

Analysis of the Phraseology of Online Newspaper Articles: A Brief Corpus-based Study on Online Newspaper Articles about Robin Williams' Passing

Andrés C. Cardozo*

Facultad de Lenguas, Universidad Nacional de Córdoba

Franco Zanfagnini**

Facultad de Lenguas, Universidad Nacional de Córdoba

Abstract

Authors such as Halliday (1985) and Sinclair (1991) have demonstrated that the organization of language has a lexis-grammar basis, which entails that vocabulary items are not isolated, but should rather be approached within a context. The importance of this approach lies in that it can display a general panorama of a particular semantic prosody, defined as “a form of meaning which is established through the proximity of a consistent series of collocates” (Louw, 2000, p.9). As pointed by Gledhill (2000), it is important to look at collocational patterns within a specific genre and a specific context in order to understand how phraseological conventions work. We believe that by looking at collocational patterns in newspaper articles related to a particular event, we can get an insight into this speech community's perception of such event. This brief study aims at analyzing the semantic prosody of online newspaper articles, more specifically the way in which collocations play a fundamental role in how judgment is passed when reporting and linguistically reconstructing events of global importance. By using concordancer software and looking at the collocations of certain high frequency words in our corpus, we will analyse the reports on the suicide of the well-known American actor Robin Williams.

Introduction

Recent advances in communication technologies have allowed for easier access to digitalised texts. Linguistics has been greatly benefited from these advances, especially because research in areas such as corpus-based linguistics can be carried out more efficiently, especially when analysing such aspects as collocational patterns within a genre. This study attempts to shed light on the way

* **Andrés Cardozo** graduated as a teacher of English in 2010, at Facultad de Lenguas, Universidad Nacional de Córdoba. He has been a lecturer in English Language and Phonetics and Phonology at tertiary and university level. He has also worked as a Spanish Language Assistant in several schools in Stockton-on-Tees, United Kingdom, as part of a bilateral British-Argentinian exchange programme. He has been part of a research team on the impact of virtual classrooms in the pronunciation class at University level. Currently, he is completing a Master's Programme in English Language with a mention in Applied Linguistics at Universidad Nacional de Córdoba. Contact: andrescardozo540@gmail.com

** **Franco Zanfagnini** graduated in 2011. He holds a degree in English Language Teaching from Facultad de Lenguas, Universidad Nacional de Córdoba (UNC). He is a Professor in the Phonetics and Phonology II chair. He has also taught in the Language I, Language II and Language III chairs. He has been part of a research project with the Phonetics and Phonology II chair called “*La Evaluación de la oralidad en la universidad ¿Evaluación impresionista u holística de pronunciación en inglés como lengua extranjera?*”, and he is finishing his Master's Dissertation on Applied Linguistics at Facultad de Lenguas (UNC). Moreover, since 2013, Franco has taught at Ciclo de Nivelación at Facultad de Lenguas teaching both Pronunciation and English Language. Contact: frank.zanfagnini@hotmail.com

in which collocations play a fundamental role in the phraseology of newspaper articles, more specifically, on how judgment is passed when reporting and linguistically reconstructing events of global impact.

Authors such as Halliday (1985) and Sinclair (1991) have demonstrated that the organisation of language has a lexis-grammar basis, which entails that vocabulary items are not isolated, but should rather be approached in context. The importance of this approach lies in that it can display a general panorama of a particular semantic prosody, seen as “a form of meaning which is established through the proximity of a consistent series of collocates” (Louw, 2000, p.9). As pointed by Gledhill (2000), it is important to look at collocational patterns within a specific genre and context in order to understand how phraseological conventions work.

We believe that by looking at collocational patterns in newspaper articles related to a particular event, we can get an insight of this speech community’s perception of such event.

This brief study aims at analysing the semantic prosody underlying the use of collocations to analyse how online versions of newspaper articles reflect the way in which writers instantiate their perceptions of a high impact event. By looking at collocations of two high frequency words in our corpus (*depression* and *suicide*), we will focus on the reports on the suicide of well-known American actor Robin Williams and analyse how collocational patterns are used to express the authors’ ideas about these two concepts.

Corpus Linguistics and Discourse analysis

As opposed to other approaches to linguistic analysis, Corpus Linguistics focuses on natural language in use, rather than on made-up linguistic samples. One of the main researchers in this approach is Michael Stubbs (1993), who assumed the task of determining a set of principles which underlie Corpus Linguistics. Such principles express that this approach sees language as a social construct, and that language should be analysed in actual occurrences which can be considered authentic samples of analysis. Among these principles, Stubbs also mentions the importance of studying whole texts and of concentrating not only on grammar but also on meaning, as they are inseparable.

In line with Corpus Linguistics, the main focus of Discourse Analysts has always been the study of language beyond the sentence with a strong emphasis on actual language in use. Moreover, authors such as Hyland (2005) or Yatsko (2013) point out that discourse analysis goes even further and tries to disclose other aspects pertaining to the individual or group under analysis. As the latter states, “discourse analysis is a semi-linguistic methodology in which analysis of text structure is a means of revealing some socio-psychological characteristics” of the individual or group under analysis. (Yatsko, “Computational Linguistics Laboratory.”)

Historically, authors have failed to reach an agreement as to how to define what Critical Discourse Analysis is. In the light of this lack of a unified criterion to characterise it, Paltridge (2006) proposes his own definition, stating that Critical Discourse Analysis does not simply explore the surface meaning reflected in language, but rather it seeks to establish links between the use of language and the social context in which such use takes place. By looking at such connections, it is possible to analyse how ideologies can be reflected in discourse, as authors make use of the language to pass on judgment in their productions. However, these ideologies are generally not individual, but rather shared across a speech community – that is to say, the genre to which a certain text belongs is likely to influence the way in which judgment is passed. As Paltridge specifies, genre is a key factor at the time of trying to impose a viewpoint; as such, genre analysts will try to delve into how language in use can share certain characteristics, even when the authors belong to different speech communities. The assumption is that genre will greatly influence the way in which ideologies are transmitted through phraseology, despite the original culture in which the texts are immersed. In

other words, culture is not the only factor which affects the way in which ideologies are transmitted within a specific genre.

Collocations

The concept of collocation plays a fundamental role in this brief study. Even though there exist many definitions of the term, Cortes' definition of collocation seems to be particularly clear. Cortes has defined collocations as "sequences of [two], three or more words that statistically co-occur in a register." (Cortes, 2004, p. 400) What is more, Cortes, and other authors such as Biber et al (1999) have coined the term *lexical bundles* to refer to, as Cortes indicates, *extended collocations* (Cortes, 2004); that is, groups of words which show a tendency to be combined in particular linguistic contexts with a relative regularity. Firth (1951) also used the terms collocation or collocability to make reference to "habitual occurrence of a word with another word or words." (Firth, as cited in Cortes, 2004) In this study, the importance of collocations lies in the measurement of the occurrence of certain word combinations so as to analyse how particular linguistic communities pass judgement on certain issues pertaining to a global reaching event.

The newspaper articles corpus

As Gledhill (2000) states, "it is now widely accepted in corpus linguistics that the context of a specialised corpus must be as explicit as possible and must display clear design criteria." In line with this idea, the authors decided to construct their own corpus, in order to fit the specific purposes of this research. Therefore, in order to carry out this research, we have resorted to gathering electronic articles retrieved from eight different newspaper websites.

We retrieved the same number of samples (10 articles) from each newspaper. The samples collected belong to internationally known papers, which we consider to be a valid selection to conduct the research because of the representativeness of the linguistic community under analysis. The newspapers chosen to build this corpus were: The Huffington Post; The New York Times; The Guardian; The Australian; The Indian Express; The Sydney Morning Herald; Los Angeles Times; and The Daily Mail. The selection of the texts was carried out following these criteria:

1. Accessibility: all of the texts chosen were retrieved electronically from free-access online versions of newspapers. Even though certain articles could not be retrieved since a fee payment was required, we resorted to including only free-access ones.
2. Reliability: all of the articles chosen belong to internationally prestigious sources.
3. Heterogeneity: the samples were retrieved from different speech communities (culture-wise) in an attempt to guarantee the minimal possible cultural bias, since the study focuses on genre rather than cultural assumptions.

The first step towards analysing the corpus obtained from the online versions of the newspapers was to copy the texts into computer text files. One of the difficulties we faced was that this was a time-consuming task since it was a long process of removing the format in which the texts were written. Another difficulty was that some of the websites accessed did not allow copying the texts in an automated fashion, so the process was even more time-consuming than expected. Once all the formatting was removed, the texts were saved in .txt documents, in order to be accessible via the selected software.

In the present study, we made use of Lawrence Anthony's free software AntConc (available online at <http://www.laurenceanthony.net/software.html>), which was most appropriate for the quantitative part of the analysis and for organising the items graphically for a simpler qualitative analysis. This computer programme was first used to retrieve a list of sorted lexical items in order to evaluate their frequency of occurrence. Based on the results displayed by AntConc, we could decide which of the most salient items we would focus on. Once this step was completed, we

generated concordances to look at the context in which such items were used. The following section will deal with the collection and analysis process in detail.

Data collection and analysis

The generated corpus has a total of 63205 words. After obtaining a high-frequency words list, we focused on the main salient lexical items. On the contrary, structural words were not taken into account since they would not ultimately serve the purpose of this study, which is to analyse how judgement is passed with the aid of content words. The words *depression* and *suicide* were chosen from the corpus for two main reasons: a) they have a similar frequency of occurrence (while the word *depression* yielded 255 hits, *suicide* yielded 230 hits) and b) we consider them to be a reliable source for analysis of the phraseology used for passing judgement. Table 1 gives detail of the salient words under analysis in comparison to the totality of the corpus.

Table 1: Frequency and proportion of selected items in relation to the main corpus.

Word Rank	Word	Frequency	Proportion
30	depression	255	0.40
35	suicide	230	0.36

The proportion might lead the reader to think that the percentage (0.40 and 0.36) is relatively low. However, taking into account the relationship between structural and lexically loaded words, these two items represent a rather high proportion of the present corpus.

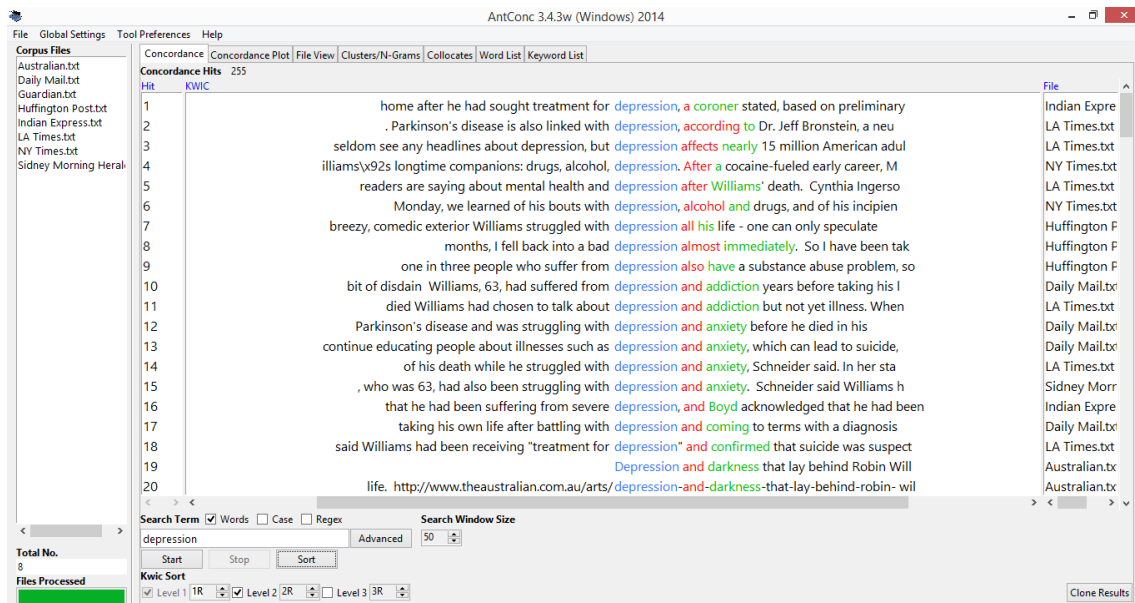
Even though some other words were of interest for the present study and analysis, we decided to narrow down the selection to fit the scope of this study. However, some other words might also be included in the future to pave the way for a more detailed and comprehensive kind of research.

Depression

In order to analyse the phraseology of this term, we used the concordancer function on AntConc so as to come up with a list of contextual occurrences. Such list was first sorted with one and two levels to the right, and one and two levels to the left. In other words, we analysed the lexical item and the relationship it established with the neighbouring items in different sentences, as exemplified in screenshots 1, 2, 3 and 4.

When analysing the right-side collocates (see Screenshot 1), it was striking to see how the word *depression* is frequently horizontally paired with other mental disorders by means of the coordinating conjunction *and*. This pairing shows that depression is not apparently perceived as the ultimate cause of Williams' death, but is often accompanied by other highly negatively-loaded words referring to conditions such as *anxiety* and *paranoia*. Such phenomenon was observed in articles from different newspapers, which might give us a hint of how this high-reaching event was perceived internationally. Some other examples are: "depression and addictions", "depression and the complications of addictions", "depression and other mental illnesses".

Screenshot 1: "depression" with a 1- and 2-to-the-right sorting.

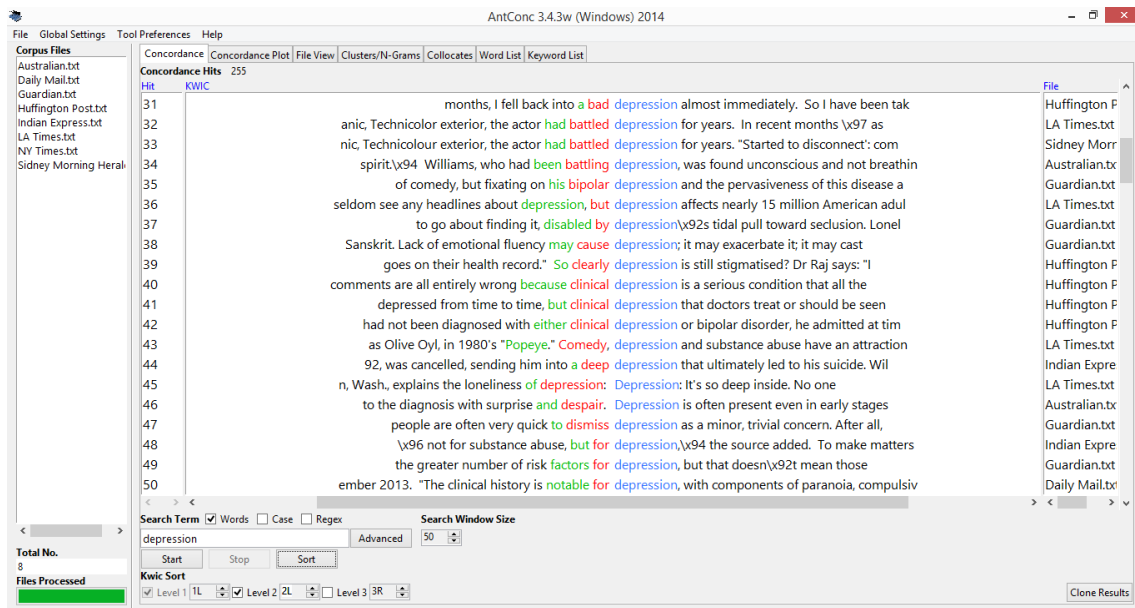


However, there is a general consensus as to seeing “depression” as a disease or an external entity which has a bulldozing effect on its sufferer, thus portraying the world-known actor as a victim rather than as a conscious agent who committed the act of suicide. For instance, “depression is a disease like any other disease”, and “clinical depression is a serious condition”. Moreover, AntConc also displayed a series of sentences which were not frequent collocates but which accentuated the fact that depression is a disease, such as “people who think depression is a choice are wrong” and “depression is no more a choice than baldness”.

Another noteworthy characteristic identified when sorting lexical items to the right was that the word *depression* has been systematically personified as an entity which can exert certain influence on the sufferer. This is achieved by phrasings which have *depression* as a head accompanied by action verbs (sometimes mediated by a modal denoting potentiality). For example, “...depression can destroy lives”, “...depression takes someone...”, “depression that cripples millions...” etc.

On the other hand, when analysing the left-side collocates (see Screenshot 2), it can be observed that there is great similarity with the right-side pairings concerning mental illnesses. We could notice that noun phrases such as “substance abuse” and “bipolar disorders” were present in couplings by using the coordinating conjunction “and”, thus reinforcing the idea that depression alone cannot be considered the sole cause of suicide.

Screenshot 2: “depression” with a 1- and 2-to-the-left sorting.



In line with the right-side analysis, *depression* is also depicted as an active force with a negative impact on the affected person. In this case, *depression* is part of a prepositional phrase virtually always headed by *from* and accompanying a form of the verb *suffer*, and, at a lower degree, in phrases headed by *to* with verbs such as *succumb*. Examples of this are: “Williams had in fact been suffering from depression” and “he was succumbing to depression”.

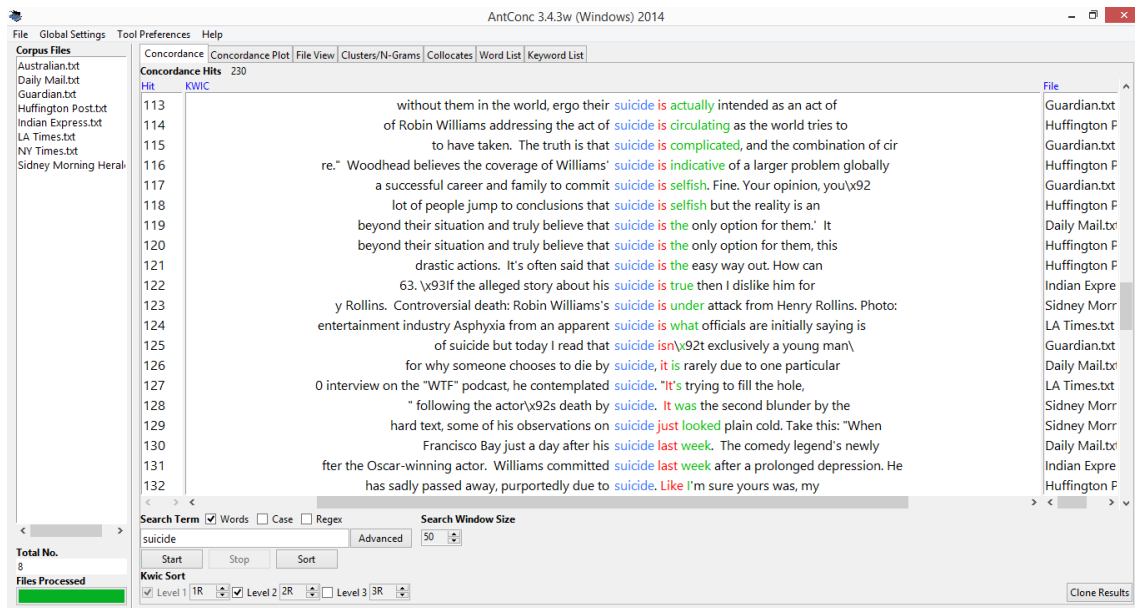
Along with the idea of depression as an external force, the most frequent left-side collocate is the verb *battle*, which entails a metaphoric perception of the concept of depression. The conceptual metaphor underlying the phrase “battle depression” is that of conceiving the disease as an enemy who needs to be defeated. This pervasive metaphoric conception of the medical condition could give an idea of the general awareness present in different newspapers articles across various cultures. Some examples of how this metaphor is instantiated are: “he struggled with his own battles of depression” and “Williams was battling severe depression of late”.

As regards the adjectives which act as pre-modifiers of *depression*, the most frequent collocate is *severe*, which denotes the critical view on the part of the writers, who apparently perceive this disease as something which cannot be countered easily. In rare cases, words such as *clinical*, *genuine*, *prolonged* and *real* collocate to the left side of the term. Whereas words like *clinical* and *prolonged* are neutral and therefore are not used to pass on judgement, words like *genuine* and *real* can reflect the idea that many people who claim to suffer from depression are not actually affected by it, and that it entails something far more serious which can have grave consequences.

Suicide

When analysing the collocates of “suicide” to the right (see Screenshot 3), it is difficult to find phraseology which denotes a viewpoint on the part of the text producers.

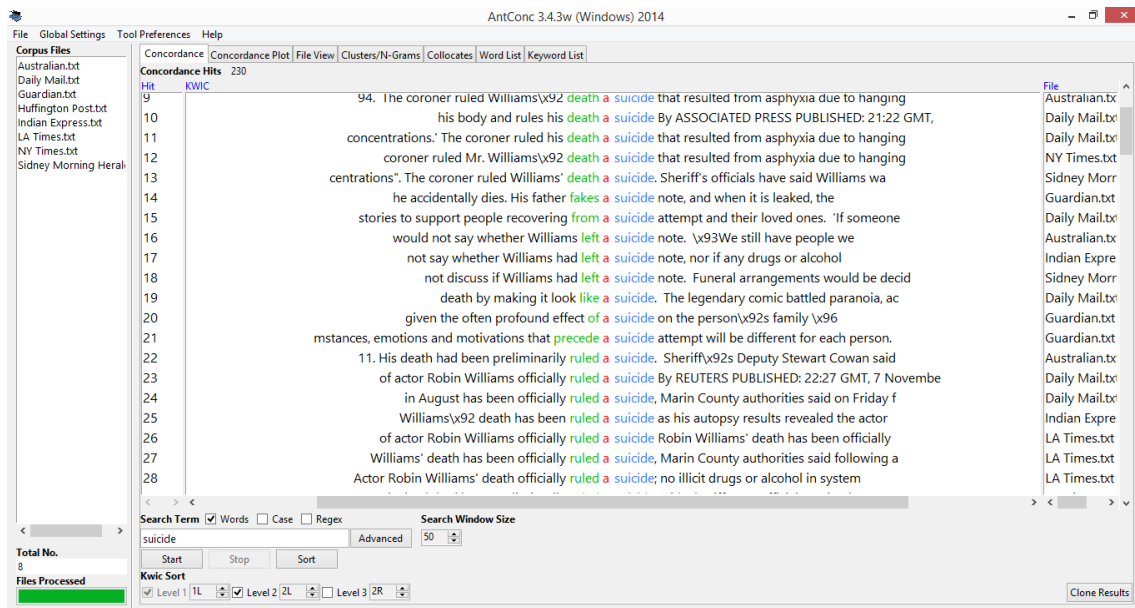
Screenshot 3: “suicide” with a 1- and 2-to-the-right sorting.



Some of the most frequent collocates of *suicide* are examples of set expressions such as *note*, *attempt*, and *risk*, which do not necessarily carry a certain ideology. However, we could find certain instances of semantic prosody, which was present mainly in relational clauses where *suicide* was given certain attributes by means of adjectives such as *selfish*, or by noun phrases such as “the easy way out” or “a permanent solution for a temporary problem”. The use of these expressions denotes the writer's standpoint, where suicide is attributed a negative connotation, and where the suicide victim might no longer be seen as a victim, but as a wrongdoer who can be judged. However, as the frequency of such instantiations is very low, this idea might not be representative of the whole corpus under analysis.

Conversely, when analysing the left-side collocates (see Screenshot 4), there is noticeably more analysis that can be carried out under our criteria. It was interesting to notice that in the articles immediately following Williams' death, the adjective *apparent* as a pre-modifier of *suicide* is frequently used. The reason for this choice might be that at the time of reporting this recent event, the case was still under scrutiny, resulting in the writers avoiding giving a categorical view on the events.

Screenshot 3: “suicide” with a 1- and 2-to-the-right sorting.



One of the most salient phraseological choices is the verb *commit*. To a certain extent, this verb denotes a degree of responsibility on the part of the agent, as can be seen in the following examples: “Williams committed suicide last week after a prolonged depression” and “he committed suicide by hanging himself”. The analysis of the word *commit* establishes a clear difference in the perception of the concepts of suicide and depression. This difference will be discussed in detail in the following section.

This assignment of responsibility can also be gauged by the use of such strategies as the genitive case and possessive adjectives. By attributing suicide to a certain agent, the article writers may distance themselves from the victim’s actions and depict the act of suicide as something which is not approved of. Expressions such as “his suicide”, “(Robin) Williams’ suicide” and “the actor’s suicide” pervade the articles, which might imply that the perception of suicide being a conscientious choice is generalised and also perceived negatively by the international communities.

Discussion

In this section we will discuss our findings and we will try to interpret and compare the data thrown by the analysis carried out earlier.

Our assumptions had been that by looking at the semantic prosody of both *depression* and *suicide*, we would be able to uncover hidden connotations which would show the different writers’ viewpoints. However, it is surprising that the majority of the lexical items which accompany *suicide* are not used to pass judgment but seem to be used in order to add information about the event.

Whereas in the analysis of the word *depression* it was observed that a high degree of reciprocity is given to the sufferer, the word *suicide* tends to portray the person as an active agent who makes a conscientious choice. Depression is therefore rendered as an external force, whereas suicide is conceptualised as an act of wrongdoing where the concept of responsibility is reinforced.

What is more, an extended hypothesis can be that the use of the phraseology of these two words can be intrinsically related to widespread people’s viewpoints (based on religious beliefs, morals or ethics) as regards depression and suicide. As controversial as people’s views can be, depression is frequently treated as a negative phenomenon which the sufferer might not have control over, whereas suicide seems to be seen as a negative egotistical action. As the concept of suicide might be against most people’s values, it is generally rejected. As previously analysed, the act of suicide can be labelled as an “easy way out”, a phrase which enacts the widespread rejection.

Conclusion

The present paper has attempted to integrate Discourse Analysis and Corpus Analysis. This approach has allowed us to analyse the passing of judgment in online newspaper articles by focusing on the phraseology of two lexical items. By analysing the instances of frequent collocations of these items, we can achieve a deeper understanding of how world views can be reflected in writing. What is more, further insights can be gained as to the impact of a world-reaching event in certain linguistics communities across cultures.

From the perspective of genre analysis, this kind of corpus research can be very helpful at the time of characterising the phraseology which underlies a certain speech community. However, the present study is considered to be only a brief analysis focused on a very limited corpus, and so it is far from drawing a definite general phraseology of the online newspaper article genre. An extended version of this study could be carried out in order to gain a more comprehensive scope and to gain a deeper insight into other aspects of the phraseology of the genre.

References

- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes* 23 (4), 397-423. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0889490603000851>
- Gledhill, C. (2000). The discourse function of collocation in research article introductions. *English for Specific Purposes* 19, 115-135. Retrieved from http://www.academia.edu/8250148/The_discourse_function_of_collocation_in_research_article_introductions
- Halliday, M. A. K. (1985). *An Introduction to Functional Grammar*. London: Arnold.
- Hyland, K. (2005). 'Digging up texts and transcripts: Confessions of a discourse analyst', in P. K. Matsuda and T. Silva (eds.), *Second Language Writing Research: Perspectives on the Process of Knowledge Construction*. Mahwah: Lawrence Erlbaum, pp 177-189.
- Louw, B. (2000). 'Contextual prosodic theory: Bringing semantic prosodies to life' in C. Heffer and H. Sauntson (eds.) *Words in Context, A Tribute to John Sinclair on His Retirement*. Birmingham: University of Birmingham, 48-94.
- Paltridge, B. (2006). *Discourse Analysis*. London: Continuum.
- Sinclair, J. M. 1991. *Corpus, Concordance, Collocation*. Oxford: OUP
- Stubbs, M. (1993). 'British traditions in texts analysis: from Firth to Sinclair'. In M. Baker, G. Francis and E. Tognini-Bonelli (eds.) *Text and Technology: in honour of John Sinclair*. Amsterdam: Benjamins. 1-36.
- Yatsko, V. (2013). 'Computational Linguistics Laboratory'. Retrieved from <http://yatsko.zohosites.com/integrational-discourse-analysis-conception.html>

