

The emergence of the modern concept of introspection: a quantitative linguistic analysis

I. Raskovsky

Department of Computer Science
University of Buenos Aires
Pabellón I, Ciudad Universitaria
Buenos Aires, C1428EGA, Argentina
iraskovsky@dc.uba.ar

D. Fernández Slezak

Department of Computer Science
University of Buenos Aires
Pabellón I, Ciudad Universitaria
Buenos Aires, C1428EGA, Argentina
dfslezak@dc.uba.ar

C.G. Diuk

Department of Psychology
Princeton University
Princeton, NJ 08540, USA
cdiuk@princeton.edu

G.A. Cecchi

Computational Biology Center
T.J. Watson IBM Research Center
Yorktown Heights, NY 10598, USA
gcecchi@us.ibm.com

Abstract

The evolution of literary styles in the western tradition has been the subject of extended research that arguably has spanned centuries. In particular, previous work has conjectured the existence of a gradual yet persistent increase of the degree of self-awareness or introspection, i.e. that capacity to expound on one's own thought processes and behaviors, reflected in the chronology of the classical literary texts. This type of question has been traditionally addressed by qualitative studies in philology and literary theory. In this paper, we describe preliminary results based on the application of computational linguistics techniques to quantitatively analyze this hypothesis. We evaluate the appearance of introspection in texts by searching words related to it, and focus on simple studies on the Bible. This preliminary results are highly positive, indicating that it is indeed possible to statistically discriminate between texts based on a semantic core centered around introspection, chronologically and culturally belonging to different phases. In our opinion, the rigorous extension of our analysis can provide not only a stricter statistical measure of the evolution of introspection, but also a means to investigate subtle differences in aesthetic styles and cognitive structures across cultures, authors and literary forms.

1 Introduction

The evolution of literary styles in the western tradition has been the subject of extended research that arguably has spanned centuries. In particular, previous work has conjectured the existence of a gradual yet persistent increase of the degree of self-awareness or *introspection*, i.e. that capacity to expound on one's own thought processes and behaviors, reflected in the chronology of the classical literary texts. This type of question has been traditionally addressed by qualitative studies in philology and literary theory. In this paper, we describe preliminary results based on the application of computational linguistics techniques to quantitatively analyze this hypothesis.

The striking differences between the Iliad and the Odyssey in the way the characters' behaviors are attributed to divine intervention, or to the individual's volition, has been pointed out by numerous scholars (Onians, 1988; Dodds, 1951; Adkins, 1970; De Jong and Sullivan, 1994). However, not until the highly influential work of Marshall McLuhan (McLuhan, 1962) and Julian Jaynes (Jaynes, 2000) was it pointed out that these changes may reflect not just artistic or even cultural tendencies, but profound alterations in the mental structure of those who wrote, collected and assimilated the stories. While

McLuhan argued for a materialistic effect of the type of medium (the linearity of written language, the holistic nature of the moving image) on the organization of thoughts (linear or integrative, respectively), Jaynes proposed a more radical hypothesis: a relatively abrupt transition from a “bicameral mind”, where one hemisphere produced god-like commands that the other followed blindly, to the modern mind with its ability of self-awareness. Moreover, Jaynes boldly suggested that this transition may have been accompanied by a physical process that altered the relationship between the hemispheres, and changed culture permanently. Since its publication, *The origins of consciousness in the breakdown of the bicameral mind* has been highly influential inside and outside scientific quarters, as well as a source of continuing controversy (Cavanna et al., 2007).

Whether brought about by nature or nurture, however, Jaynes presents compelling arguments about the effects of this transition, including stylistic changes throughout the other foundational text of the western world, the Bible. Simply put, a less radical version of Jaynes’ hypothesis would state that, within the judeo-greco-christian cultural tradition, there exists an “arrow of time” pointing to increasing *introspection*. The question we set out to answer in the present manuscript is to what extent it is possible to analyze, quantitatively, this hypothesis.

The widespread availability of classic and modern literary texts has paved the road to a wide variety of linguistic studies. Matters of literary style and structure are necessarily more controversial, although the recent work of F. Moretti (Moretti, 2005) has shown that it is indeed possible to quantify the subtle variations in the structure of the novel over temporal periodizations and geographical locations. In any event, given that our intention is to complete a preliminary study of feasibility, we focus here on capturing the textual traces of words or lexical structures that can be reasonably argued to reflect introspective thinking on the part of the characters, using techniques from machine learning and computational linguistics.

2 Materials and methods

We downloaded selected texts representative of different ages in literature from the MIT classic texts archive (Daniel C. Stevenson, 2010), based on references in Jaynes’ book (Jaynes, 2000). The selected texts are: the Iliad and the Odyssey (approx. 1200 BC to 900 BC), The Bible (approx. 1400 BC to AD 200), Lucretius’ On the Nature of Things (99 BC - 55 BC), St. Augustine’s Confessions (AD 397 - AD 398), Shakespear’s The Merchant of Venice (AD 1596 - AD 1598), Hamlet (approx. AD 1600), Macbeth (AD 1603 - AD 1607) and Othello (AD 1603), Cervantes’ Quixote (AD 1605 - AD 1615), Jean Austen’s Mansfield Park (AD 1814), Emma (AD 1815) and Persuasion (AD 1816) and Proust’s Time Regained (AD 1927).

On this preliminary study, we focused on extremely simple techniques to test our hypothesis. We have implemented a series of basic routines to analyze the frequency of certain words related to introspection, selected by hand. We used very simple regular expressions to search over the text: `think+`, `thought`, `myself`, `mind+`, `feel+` and `felt`. The search was conducted on 10,000-words windows starting from the beginning of the text moving towards the end in 2,000-words steps. Also, the appearance of references to God in the Bible was measured. In this case, we looked for: `lord`, `god` and `almighty`; all searches done case insensitive. In order to control for the possible increase of these selected words as a trivial consequence of an increase in the overall linguistic richness or expressivity of the text, we also computed the total number of distinct words for each step.

As an alternative approach, we applied a data-driven method to extract the semantic structure of texts, namely topic modeling (Blei, 2009). We utilized the implementation of the *mallet* package (McCallum, 2002), an off-the-shelf tool, generating 100 topics through 10,000 Gibbs sampling rounds. The topics were then manually inspected for their semantic relevance to the issue at hand, i.e. introspection.

3 Results

The preliminary results are highly positive, indicating that it is indeed possible to statistically discriminate between texts based on a semantic core centered

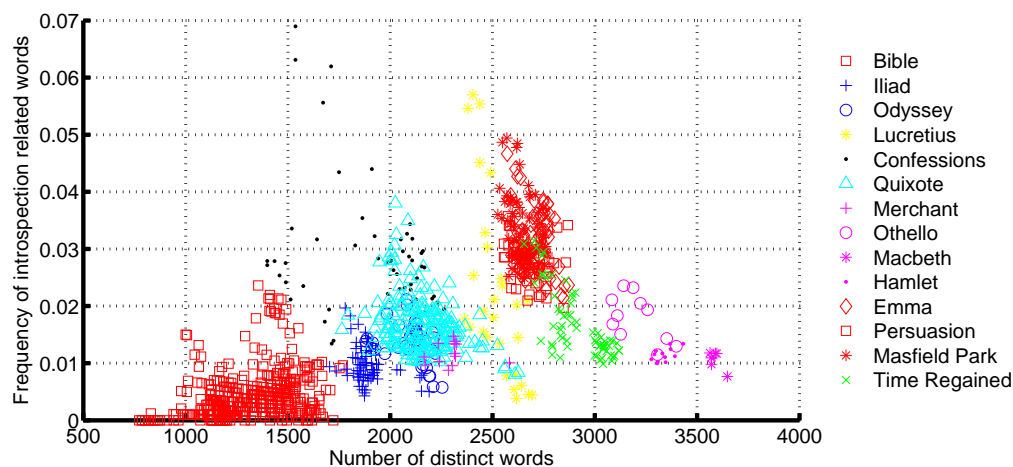


Figure 1: Frequency of words related to introspection versus the amount of different words. Each text is identified by a unique color; each point represents a 10,000-words window.

around introspection, chronologically and culturally belonging to different phases. In figure 1 we show the frequency of words related to introspection versus the amount of different words, for all the texts we chose. Each text is identified by a unique color; each point represents a 10,000-words window. To summarize this information and provide statistical value to our analysis, we present in figure 2 the mean and standard deviation for each of the selected texts.

We clearly observe how different texts are disjoint in the graph, both in amount of different words used in each window, as well as the frequency of introspection. This is the case for the *Iliad* and the *Odyssey*, confirming that our preliminary measure captures the semantic differences between both pieces. We also observe a trend between some of the sections of the Bible (we will return to this below), to the Homeric texts, Lucretius, Cervantes and Austen, roughly following a chronological order. St. Augustine’s *Confessions* is an exception of this trend, as it shows a higher frequency than Cervantes. However, given that *Confessions* is considered the first auto-biography work in the Western tradition, the high value of our introspection measure is to some extent a validation of its pertinence. A more noticeable exception is Shakespeare’s oeuvre, that seems to consist of very differentiated clusters for each piece. Taken together, however, the ensemble average of his work seems to fall in line with the global temporal order. It is beyond the scope of our manuscript to discuss the nuances of the work

of *The Bard*, but our analytic approach may provide new tools to the ongoing Shakespearean scholarship. Finally, our analysis seems to really break down for Proust, as one intuitively would expect a much higher measure of introspection, even more so considering that he displays significant richness in terms of the number of distinct words in the text. This failure is clearly an indication of the limitations of our current approach, which as it stands may only not be applicable to modern or contemporary literature.

The Bible is of particular interest for this work, as it was written in parts along a wide time interval (taking into account the Old and New Testaments). It enables us to analyze the “arrow of time” of introspection within a relatively coherent framework, even though a vast and in most cases unknown host of writers and compilers gave this text its present shape, and the relationship between textual linearity and chronological order is certainly not simple. Be it as it may, for our purposes we only require that this relationship be monotonic in a statistical sense, which we assume to be the case for the Bible. In figure 3 we show the frequency of introspection along different *periods* of the Bible. The text was divided into 6 pieces of the same length; purposefully, no semantic division was performed. Introspection increases towards the more modern sections of the text, confirming our initial hypothesis. Note that the lexical richness (number of distinct words) of the last period, that includes the New Testament, seems

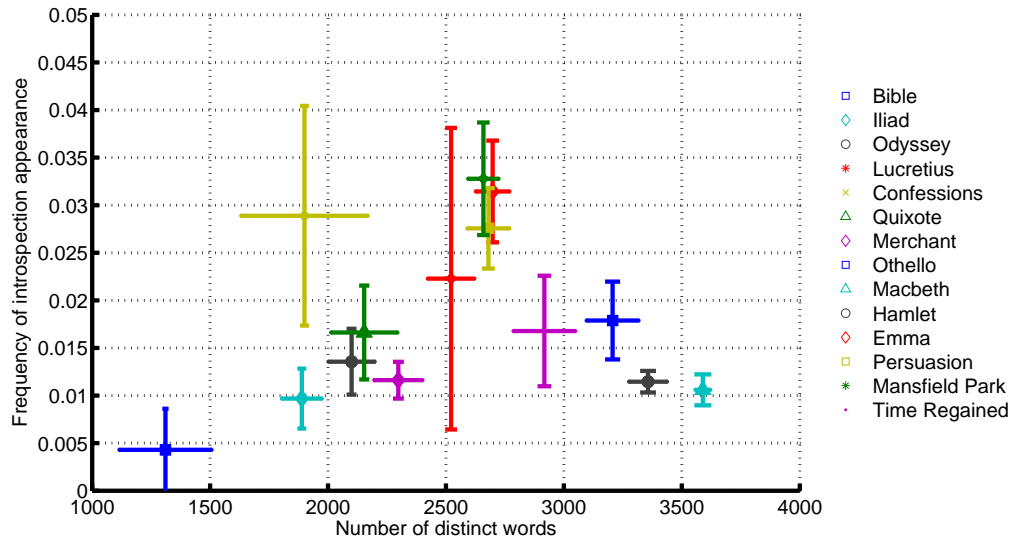


Figure 2: Frequency of words related to introspection versus the amount of different words. Each text is identified by a unique color. Error bars shows the standard deviation in both axis.

to be part of the same plateau as the 3rd, 4th and 5th. Another interesting aspect is how introspection and citations to God evolve along the lexical, pseudo-chronological order the the Bible’s test. Figure 4 shows the appearance of introspection and mentions to God over 10,000-words windows. We observe a significant increase in introspection as the lexical order moves ahead, while at the same time the references to God show a weaker, yet clearly discernible trend to diminish. This is precisely the analytic counterpart of the phenomenology on which Jaynes based his hypothesis.

Another avenue of research involves developing and applying more sophisticated tools of textual analysis, in order to capture the presence of relevant passages of introspection using data-driven approaches. In particular, we have done a preliminary study using Topic Modeling (Blei, 2009), a technique that uses probabilistic models to uncover the underlying semantic structure of a text or a collection of documents. In topic modeling, a *topic* is a collection of keywords that are automatically extracted as highly descriptive of a document. As described in the Methods section, we utilized the *mallet* package implementation, choosing 100 topics to be uncovered. This package produces *approximate* inference and therefore different runs may yield different results. In each run, we observed a handful of topics (between 3 and 6, approximately) that

contained words related to introspection, such as the *mind*, *think* and *feel* roots. The following topic, selected as a representative from one of the *mallet* runs, was identified based only on the presence of *mind*, although some of the other words may also be relevant for the purpose of revealing introspective activity (*soul* and *desire*):

soul	love	yea	desire
mind	hate	sought	loveth
measure	fair	pleasant	nay
keepeth	hungry	satisfied	excellent
occasion	rejoicing	desired	

Figure 5 presents the frequency with which this topic of interest is considered the main topic by *mallet*, as a running average for every 100 lines of the Bible. A simple linear regression shows that this topic becomes more frequent towards the end of the text, and mirrors the results obtained with the more hand-crafted approach. This result, while preliminary (there is a good number of parameters to explore in setting up topic modeling), is highly promising, as topic modeling provides a link with the vast literature of statistical semantic analysis.

4 Conclusions

Previous work in the evolution of literary styles in the Western tradition has conjectured the existence of a gradual yet persistent increase of the degree of

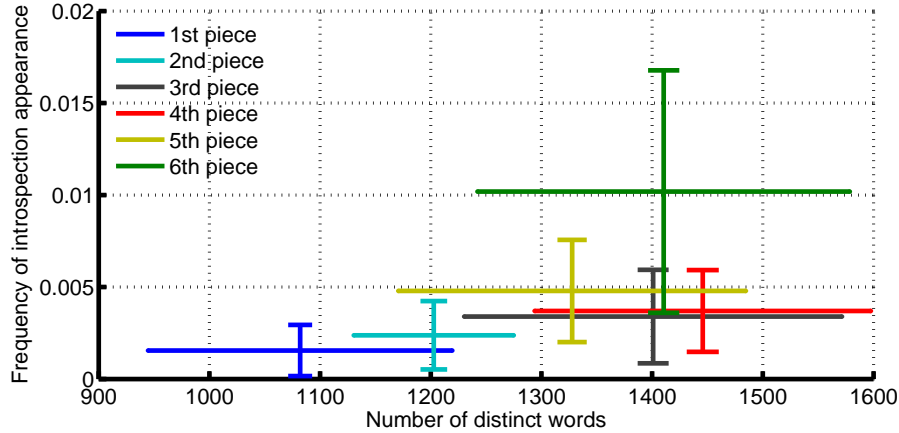


Figure 3: Frequency of words related to introspection versus the amount of different words in the bible, divided in 6 pieces of the same length.

introspection. In particular, the ideas of Marshall McLuhan and Julian Jaynes suggest the hypothesis that these changes reflect profound and permanent alterations in the cognitive structures of the culture. We set out to investigate to what extent it is possible to analyze, quantitatively, this hypothesis. We focused on Homeric and Biblical texts, given their cultural preeminence, and utilized relatively simple analytic techniques to measure the degree of introspection along the texts, assuming they reflect, however imperfectly, a certain chronological order.

The result of measuring word frequencies is highly positive, indicating that it is indeed possible to statistically discriminate between texts based on a semantic core centered around introspection, chronologically and culturally belonging to different phases. We clearly observe how different texts are disjoint in the graph, both in amount of different words used in each window, as well as the frequency of introspection. However, our analysis seems to really break down for Proust, as one intuitively would expect a much higher measure of introspection. This failure is clearly an indication of the limitations of our current approach, which as it stands may only not be applicable to modern or contemporary literature.

Our analysis on the Bible is of particular interest. It enables us to analyze the “arrow of time” of introspection within a relatively coherent framework, and the relationship between textual linearity and chornological order is certainly not simple. In-

trospection increases towards the the more modern sections of the text, confirming our initial hypothesis. Another interesting aspect is how introspection and citations to God evolve along the text. We observe a significant increase in introspection as the lexical order moves ahead, while at the same time the references to God show a weaker, yet clearly discernible trend to diminish.

As an alternative approach, in order to capture the presence of relevant passages of introspection using a data-driven method, we applied topic modeling, an off-the-shelf tool. We observed a handful of topics (between 3 and 6, approximately) that contained words related to introspection, such as the `mind`, `think` and `feel` roots. The analysis of the bible under this technic mirrors the results obtained with the more hand-crafted approach.

While the analysis presented can only be considered an initial step towards a systematic characterization of the textual correlate of the concept of introspection, the simplicity of our methods and the clarity of the results support our initial hypothesis, and validate our approach. In our opinion, the rigorous extension of our analysis can provide not only a stricter statistical measure of the evolution of introspection, but also a means to investigate subtle differences in aesthetic styles and cognitive structures across cultures, authors and literary forms (i.e. the novel, cf. (Moretti, 2005)).

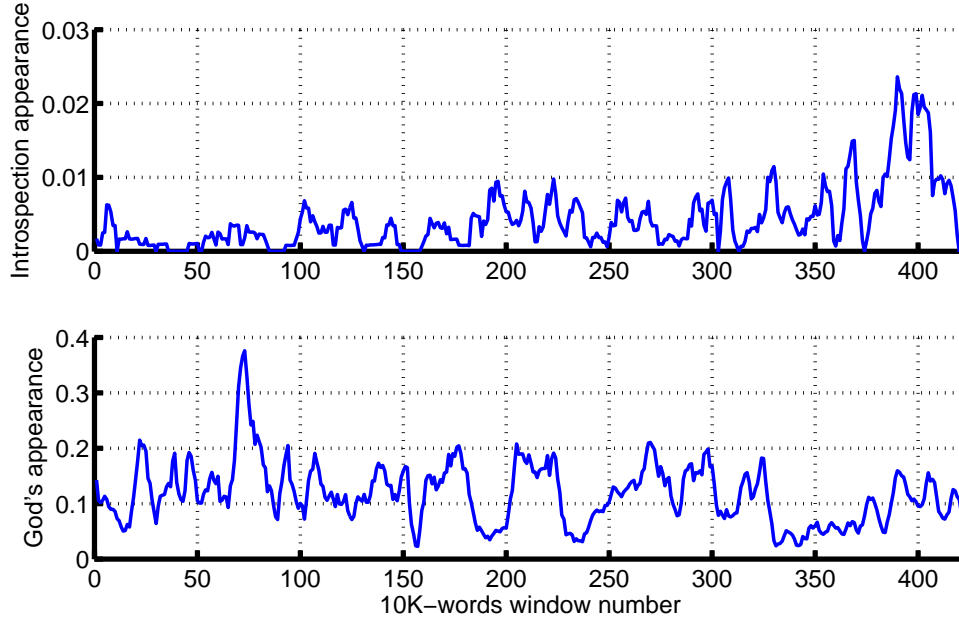


Figure 4: Frequency of words related to introspection and God versus the 10,000-words window number.

5 Outlook

Given the necessarily broad, integrative nature of any approach to introspection, there is a number of different alternatives we are currently exploring to expand our analysis, with an emphasis on interdisciplinary perspectives.

A first step is systematizing the routines for filtering, processing and analysis of the texts. We will incorporate more terms related to introspection captured in the structure of Wordnet, and eventually also incorporate as part of the analysis elements of the graphical structure that underlies this database (Sigman and Cecchi, 2002). Some interesting developments in this area are the measures of semantic similarity between concepts (Budanitsky and Hirst, 2006; Pedersen et al., 2004; Patwardhan et al., 2003). This measure may result useful for classifying the different topics acquired using topic modeling, taking into account the *similarity* of the words related to introspection, as an extension to the semantic relationships established by Wordnet and the various dictionaries and thesaurii currently available as databases.

We will also incorporate the notion of *concept drift* to our topic modeling, expecting it to account for the temporal evolution of the use of introspec-

tion. A promising proposal for this purpose is that of Dynamic Topic Modeling. We are particularly interested in approaches that require minimal a priori intervention; we expect that a dynamic model with an unconstrained number of topics, as opposed to the fixed number of topics proposed in the original paper (Blei and Lafferty, 2006), may lead more naturally to the identification of potential transitions along the text. This approach is not straightforward to implement, and may require the development of an appropriate statistical model.

Another step will be a more careful and principled selection and categorization of our text corpus. While the techniques at hand enable the analysis of massive amounts of data, we will select our texts based on their cultural and historical relevance in a more systematic way. Comparing different cultures and ages results in an interesting challenge. We are specifically interested in the replication of the results presented here in the case of the aboriginal American cultures. The concept of introspection appears in many classic American texts such as the Popol Vuh and the Chilam Balam; however, their compilation by European scholars and translation to different languages may not keep the essence of the original texts. A robust systematization of our technique

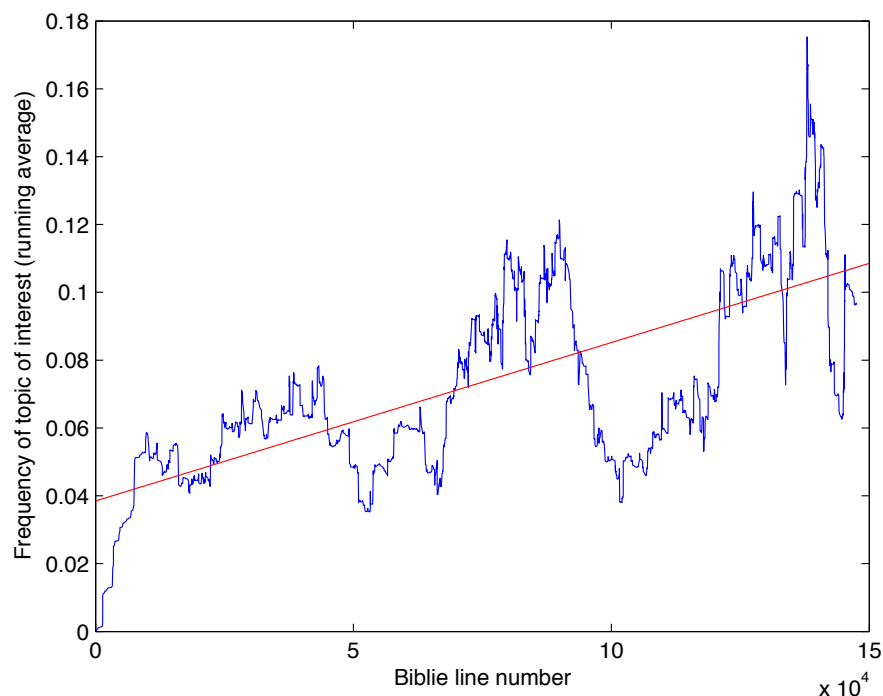


Figure 5: Frequency of Introspection topic as the main topic for each section of the Bible, as a running average every 100 lines.

will allow us to analyze texts in different languages easily. We look forward to compare the measures of introspection between texts in their native languages in contrast with its appearance in their translations. Moreover, this may help with the conservation of selected concepts along translations of texts. This line of research will require the interaction with experts in early American philology.

Finally, it is important to note that the analytical techniques proposed here, namely the quantification of psychological concepts embedded in the text, can be used as tools for pedagogical and psychiatric evaluation (Lombardo et al., 2007). This will require a concerted effort with psychologists and psychiatrists to collect and organize personal narrations by patients, as well as the compilation of texts already available in the literature, in particular by people suffering from schizophrenia and depression.

In summary, we believe the results presented here will provide a rich source of multi-disciplinary follow-up and derived lines of research around statistical measurements of psychological features in text, within and beyond the concept of introspection.

References

- Adkins, A. (1970). *From the many to the one: a study of personality and views of human nature in the context of ancient Greek society, values and beliefs*. Constable & Company Limited.
- Blei, D. (2009). *Text Mining: Theory and Applications* (A. Srivastava and M. Sahami editors), chapter Topic Models. Taylor and Francis.
- Blei, D. and Lafferty, J. (2006). Dynamic topic models. In *ICML '06: Proceedings of the 23rd international conference on Machine learning*, pages 113–120, New York, NY, USA. ACM.
- Budanitsky, A. and Hirst, G. (2006). Evaluating wordnet-based measures of lexical semantic relatedness. *Computational Linguistics*, 32(1):13–47.
- Cavanna, A., Trimble, M., Cinti, F., and Monaco, F. (2007). The “bicameral mind” 30 years on: a critical reappraisal of Julian Jaynes hypothesis. *Functional Neurology*, 22(1):11–15.
- Daniel C. Stevenson, W. A. (last accessed: Feb. 27th, 2010). Mit classics. <http://classics.mit.edu/>.
- De Jong, I. and Sullivan, J. (1994). *Modern critical theory and classical literature*. Brill Academic Pub.

- Dodds, E. (1951). *The Greeks and the Irrational*. Berkeley: Univ. of California Press.
- Jaynes, J. (2000). *The origin of consciousness in the breakdown of the bicameral mind*. Mariner Books.
- Lombardo, M., Barnes, J., Wheelwright, S., and Baron-Cohen, S. (2007). Self-Referential Cognition and Empathy in Autism. *PLoS ONE*, 2(9):e883.
- McCallum, A. (2002). Mallet: A machine learning for language toolkit. <http://mallet.cs.umass.edu>.
- McLuhan, M. (1962). *The Gutenberg galaxy: The making of typographic man*. Routledge & Kegan Paul.
- Moretti, F. (2005). *Graphs, Maps, Trees: Abstract models for a literary history*. Verso Books.
- Onians, R. (1988). *The origins of European thought about the body, the mind, the soul, the world, time, and fate: new interpretations of Greek, Roman and kindred evidence also of some basic Jewish and Christian beliefs*. Cambridge University Press.
- Patwardhan, S., Banerjee, S., and Pedersen, T. (2003). Using measures of semantic relatedness for word sense disambiguation. *Lecture notes in computer science*, pages 241–257.
- Pedersen, T., Patwardhan, S., and Michelizzi, J. (2004). Wordnet:: similarity-measuring the relatedness of concepts. In *Proceedings of the National Conference on Artificial Intelligence*, pages 1024–1025.
- Sigman, M. and Cecchi, G. (2002). Global organization of the wordnet lexicon. *Proceedings of the National Academy of Sciences*, 99(3):1742.