Using Distributional Semantics to Trace Influence and Imitation in Romantic Orientalist Poetry

Nitish Aggarwal[*] **Justin Tonra**[°] **Paul Buitelaar**[*]
[*]Insight Centre for Data Analytics
National University of Ireland, Galway, Ireland
firstname.lastname@deri.org

[°]University Fellow in English, School of Humanities
National University of Ireland, Galway
justin.tonra@nuigalway.ie

Abstract

In this paper, we investigate whether textual analysis can yield evidence of shared vocabulary or formal textual characteristics in the works of 19th century poets Lord Byron and Thomas Moore in the genre of Romantic Orientalism. In particular, we identify and trace Byron's influence on Moore's writings to query whether Moore imitated Byron, as many reviewers of the time suggested. We use a Distributional Semantic Model (DSM) to analyze if there is a shared vocabulary of Romantic Orientalism, or if it is possible to characterize a literary genre in terms of vocabulary, rather than in terms of the particular plots, characters and themes. We discuss the results that DSM models are able to provide for an abstract overview of the influence of Lord Byron's work on Thomas Moore.

1 Introduction

Literary criticism has often marshalled the serendipitous discovery in the service of constructing an argument or a critical judgment. Such serendipity can take material or cognitive form, and provide the raw materials for analysis and conjecture. In literary criticism, arguments are often based upon evidence gleaned from close reading of a text in support of a hypothesis, but quantitative methods have shown how literary texts can yield evidence that is not immediately discernible to the human eye for a similar interpretive purposes. In literary studies, computers have assisted in the collection of such data with varying degrees of complexity and sophistication for about half a century. How can we use the information from such computing processes for creating new knowledge, or, in literary-critical terms, for articulating the meaning in a text? To what degree can literary criticism and computing enrich one another? Is algorithmic criticism derived from algorithmic manipulation of text (Ramsay, 2011) possible?

Inspired by general questions such as these, this paper discusses a particular project that uses Distributional Semantics to trace influence and imitation between two particular poets writing in the genre of Romantic Orientalism. Our intuition is that if text analysis can yield evidence of shared vocabulary to trace influence between poets, we can build a network of different authors with their degree of influences. This can help a reader in finding a similar literature and in discovering implicit information.

In the period from 1813 to 1817, friends and fellow-poets Lord Byron and Thomas Moore wrote a series of long poems which are now seen as representative of Romantic Orientalism (a subset of Romantic literature recognisable by its Oriental and Middle-Eastern themes and settings). Throughout this period, an unusual pattern of coincidence is evident in the writings of the two poets, with correspondence between the poets describing similar plots, settings, and characters names in their respective works. The publication of Byron's quartet of Oriental tales in 1813 and 1814 (*The Giaour, The Bride of Abydos, The Corsair, Lara*) anticipated much of the substance of Moore's work, and delayed the publication of his own suite of four Oriental poems, *Lalla Rookh*, until 1817. On the publication of the latter, many reviewers accused Moore of imitating Byron's work, correctly fulfilling Moore's own prediction of 1813, that he would be seen as "an humble follower–a Byronian" (Moore, 1964). Subsequent critics have generally acknowledged Byron as a direct influence on Moore, but the basis of these acknowledgements

is usually subjective critical interpretation of plot, character, and poetic form in the published texts (see, in particular, (Vail, 2001), (Sultana, 1989), (Gregory, 2008)). More general accounts of Byron's and Moore's literary association can be found in (Hamilton, 1948), (Jordan, 1948), (Tessier, 2014).
The purpose of this project is to investigate further the possible causes for the unusual pattern of coincidence in the writings of these two poets during this time. A Computational Linguistics approach to Byron's and Moore's Orientalism was identified as a potentially productive way of studying coincidence, influence, and imitation between their writings and how they related to the genre of Romantic Orientalism. Fresh empirical insight into this topic is desirable because of the difficulty of thinking about and articulating these issues in a way that is not speculative or nebulous. Such methodologies have not been applied to these texts, and offer the possibility of yielding fresh perspectives on questions about the texts, and the genre of Romantic Orientalism: is it defined by a limited vocabulary which inevitably leads to similarities and coincidences between its practitioners? Does writing within a specific genre impose topical or semantic constraint upon the author?
The motivations for the project emerge from a conviction that Computational Linguistics techniques may reveal evidence of shared vocabulary or formal textual characteristics in the works of Byron and Moore during the period 1813-17. The questions that the project seeks to answer include: can we identify and trace Byron's influence on Moore's writings? Did Moore imitate Byron, as many reviewers of the time suggested? Is there a shared vocabulary of Romantic Orientalism? Is it possible to characterise a literary genre in these terms, rather than in terms of plot, character, theme, etc.? The basis for such enquiries must go beyond a subjective comparison of the poems: the method that has characterised literary-critical approaches to these texts to date.

2 Distributional semantics

How related are love and emotion? Reasoning about semantic relatedness of natural language text is not a very difficult task for a human because of sufficient background knowledge and other related information to understand the semantics of natural language text. However, for computers, it is still an open issue to provide significant background knowledge to understand the complex structure of natural language. One plausible way to provide such background knowledge is taking the usage of given text in large contextual space into account.
Semantic relatedness of two given terms (text fragments, phrases or words) can be obtained by calculating the correlation between two high dimensional vectors of a Distributional Semantic Model (DSM), which is based on the assumption that semantic meaning of a text can be inferred from its usage in context (Harris, 1954), i.e. its distribution in text. DSM builds this semantic representation through a statistical analysis over the large contextual information in which a term occurs (see for details (Landauer, 1998), (Blei, 2003)). One recent popular model to calculate this semantic relatedness by using the distributional semantics is Explicit Semantic Analysis (ESA) proposed by (Gabrilovich and Markovitch, 2007), which attempts to represent the semantics of the given term by a high dimensional vector in explicit concept space such as Wikipedia concepts. Every explicit concept represents a dimension of the ESA vector, and associativity weight of a given term with the explicit concept reflects the vector dimension weight. For instance, for a given term t, ESA builds a vector v, where $v = \sum_{i=0}^{N} a_i * c_i$ and c_i is i^{th} concept from the explicit concept space, and a_i is the associativity weight of term t with the concept c_i. Here, N represents the total number of concepts. The semantic relatedness score is calculated by taking cosine between the corresponding high dimensional vectors.

3 Approach

Section 1 described our aim to investigate whether textual analysis techniques can yield evidence about Byron's influence on Moore's writings by analyzing the four long poems (published in 1813-14) by Lord Byron and a collection of four long poems (published in 1817) by Thomas Moore. To analyze this influence, we calculate semantic relatedness scores between Byron's poems and Moore's poems. We split

these poems in line-groups[1] and obtain 227 line-groups from Byron's poems and 246 line-groups from Moore's poems. We calculate ESA scores of every line-group of Byron's poems with every line-group of Moore's poems. All the line-group pairs can be sorted according to their relatedness scores, which can provide highly related line-group pairs. After getting these highly related pairs, we can manually analyze them, and if manual analysis confirms the high relatedness of the pairs provided by ESA, then it may indicate some degree of influence or imitation between the poets. Also, these results will conclude that text analysis techniques can reduce the human effort in analyzing the influence between work by different authors.

4 Evaluation

4.1 Experiment

We built two ESA models; one by using Wikipedia and the other by using a corpus of poetry primarily from the eighteenth and nineteenth centuries[2]. In the first model, we take every Wikipedia article as a dimension of the ESA vector, and TF-IDF weight of a given text with article content is considered as the associativity strength with the corresponding dimension. We use modified ESA (Aggarwal, 2012) which builds the ESA vector by taking all words of a given text together rather than taking them individually. Wikipedia may not have a good coverage of the vocabulary of poems in Romantic Orientalism, which led us to try another ESA model that utilizes a Poetic Corpus. This corpus consists of 892 long poems and some of the poems contain more than 7K lines. Therefore, we split each poem with their line-groups and obtain 22K different line-groups. Similar to Wikipedia-based ESA, we take every line-group as a dimension of the ESA vector, and TF-IDF weight of the given text with line-group is considered as associativity strength with the corresponding dimension.
We use both ESA models: Wikipedia-based ESA and Poetic Corpus-based ESA to calculate the semantic relatedness scores of every line-group of Byron's poems with every line-group of Moore's poems. Both the results obtained by these two models are analyzed manually to check if Poetic Corpus-based ESA outperforms Wikipedia-based ESA as Poetic Corpus has better vocabulary coverage for Romantic Orientalism poems. We described in section 3 that we obtain 227 line-groups from Byron's poems and 246 line-groups from Moore's poems that means 56K line-group pairs. Manual analysis of 56K line-group pairs will take a very long time, therefore, we analyze only a small subset of the 56K pairs. To select the sample, we categorize the line-group pairs in three different categories: Highly-Related, May-be-Related and Not-Related. In the ranked list of line-group pairs, the top 1K are considered Highly-Related, the pairs ranked between 25K to 26K are considered May-be-Related, and the bottom 1K are taken as Not-Related. We randomly selected 5 line-group pairs from each category and manually analyzed the results obtained from ESA. Hence, we analyzed 15 pairs obtained according to Wikipedia-based ESA and 15 pairs according to PoeticCorpus-based ESA.

4.2 Results and Discussion

Manual (close-reading) analysis of 15 line-group pairs from the Wikipedia-based ESA took place first. At first glance, the pairs identified as Highly-Related were indeed quite closely related, particularly in terms of their narrative content. While some individual line-groups appeared in more than one pairing identified by the model, the pair exhibited a frequently occurring narrative scenario where a female character addressed her male lover before the departure or death of one of the parties. The model also succeeded in identifying this scenario in poems by both Byron and Moore. The scenes are unsurprisingly united by the presence of strong emotional language and imagery on the theme of love. However, the recognition of a leavetaking (whether in death or departure) in the scenes is also noteworthy, as is the fact that the identified line-groups are comprised of direct quotations from characters (as opposed to poetic narrative).

[1]Line-groups in poetry are similar to paragraphs in prose. On the printed page, a line of white space separates one line group from the next. Like paragraphs, they vary in length, and are often semantically, syntactically, or thematically self-contained.

[2]The poems in this corpus come from Women Writers Project (1560-1845), Eighteenth-Century Collections Online (1701-1800), and poetic corpora shared by Ted Underwood (1701-1899)

Subjected to manual analysis, pairs of line-groups in the May-be-Related and Not-Related categories exhibit varying degrees of relatedness. Most are lacking the immediate recognition of narrative similarity evident in the Highly-Related pairs, with some pairs containing vastly different narrative scenarios. Many of the consistencies from the Highly-Related category are also absent: some pairs vary greatly in length, and some contain a mix of narrative and quotation. One example from the manually-analysed examples proved to be a potential anomaly: a pair determined by the model to be Not-Related (i.e. in the bottom 1K of pairs in terms of relatedness) might easily be considered related in that both line-groups are florid poetic descriptions of a pastoral landscape.

The results of the Poetic Corpus-based ESA model were similar, if a little more refined. Interestingly, the line-group pairs in the Highly-Related category were largely similar to those resulting from the Wikipedia-based ESA. They were comprised of direct quotation (rather than narrative), and featured a character speaking to their lover in strong emotional language. In some cases (though not consistently) greater linguistic similarities between the pairings were more evident than in the results of the Wikipedia-based ESA. This was an anticipated consequence of using the Poetic Corpus-based ESA, where the model would be more likely to recognise the more unconventional features of nineteenth-century poetic diction than the Wikipedia-based ESA.

From a literary-critical perspective, however, identification of the Highly-Related pairs by a computer is no great advance on the capabilities of human scholarship. A traditional scholar can just as easily recognise the similarities in the scenes identified by both the Wikipedia- and Poetic Corpus-based ESAs in the course of reading the eight poems by Byron and Moore. Their narrative similarity is the most prominent characteristic that contributes to their relatedness. This identification can be made by the lone scholar because the dataset is relatively small in this project, and the time needed to read and analyse it is not prohibitive. The potential value of this kind of automated semantic-relatedness identification is increased when it is applied in a more exploratory fashion to larger datasets, and to poetic corpora whose scales are beyond the reasonable comprehension of the individual scholar. In this scenario, a potential application of the process would involve identifying and mapping the patterns and networks of relatedness in large-scale poetic corpora. For the present purposes of this project–studying imitation and influence in the texts of Byron and Moore–semantic relatedness measurements have been of limited value on their own, but have offered promise in other areas. The first aspect of their success has been in identifying sentiment analysis as a potential next step in drilling down into the texts to further reveal the essence of their similarity. The second is revealing a wider application of semantic relatedness in examining broader patterns of similarity within the history of poetry.

5 Conclusions and Future Work

We developed a method to identify influence and imitation in Romantic Orientalism poetry. We built two Explicit Semantic Analysis (ESA) models by using Wikipedia and a Poetic Corpus. The results from the analysis conducted with the Poetic Corpus-based ESA were a slight improvement on those resulting from the Wikipedia-based ESA. This was as anticipated, and results might be improved even further with a refined Poetic Corpus comprised of works from a more concentrated time period, which are more likely to share linguistic similarities with the Byron and Moore poems.

The performance of ESA model depends on several parameters (Aggarwal, 2014) that are included in the model, therefore, future work will include an investigation of ESA model in literature research. Also, we are planning to use an improved version of the ESA (Polajnar, 2013) model which reduce the orthogonality problem in the model. The value of ESA to the particular task of tracing imitation and influence in the Romantic Orientalist poetry of Byron and Moore has been limited thus far, but it has provided evidence of linguistic similarities in the expression of emotion. The next step for further investigation of imitation and influence between the two poets will involve the use of sentiment analysis. ESA was successful in identifying line groups that were closely related in terms of their narrative content and in their use of similarly emotional language. For such a small dataset, this does not represent a significant improvement on close reading, as similar results could have been obtained in this manner quite quickly. But the automated identification of semantic relatedness demonstrated in this project has

potentially valuable applications for exploring broader literary corpora. For instance, a semantic mapping of transnational and transhistorical poetic relatedness is a possible future venue for our research.

Acknowledgments

This work has been funded in part by a research grant from Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289 (INSIGHT) and by the EU FP7 program in the context of the project LIDER (610782).

References

Harris, Zellig, *Distributional structure*, 1954. Word 10 (23): 146-162.

Gabrilovich, Evgeniy and Markovitch, Shaul, *Computing semantic relatedness using Wikipedia-based explicit semantic analysis*, 2007. Proceedings of the 20th international joint conference on Artifical intelligence Hyderabad, India 1606–1611

Landauer, Thomas K and Foltz, Peter W and Laham, Darrell, An introduction to latent semantic analysis, Discourse processes, 25, 2-3, 259–284, 1998, Taylor & Francis

Blei, David M and Ng, Andrew Y and Jordan, Michael I Latent dirichlet allocation, the Journal of machine Learning research, 3, 993–1022, 2003, JMLR. org

Aggarwal, Nitish and Asooja, Kartik and Buitelaar, Paul DERI&UPM: Pushing corpus based relatedness to similarity: Shared task system description., Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation. Association for Computational Linguistics, 2012.

Ramsay, Stephen *Reading Machines: Toward an Algorithmic Criticism*, Urbana: University of Illinois Press, 2011. Print.

Moore, Thomas The Letters of Thomas Moore, Ed. Wilfred S. Dowden. 2 vols. Oxford: Clarendon Press, 1964. Print.

Gregory, Allan, "Thomas Moore's Orientalism." *Byron and Orientalism*, Ed. Peter Cochran. Newcastle upon Tyne: Cambridge Scholars, 2008. 173-82. Print.

Hamilton, Ian, "Byron and the Best of Friends." *Keepers of the Flame Literary Estates and the Rise of Biography.* London: Hutchinson, 1992. Print.

Jordan, Hoover H., "Byron and Moore." *Modern Language Quarterly*, 9.4 (1948): 429-39. Print.

Sultana, Fehmida, "Romantic Orientalism and Islam: Southey, Shelley, Moore, and Byron.", Unpublished dissertation. Tufts University, 1989. Print.

Polajnar, Tamara and Aggarwal, Nitish and Asooja, Kartik and Buitelaar, Paul, *Improving ESA with document similarity*, *Advances in Information Retrieval*, 582–593, 2013, Springer

Aggarwal, Nitish and Asooja, Kartik and Buitelaar, Paul *Exploring ESA to Improve Word Relatedness*, *Third Joint Conference on Lexical and Computational Semantics (*SEM)*, 2014

Tessier, Therese, "Byron and Thomas Moore: A Great Literary Friendship.", *The Byron Journal* 20 (1992): 4658. MetaPress. Web. 22 Jan. 2014.

Vail, Jeffery W., *The Literary Relationship of Lord Byron & Thomas Moore*, Baltimore: Johns Hopkins University Press, 2001. Print.