) “Serenus **in a short time** began to find his danger...” (*The Adventurer* 62)
- (20) “Among those whose reputation is exhausted **in a short time** by its own luxuriance...” (*The Rambler* 106)
- (21) “Thus, **in a short time**, I had heated my imagination to such a state of activity...” (*The Rambler* 101)

In (19) Serenus moves from a state of blissful ignorance to a sense of peril; in (20) reputations are tarnished; and in (21) a state of calm is replaced by an overheated imagination.

Or the change may involve a move from an inferior state to a better one, as in:

- (22) “... of those who break the ranks and disorder the uniformity of the march, most return **in a short time** from their deviation...” (*The Rambler* 135)
- (23) “...he therefore studied all the military writers both ancient and modern, and, **in a short time**, could tell how to have gained every remarkable battle that has been lost...” (*The Rambler* 19)

The change in (22) is from disorder to order; and in (23) ignorance is replaced by knowledge.

2.4 Further work

This paper reports on what is only the first stage of the current project; a full comparison of the three writers’ styles will only be possible when Addison’s and Swift’s essays have been compared with their respective reference corpora (‘Johnson/Swift’ and ‘Johnson/Addison’ respectively).

There is also the question of similarities to be considered. What do the three writers’ essays share from a lexical viewpoint? Investigating this will require a contemporaneous reference corpus that consists of something other than periodical essays, and for this I am considering using a corpus of eighteenth-century fiction that I have compiled.

Finally, it would be interesting to extend the analysis to look at key parts of speech and key semantic categories. However, to do this properly all of the essays would have to be tagged fully before we would be able to analyze them with the *Wmatrix* software. To ensure that all of the words are tagged with part-of-speech information and semantic tags the lexicon needs to be extended to cover the words not in the standard *Wmatrix* lexicon. Divergent (from author to author) and antique spellings also need to be recognized by the tagger. Work on compiling a suitable lexicon is currently in progress.

3. Conclusion

My aim has been to show how techniques from corpus stylistics can be used to investigate the language of Johnson’s periodical essays. At best such an analysis can reveal patterns that would not be evident using a more traditional, close reading approach to the texts; at the very least the method produces empirical evidence for patterns that the analyst may have originally identified through informed intuition. Application of these techniques has thrown up a number of different, but interlinked results.

First, where the results from the initial keyword analysis revealed that Johnson not only had a preference for using the prepositions BY and WITHOUT and the conjunctions AND and OR,

the follow-up key collocate analysis showed that these prepositions and conjunctions tended to co-occur, so that conjoined BY and WITHOUT-phrases were common in the essays. The prominence of conjunctions probably reflects the large number of parallel structures that is such a striking aspect of Johnson's style.

The analysis of content keywords showed that the concept of HAPPINESS was a key concern of Johnson. The word HAPPINESS was the top content keyword and 'happiness-contentedness' was the largest key content category. Keywords can and should be investigated further by identifying key collocates and noting common uses by reading off the words in context using concordances. Each approach is likely to bring to the surface something different, although even here strong conceptual links between lexical items may be evident – for example, not only were HAPPINESS and LIFE key keywords but LIFE was the main content collocate of HAPPINESS.

Finally, a mixture of quantitative and qualitative analyses of four-word clusters may also throw light on differences in the texts, particularly as regards discourse structures.

Bibliography

- 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.
- Fischer-Starcke, B. 2009. Keywords and frequent phrases of Jane Austen's *Pride and Prejudice*: A corpus-stylistic analysis. *International Journal of Corpus Linguistics* 14 (4): 492–523.
- Fletcher, W.H. 2002. *kwNgram*. Available at: <http://www.kwicfinder.com/kfNgram/kfNgramHelp.html> (accessed August 2012).
- Mahlberg, M. 2007. Clusters, key clusters and local textual functions in Dickens. *Corpora* 2 (1): 1–31.
- . -. 2009. Corpus stylistics and the Pickwickian watering-pot. In *Contemporary Corpus Linguistics*, ed. P. Baker, 47–63. London: Continuum.
- McIntosh, C. 1998. *The Evolution of English Prose 1700–1800: Style, Politeness, and Print Culture*. Cambridge: Cambridge University Press.
- Rayson, P. 2008. From key words to key semantic domains. *International Journal of Corpus Linguistics* 13 (4): 519–49.
- Scott, M. 2002. Comparing corpora and identifying key words, collocations, and frequency distributions through the *WordSmith Tools* suite of computer programs. In *Small Corpus Studies and ELT: Theory and Practice*, ed. M. Ghadessy, A. Henry and R.L. Roseberry, 47–67. Amsterdam/Philadelphia: John Benjamins.
- . -. 2007. *WordSmith Tools 4.0*. Oxford: Oxford University Press.
- Starcke, B. 2008. I don't know – differences in patterns of collocation and semantic prosody in phrases of different lengths. In *Language, People and Numbers: Corpora in Society – Theoretical and Applied*, ed. A. Gerbig and O. Mason, 199–216. Amsterdam: Rodopi.
- Wimsatt, W.K. 1941. *The Prose Style of Samuel Johnson*. New Haven: Yale University Press.
- . -. 1948. *Philosophic Words: A Study of Style and Meaning in the Rambler and Dictionary of Samuel Johnson*. New Haven: Yale University Press.
- The Rambler* <http://www2.lib.virginia.edu/etext/index.html>
- Project Gutenberg (<http://www.gutenberg.org/>)