Using Key Words as Nodes of Cohesive Networks
Lorenzo Mastropierro (University of Birmingham, UK)

This paper explores the potential of key word analysis to shed light on lexical cohesion in literary texts, showing how key words can support the study of cohesive networks. Through the analysis of H. P. Lovecraft's *At the Mountains of the Madness*, it explores how a corpus linguistic approach can further our understanding of lexical cohesion as an incremental textual feature that contributes to literary meaning.

Key word analysis has been extensively used for the study of literary texts (for example, Stubbs 2001, Stubbs 2005, Toolan 2009, Culpeper 2009, Walker 2010, Mahlberg & McIntyre 2011). However, an aspect of key words that does not seem to have been explored yet is their role in building cohesion: key words, as a form of repetition, contribute cohesion to a text. Moreover, considering key words as items that occur repeatedly as the text goes on also emphasises the directionality of the repetition and the fact that meaning is incremental, reflecting the progression of the text. In contrast, in corpus linguistics the emphasis is on the amount of data and meaning is discussed from a cumulative, rather than directional, point of view. With this paper we aim therefore to explore the potential of key words to act as nodes of cohesive networks that develop incrementally throughout a text. As Thornbury (2010: 279) explains, "while a list of keywords is not in itself a semantic network, it provides the raw data out of which such a network can be constructed".

In order to test this potential, we carry out a study of cohesive networks in *At the Mountains of Madness*. First, we use a log-likelihood test ($p \leq 0.0000001$) to generate a list of key words comparing the text (41,431 words) with a reference corpus of coeval American fiction (over 2M words). We then select key words to examine in detail as potential nodes of cohesive networks. Two interrelated criteria are used for the selection of the key words: (i) frequency – the selected key words must occur frequently throughout the text, and (ii) thematic relevance – the selected key words must relate to the themes of the short novel. Frequency is important because a word that is repeated frequently across a text contributes importantly to the cohesion of that text. Thematic relevance is important because themes and "aboutness" are textual features that develop incrementally throughout the text. A word that is both thematically relevant and frequent, as a content key word can be, is therefore a possible candidate to represent a node of a cohesive network. For the purpose of exemplifying the method of cohesive networks identification, we focus on two key words that meet both the aforementioned criteria, *mountains* (48 occurrences) and *peaks* (29 occurrences), using traditional literary criticism to inform our understanding of what is thematically relevant in the text.

With the help of the concordancer, we analyse the key words' patterns and their local textual functions (Mahlberg 2005, 2006). We identify two textual behaviours that are shared by both key words: a semantic preference of "Height", as seen in *great, highest, steep, titan, elevation, gigantic, greatest, loftiest, mighty*, etc., and a negative semantic prosody referred to as "Mysteriousness", that characterises *mountains* and *peaks* as a frightfully mysterious place, as seen in the co-occurrence of words such as *madness, horrible, evil, nameless, terrible, unknown, brooding, cryptic, forbidden, frightful, horror, looming, mysterious, mystery*, etc. These patterns are built incrementally with every occurrence of the key words: when *mountains* and *peaks* appear, they co-occur with the words that form the shared semantic preference/prosody. Each occurrence of these key words thus connect together the instantiations of the semantic preference/prosody, creating a net of connections that builds incrementally and spans across the whole text. These connections are conceptualised as a

cohesive network based on the repetition of the key words and the semantic preference/prosody. Figure 1 below shows an example of the network. The figure shows 12 random sentences in which *peaks* and *mountains* occur, sorted in the order they appear in the text. The nodes of the network (the key words) are underlined, while the words belonging to the shared semantic preference and prosody are in bold. Every successive instantiation of the semantic preference/prosody builds on the previous ones, through the link that the repetition of the key words establishes.

> (1) The last lap of the voyage was vivid and fancy-stirring, **great barren** <u>peaks</u> of **mystery looming** up constantly against the west as the low northern sun of noon or the still lower horizon-grazing southern sun of midnight poured its hazy reddish rays over the white snow […]. (2) […] since an aeroplane survey of the nearly exposed rock surfaces shewed an entire absence of those Archaean and primordial strata for which he was looking, and which formed so great a part of the **colossal** <u>peaks</u> that **loomed** up at a tantalising distance from the camp. (3) […] but at this time and place, with those **dark**, **unknown** mountain <u>peaks</u> **soaring stupendously** ahead, that anomalous elder-world discovery in our minds, and the pall of probable disaster enveloping the greater part of our expedition, we all seemed to find in it a taint of latent malignity and infinitely evil portent. (4) The **unknown** <u>mountains</u> ahead rose **dizzyingly** up like a fearsome rampart of giants […]. (5) One had to be careful of one's imagination in the lee of those **overshadowing** <u>mountains</u> of **madness**. (6) In spite of all the prevailing horrors we were left with enough sheer scientific zeal and adventurousness to wonder about the **unknown** realm beyond those **mysterious** <u>mountains</u>. (7) As we drew near the **forbidding** <u>peaks</u>, **dark** and **sinister** above the line of crevasse-riven snow and interstitial glaciers, we noticed more and more the curiously regular formations clinging to the slopes; […] (8) The touch of **evil mystery** in these barrier <u>mountains</u>, and in the beckoning sea of opalescent sky glimpsed betwixt their summits, was a highly subtle and attenuated matter not to be explained in literal words. (9) […] a **frightful** line of <u>peaks</u> had shot suddenly up amidst the most appalling din and chaos— and earth had received her **loftiest** and most **terrible** <u>mountains</u>. (10) This vast nighted gulf had undoubtedly been worn by the great river which flowed down from the **nameless** and **horrible** westward <u>mountains</u>, […]. (11) Perhaps we were mad—for have I not said those **horrible** <u>peaks</u> were <u>mountains</u> of **madness**? (12) For this far violet line could be nothing else than the **terrible** <u>mountains</u> of the **forbidden** land— **highest** of earth's <u>peaks</u> and focus of earth's **evil**; […].

Figure 1. Example of cohesive network

In Sentence (1), *peaks* occurs with *great*, *barren*, *mystery*, and *looming*. When the reader finds *peaks* again in the text in Sentence (2), a connection is established with the previous occurrence. In this case, *peaks* occurs with *colossal* and *loomed*, which relate to *great* and *looming* respectively. In Sentence (4), *mountains* recalls the previous uses of *peaks*; as both terms refer to the same fictional place. Here, the adjective *unknown* links back to the earlier occurrence of the same word (in Sentence (3)) or to the use of related items (*mystery*, in Sentence (1), for example). These connections are established every time *peaks* and *mountains* occur, and every occurrence reinforces the link in an incremental fashion. In Sentence (9), towards the end of the short story, *peaks* occurs with *frightful*, while *mountains*

occurs with *loftiest* and *terrible*. *Frightful* and *terrible* refer back to all of the occurrences of the "Mysteriousness" semantic prosody, whereas *loftiest* builds on the repetition of the "Height" semantic preference.

We argue that this network, in addition to creating lexical cohesion in the text, participates in the process of building up the fictional world. The representation of *mountains* and *peaks* is in fact cohesive and consistent throughout the text, and this contributes to define the spatial locales of *At the Mountains of Madness*. As recognised by many literary critics (e.g. Ralickas 2007, Kneale 2006), spaces and places in this short novel are a fundamental aspect to establish the atmosphere of malevolence and awe that characterises the text.

Overall, building on corpus linguistic models, cohesion can be conceptualised as the sum of relationships between lexical items rather than between individual words. The approach we present in this paper shows how it is possible to study cohesion as a feature that spans across the whole of a text. We argue that key words can function as nodes of cohesive networks, and that their repetition contributes incrementally to the construction of literary meanings.

References

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.

Kneale, J. (2006). From beyond: H. P. Lovecraft and the place of horror. *Cultural Geographies, 13*, 106-126.

Mahlberg, M. (2005). *English General Nouns: A Corpus Theoretical Approach*. Amsterdam: John Benjamins.

Mahlberg, M. (2006). Lexical cohesion: Corpus linguistic theory and its application in ELT. Special issue of the *International Journal of Corpus Linguistics, 11*(3), 363-383.

Mahlberg, M., & McIntyre, D. (2011). A case for corpus stylistics: Ian Fleming's *Casino Royale*. *English Text Construction, 4(*2), 204-227.

Ralickas, V. (2007). "Cosmic horror" and the question of the sublime in Lovecraft. *Journal of the Fantastic in the Arts, 18*(3), 364-398.

Stubbs, M. (2001). *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford: Blackwell Publishing.

Stubbs, M. (2005). Conrad in the computer: Examples of quantitative stylistic methods. *Language and Literature, 14*(1), 5-24.

Thornbury, S. (2010). What can a corpus tell us about discourse? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 270-287). London: Routledge.

Toolan, M. (2009). *Narrative Progression in the Short Story. A Corpus Stylistic Approach*. Amsterdam: John Benjamins.

Walker, B. (2010). Wmatrix, key concepts and the narrators in Julian Barnes's *Talking it Over*. In D. McIntyre & B. Busse (Eds.), *Language and Style* (pp. 364-387). Basingstoke: Palgrave Macmillan.