
Word Vectors in the Eighteenth Century

Ryan James Heuser
heuser@stanford.edu
Stanford University, United States of America

Introduction

This talk explores how new vector-based approaches to computational semantics both afford new methods to digital humanities research, and raise interesting questions for eighteenth-century literary studies in particular. New semantic models known as “word embedding models” have generated excitement recently in the natural language processing and machine learning communities, due to their ability to represent and predict semantic relationships as complex as analogy. “Man” is to “woman” as “king” is to what?, one can ask of the model; “queen,” it will most likely reply. These models formulate analogical and other semantic relationships by computing mathematical vectors for words, such that, if $V(x)$ denotes the vector for the word x , then the above analogy can be expressed as $V(\text{woman}) - V(\text{man}) + V(\text{king}) \approx V(\text{queen})$. Although these models have a longer history—vector space semantics dates from the ‘70s, having been first developed for the SMART information retrieval system (Salton, 1971) by Gerard Salton and his colleagues (Salton et al, 1975)” (Turney and Pantel, 2010)—new innovations in their speed and accuracy (see Note [1]) have renewed researchers’ interests—a development begun, in part, by Google, when researchers there unveiled newly efficient algorithms in 2013, packaged in software they released called word2vec. (The word2vec algorithm was originally described by Mikolov et al, 2013. It introduced the neural network to vector space semantics, providing an efficient means by which to compute word vectors. The GloVe algorithm from the Stanford NLP Group eschews the neural network approach, instead performing a novel method of dimensionality reduction on word collocation counts).

“Word vectors,” as these new methods are sometimes informally called, have already enabled published research into questions relevant to humanistic research, such as a recent landmark paper from researchers in the Stanford NLP Group into patterns of

semantic change across centuries of discourse (Hamilton et al). However, unfortunately, word vectors have so far rarely appeared in research from the digital humanities community itself. Moreover, what work that does exist has so far been primarily circulated through blogs, rather than through published proceedings or articles. Ben Schmidt, for instance, has written an influential introduction to word vectors in his blog post “Vector Space Models for the Digital Humanities” (2015a), which also includes a documented R package for computing them. Also notable is his post, “Rejecting the gender binary” (Schmidt, 2015b), which uses word vectors to dissect the polysemy of words; as well as Michael Gavin’s post, “The Arithmetic of Concepts” (2015), which explores the conceptual implications of adding and subtracting word vectors.

On the whole, the current research landscape of word vectors in the digital humanities resembles the landscape of topic modeling years ago, when the original LDA algorithm (published in 2003 [Blei et al]), before appearing in landmark published DH studies such as Matt Jockers’ *Macroanalysis* (2013), was employed for humanistic research as early as 2006 by researchers working outside or tangentially to the digital humanities (Newman and Block).

Given this scarcity of digital-humanities research on word vectors, work that seeks equally to explain, interpret, and demonstrate their potential seems particularly useful. With these goals in mind, this paper attempts first to unpack for a digital-humanities audience how word vectors work, with reference to the canonical analogy cited above: “man is to woman as king is to queen.” Second, in order to interpret word vectors’ conceptual implications for eighteenth-century literature, I move away from this canonical analogy to one central to a particularly influential argument in the period: “Learning is to Genius as Riches are to Virtue.” Lastly, I turn from this close reading of word vectors to methods of distant-reading analogies that lie implicit in eighteenth-century literature.

Explaining Word Vectors

How do word vectors work? In the interests of space, I have omitted this section of my talk from the abstract. Readers curious about the mechanics of word vectors can read more on my blog, which also links to a number of other explanatory resources (Heuser, “Methods”).

Close-reading Word Vectors

Word vectors provide a persuasive computational means for the semantic representation and analysis of analogies. They combine a mathematical elegance

with an intuitive interpretability to yield what is, potentially, a method useful not only for large-scale semantic analysis, but also for smaller-scale explorations of particular analogies in literature, and their specific forms of analogical argumentation. For instance, analogy lies at the heart of Edward Young’s essay *Conjectures on Original Composition (1759)*, which argued for the superior aesthetic interest of modern, “original” composition over the neoclassical imitation of the ancients. Crucially, Young makes his argument through analogy, identifying several other conceptual contrasts as analogues to his central one between original and imitative composition:

<i>Type of opposition</i>	<i>Associated with original composition</i>	<i>Associated with imitative composition</i>
<i>Attributes of a Poet/Author</i>	Genius	Learning
<i>Forms of social organization</i>	Organic growth	Mechanistic commerce
<i>Forms of social value</i>	Virtue	Riches

Table 1. Table of conceptual analogies leveraged by Edward Young to argue for original over imitative composition (*Conjectures on Original Composition, 1759*).

“I would compare Genius to Virtue, and Learning to Riches,” Young writes; “[a]s Riches are most wanted where there is least Virtue; so Learning where there is least Genius.” In this way, Young’s valuation of “Genius” over “Learning,” and of original over imitative composition, become *ethically justified* through their analogy with another, more obviously moral contrast between “Virtue” and “Riches.”

But what is the logic behind this analogy? Here, word vectors provide the close reader with a framework, language, and method of exploring the semantic implications at work in an analogy. In terms of vectors, we can ask, what does $V(\text{Virtue}) - V(\text{Riches})$ (Also sometimes expressed here, in a shorthand, as $V(\text{Virtue} - \text{Riches})$) mean, and is it in fact correlated with $V(\text{Genius}) - V(\text{Learning})$ in the broader discourse of the period? Asking this question of a word2vec model trained on the 80 million words of eighteenth-century literature in the ECCO-TCP corpus, we find that “Riches” are to “Virtue” as “Learning” is to...

```
In [3]: analogy(model, 'riches', 'virtue', 'learning')
Out[3]:
[(u'morality', 1.0672287940979004),
 (u'piety', 1.0626451969146729),
 (u'science', 1.0292117595672607),
 (u'philosophy', 1.025746345200195),
 (u'prudence', 1.0140740871429443),
 (u'genius', 0.9834112524986267), # <-- 6th closest term
 (u'wisdom', 0.9778728485107422),
 (u'morals', 0.9766285419464111),
 (u'modesty', 0.9748671650886536),
 (u'humanity', 0.972758948802948)]
```

Figure 1. “Riches” are to “Virtue” as “Learning” is to what?, asked of a word2vec model trained on 80 million words of eighteenth-century literature (the ECCO-TCP corpus).

“Genius” is the sixth closest word vector, or the sixth most likely solution, to this analogy. How to test the significance of this result is not immediately clear, but, out of tens of thousands of possibilities, it’s certainly provocative: it raises the possibility that word vectors might provide computational assistance to close readings. Indeed, the other words in this list amplify the semantic profile of this analogy in a way that might help to clarify its underlying implications. For instance, the contrast between the *intrinsic* form of value in “Virtue” and the *extrinsic* form of value in “Riches” seems underscored for me by the contrast here between the extrinsic writerly attribute of “Learning,” associated with an Oxbridge education, and the intrinsic attributes of morality, genius, and wisdom.

Ultimately, however, what does it mean to close-read word vectors? This is a question raised by Gabriel Recchia in a blogpost responding to my interpretation above as it first appeared on my blog (Recchia; Heuser, “Concepts”). Recchia’s post explores other vector operations that even more reliably yield “genius,” namely $V(\text{learning}) + V(\text{virtue})$ and $V(\text{talents}) + V(\text{abilities}) + V(\text{erudition})$. To me, however, these alternative “paths” to genius do not exclude one another; instead, each contributes to our understanding of the semantics of genius in the period. My goal with this interpretation is not to “prove” Young’s analogy, but rather to suggest that, by “amplifying” a particular analogy through its semantic associations across a corpus, word vectors help contextualize our interpretations of particular analogies in literature. As Recchia writes, “the computational exercise has helped us focus our search.”

Distant-reading Word Vectors

If, then, vectors help us explore this *micro*-analytic scale of interpretation, they also help us scale those same interpretive models up to the level of macroanalysis. For instance, inspired by the foregoing close-reading of Young’s complex web of analogies (Table 1),

we might continue Young’s project of obsessive analogization by way of a distant reading. By defining vectors for a range of common eighteenth-century contrasts (Table 2), and then measuring the correlation between them, we can in fact construct *another* complex web of analogies—this time gleaned computationally, from a large-scale archive of the period’s discourse.

Ancient(s) <> Modern(s)	Private <> Public
Beautiful <> Sublime	Romances <> Novels
Body <> Mind	Ruin <> Reputation
Comedy <> Tragedy	Simplicity <> Refinement
Folly <> Wisdom	Tradition <> Revolution
Genius <> Learning	Tyranny <> Liberty
Human <> Divine	Virtue <> Honour
Judgment <> Invention	Virtue <> Riches
Law <> Liberty	Virtue <> Vice
Marvellous <> Common	Whig <> Tory
Parliament <> King	Woman <> Man
Passion <> Reason	

Table 2: Common eighteenth-century contrasts, each expressed as a vector contrast. For instance, Virtue <> Vice denotes the vector $V(\text{Vice}-\text{Virtue})$. Contrasts were gleaned manually while reading Fielding’s *Tom Jones* (1748), as well as a number of essays from the period; they are not meant to be exhaustive. This is an admittedly unsatisfactory method; I am currently exploring ways to discover conceptual contrasts computationally.

Looking at a particularly strong correlation among the contrasts in Table 2, between $V(\text{Simplicity}-\text{Refinement})$ and $V(\text{Virtue}-\text{Vice})$, we can see how their correlation emerges from the way in which both contrasts carry similar semantic associations across the same set of words (Figure 3).

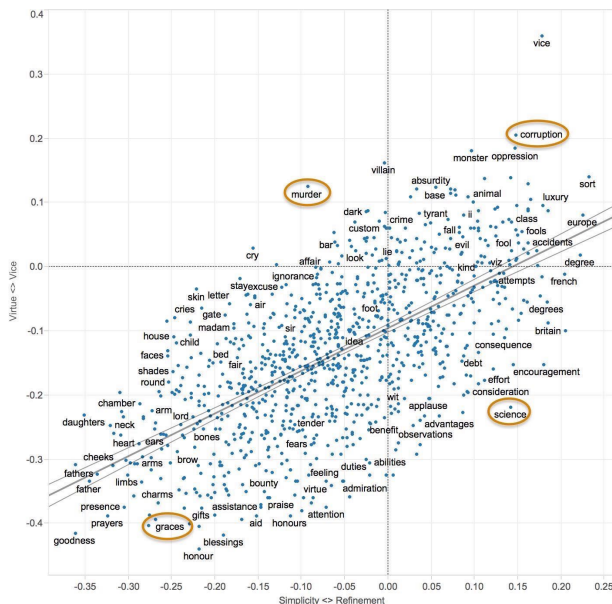


Figure 2. 1,000 most frequent nouns in the ECCO-TCP corpus. On the x-axis is their cosine similarity with the $V(\text{Simplicity}-\text{Refinement})$ vector: if above 0, then associated more with refinement; if below, more with simplicity. Conversely, on the y-axis, above 0 means associated more with Vice; below 0, more with Virtue.

In other words, this graph shows that there are more words for simple virtues (e.g. “graces”) than refined virtues (e.g. “science”), and more words for refined vices (e.g. “corruption”) than simple vices (e.g. “murder”). This correlation between their semantic associations ($R^2 = 0.41$) reveals, then, an analogy emerging from the period’s broader discursive practices—Simplicity is to Refinement as Virtue is to Vice—even as that analogy might appear only implicitly in particular essays, such as in Hume’s “Of Simplicity and Refinement in Writing” (1742), when Hume loosely associates refinement with the moral decline of post-Augustan Rome.

This macro-analytic approach to discovering implicit discursive analogies allows us to visualize the ways in which the frequent conceptual contrasts in eighteenth-century literature are implicitly analogized in the discourse, and how those implicit analogical relationships may have helped to structure what Peter De Bolla has called the “conceptual architecture” of the period (Figure 4).

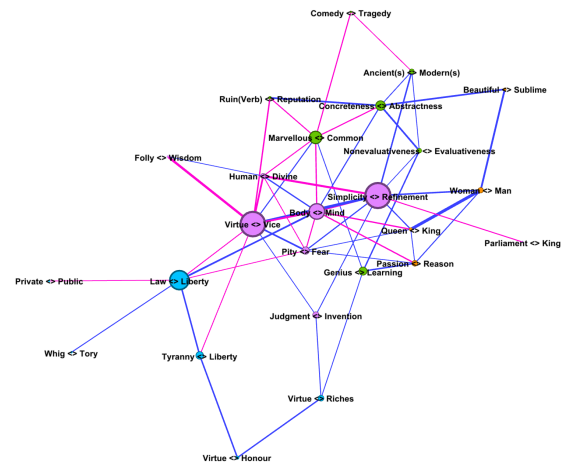


Figure 3. Semantic contrasts are connected in this network if the R^2 value of their correlation, across the 1,000 most frequent nouns (as in Figure 3), is greater than 0.1. Blue lines read in the natural order (e.g. Simplicity is to Refinement as Woman is to Man); red lines read in reverse order (e.g. Simplicity is to Refinement as the King is to Parliament). Nodes are sized by betweenness centrality, and colored by network community. Edges are sized by the R^2 value.

From this network of correlated contrasts, we can see which of them, for instance, are implicitly generated in the period's discourse. "Woman" is to "man," for instance, as "queen" is to "king"—but also as the beautiful is to the sublime, as simplicity is to refinement, and as passion is to reason. Similarly, we can see which contrasts are moralized in the period: "virtue" is to "vice" as wisdom is to folly, as pity is to fear, as the mind is to the body. Moreover, the contrasts of virtue and vice, and simplicity and refinement, might actually play a central role in such a conceptual architecture of analogies, as seen from their centrality within the network.

Conclusion

I hope to have demonstrated some of the ways in which word vectors might be useful for the digital humanities, and particularly for eighteenth-century literary studies, both by demonstrating how they might help us to close-read specific analogical maneuvers, as well as distant-read analogies as they emerge from patterns in their usage across a literary discourse.

Notes

- [1] According to statistics provided in the original paper for the Stanford NLP group's "GloVe," a competing algorithm to word2vec, a word2vec model trained on a large English-language corpus can accurately solve 65% of analogies in a test dataset, and GloVe 75% (Pennington et al, Table 2). As a rough comparison to the accuracy we would expect from human subjects, we might look to the Miller Analogy Test from Pearson—an admittedly unrelated analogy test, which is given to some graduate student applicants. In the MAT of 2002-3, to accurately solve 65% or more of its 100 analogies places a student above the 80th percentile (Pearson). Although not directly comparable, these statistics make more probable the assessment that word vectors are capable of capturing semantic relationships at a level competitive with human subjects.

Bibliography

- Blei, D., Ng, A., and Jordan, M.** (2003). "Latent Dirichlet allocation." *Journal of Machine Learning Research* 3.4-5 (2003): 993-1022.
- Bolla, P. D.** (2013) *The Architecture of Concepts: The Historical Formation of Human Rights*. Fordham UP.
- Newman, D., and Block, S.** (2006). "Probabilistic topic decomposition of an eighteenth-century American newspaper." *Journal of the American Society for Information Science and Technology* 57.6 (2006): 753-767.
- Gavin, M.** (2015) "The Arithmetic of Concepts: a response to Peter de Bolla." *Modeling Literary History*. 19 Sep 2015. Web. Accessed 1 Nov 2016.
- Hamilton, W. L., Leskovec, J., and Jurafsky, D.** (2016). "Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change." arXiv preprint arXiv:1605.09096. Submitted 30 May 2016. Web. Accessed 1 Nov 2016.
- Heuser, R.** (2016a). "Word Vectors in the Eighteenth Century, Episode 1: Concepts." *Adventures of the Virtual*. 14 Apr 2016. Web. Accessed 7 Apr 2017. <<http://ryan-heuser.org/word-vectors-1>>
- Heuser, R.** (2016b). "Word Vectors in the Eighteenth Century, Episode 2: Methods." *Adventures of the Virtual*. 1 Jun 2016. Web. Accessed 7 Apr 2017. <<http://ryan-heuser.org/word-vectors-2>>
- Jockers, M.** (2013) *Macroanalysis: Digital Methods and Literary History*. U of Illinois P.
- Mikolov, T, et al.** (2013) "Distributed representations of words and phrases and their compositionality." *Advances in neural information processing systems* 26 (2013).
- Pearson.** (2003) "Candidate Information Booklet." *Miller Analogies Test*. Web. Accessed 1 Nov 2016. <http://images.pearsonclinical.com/images/pdf/milleranalogies/matcib2002_03.pdf>
- Pennington, J., Socher, R., and Manning, C.** (2014). "Glove: Global Vectors for Word Representation." *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (2014), 1532-1543.
- Recchia, G.** (2016) "'Numberless Degrees Of Similitude': A Response To Ryan Heuser's 'Word Vectors In The Eighteenth Century, Part 1.'" *Gabriel Recchia's Blog*. 11 Jun 2016. Web. Accessed 7 Apr 2017. <<http://www.twonewthings.com/gabrielrecchia/2016/06/11/numberless-degrees-of-similitude-word-vectors/>>
- Řehůřek, R.** (n.d.) "models.word2vec - Deep learning with word2vec." *gensim*. Web. Accessed 1 Nov 2016. <<https://radimrehurek.com/gensim/models/word2vec.html>>
- Salton, G.** (1971). *The SMART retrieval system: Experiments in automatic document processing*. Prentice-Hall.

Salton, G., Wong, A., and Yang, C. (1975). "A vector space model for automatic indexing." *Communications of the ACM* 18.11, 613-620.

Schmidt, B. (2015a). "Vector Space Models for the Digital Humanities." *Bookworm*. 25 Oct 2015. Web. Accessed 1 Nov 2016.

Schmidt, B. (2015a).. "Rejecting the gender binary: a vector-space operation." *Bookworm*. 30 Oct 2015. Web. Accessed 1 Nov 2016.

Turney, P. D. and Pantel, P. (2010) "From Frequency to Meaning: Vector Space Models of Semantics." *Journal of Artificial Intelligence Research* 37: 141-188.