

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Nebraska Anthropologist

Anthropology, Department of

2012

Statement Analysis of the Dunbar Letters to the American Board of Commissioners for Foreign Missions

Amanda F. Callahan-Mims
University of Nebraska-Lincoln

Follow this and additional works at: <http://digitalcommons.unl.edu/nebanthro>



Part of the [Arts and Humanities Commons](#)

Callahan-Mims, Amanda F, "Statement Analysis of the Dunbar Letters to the American Board of Commissioners for Foreign Missions" (2012). *Nebraska Anthropologist*. 173.
<http://digitalcommons.unl.edu/nebanthro/173>

This Article is brought to you for free and open access by the Anthropology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Nebraska Anthropologist by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Statement Analysis of the Dunbar Letters to the American Board of Commissioners for Foreign Missions

Amanda F. Callahan-Mims

Abstract: Statement analysis is used by forensic investigators to evaluate the truthfulness of written and verbal statements. Currently, it is most commonly used on contemporary documents to search for deception; often, these inquiries focus on persons under investigation for a crime. Recently, the use of statistical formulae was introduced to the field of forensic linguistics. Although initially limited primarily to law enforcement, implementation of forensic procedures such as statement analysis has since produced successful results within the field of archaeology. It thus seems likely that linguistic analysis may produce favorable results when applied to historic documents.

Introduction

The purpose of this research is to evaluate letters and journal entries written by John Dunbar between 1834 – 1849 using statistical analysis. The authenticity of these letters and journal entries is important, because Dunbar's work has been cited at least twenty-eight times by scholars. Though not intentional, John Dunbar provides much of what is known about the Contact period Pawnee. Among other things, his letters contain valuable information about migration, organization, hunting and subsistence strategies, social and religious order, processing techniques, information on Indian agents, census counts, daily activities, division of labor, and epidemics. The main goals of this research are to determine whether statistical linguistic analysis of historical documents is a valid method to establish authorship of such texts, or to identify potential transcription errors within the letters. Additionally, it would be beneficial to identify any temporal trends that may be present, and to account for changes in consistency over the fifteen year period of the letters.

Background

Statement analysis is used by forensic investigators to evaluate the truthfulness of written and verbal statements. Currently, it is most commonly used on contemporary documents to search for deception; these inquiries typically focus on persons under investigation for a crime. The examination of written documents is often the easiest, most readily available source of data during such inquiries. There are several types of analysis that focus on different aspects of the document. Often, the eight parts of speech are evaluated, as well as: verb tense, key words, passiveness, vernacular, and resistance (Adams 1996). Recently, statistical analyses using linear regression were introduced to the field of forensic linguistics. Commonly used methods include: standard deviations from the mean, the difference of the means, Chi-square test, and the t test (Olsson 2004).

The implementation of such procedures has produced successful results within the field of archaeology, and vice versa, within the field of forensic linguistics. Utilizing a multidisciplinary approach allows the investigator to employ methods that yield data that can be analyzed in a myriad of ways. Given the successfulness of linguistic analysis within the law enforcement community, it seems likely the procedure may produce favorable results when applied to historic documents.

The Choice of R squared as the formula for Analysis

Because language, sentence length, spelling, punctuation and grammar can vary depending on location, education level, and historical timeframe, any evaluation that involves the eight parts of speech, verb tense, key words, passiveness, vernacular and resistance is likely to produce biased results when applied to historical texts. The remaining methods that involve statistical formulae require an equal number of words, sentences, or paragraphs for a valid comparison. This limitation is problematic when the underlying historical documents, such as those attributed to Dunbar, are of different lengths.

Linear regression techniques using R squared values predict how effective one variable is at predicting another. In contrast to the other methods discussed above, R squared does not require equal-length documents, nor does it rely upon linguistic anomalies that may vary by location, education level, or historical timeframe. As a result, R squared analysis appears to be the best method for analyzing Dunbar's letters and journal entries to interpret consistency, or "fit."

Research Goals

The purpose of this research is to evaluate letters and journal entries written by John Dunbar between 1834 – 1849. The validity of these letters and journal entries is important because Dunbar's work has been cited at least twenty-eight times by scholars. Because the letters were written over a hundred and eighty years ago, and then transcribed, it is impossible to know if the transcription was performed without interference from the transcriber. It is important to establish that the data in the letters and journal entries remained as John Dunbar wrote it, because it is so commonly used as a primary source of information.

Though not intentional, John Dunbar provides much of what is known about the Contact period Pawnee. His letters contain valuable information about a range of topics, including migration, organization, hunting and subsistence strategies, social and religious order, processing techniques, information on Indian agents, census counts, daily activities, division of labor, and epidemics. The letters and journal entries contain the only long term observations about the Pawnee during this critical period.

The main goals of the research are to determine if the linguistic procedure of utilizing R squared analysis on historical documents is a valid method to establish authorship of historic texts. Additionally, it would be beneficial to identify potential temporal trends, and to account for changes in consistency over the fifteen year period of the letters. Consistency in writing style will be tested by analyzing the average words per sentence per paragraph to evaluate overall uniformity through the letter.

About the texts

This analysis relies on two historical texts. The first is the Dunbar - Allis letters on the Pawnee, which were originally published by the Kansas State Historical Society, volume 14, in the 1915 - 1918 edition. In 1985, the Historical Society reprinted the letters with a forward by Waldo Wedel. In addition to the Dunbar letters, the 1985 text contained letters and journal entries from the two Missionaries to the American Board of Commissioners for Foreign Missions (ABCFM). The second text analyzed is a short book entitled History of the American Mission to the Pawnee, which was published by the Massachusetts Sabbath School Society in 1838. Written in a conversational format, the text is styled as a conversation between Hugh Clifford and his cousins Ann, Helen and Robert Barton.

Hypothesis

The expectation is that the average number of sentences per paragraph will correlate with the number of paragraphs per letter or journal entry. One of two outcomes was predicted to occur. The first prediction is that the average number of words per sentence per paragraph would be fairly consistent; resulting in an R squared value of close to one. The second prediction is that the number of words is not similar, resulting in a bell curve trend. This would occur, for example, when the first and last paragraphs were not only shorter, but the average sentence length was also shorter. This prediction is based on the tendency to open a letter with a quick introduction and close with a brief summary of the contents.

The ABCFM

The first American Christian foreign ministry, ABCFM, or the American Board of Commissioners for Foreign Missions, was officially chartered in 1812. In addition to its work in the Americas, ABCFM missionaries worked in India, Sri Lanka, Hawaii, China, Singapore, Thailand, Greece, Cyprus, Turkey, Syria, Iran, and Africa. It had the first unmarried female missionary to the American Indians, and also overseas (Golossanov 2006). Jeremiah Evart, who served as treasurer and secretary, led ABCFM's efforts to oppose Indian Removal policies, including the Indian Removal Act of 1830 (Andrew 1992).

About John Dunbar

John Dunbar was born in Massachusetts in 1804. He graduated from Williams College at age 28 and was ordained as a minister at age 30. It was through the Auburn Theological Seminary that he offered himself as a candidate to go west across the Rocky Mountains as a missionary to the Indians. However, it was not until another candidate, G. W. Schuyler, declined that Dunbar was selected for this position (Wedel 1985).

Originally, Dunbar intended to work with the Nez Perce. Upon his arrival in St. Louis, he and his companion Samuel Allis learned the tribe had already departed in order to make it across the mountains safely. Another opportunity to cross them would not occur until the next year. In July of 1834, Dunbar suggested that he and Allis provide mission services to the Pawnee. Although the Baptist and Methodist missionaries in Leavenworth had intended to set up missions there, they

instead deferred to Dunbar and Allis, who were Presbyterian. In September of 1834, Dunbar and Allis were introduced to the Pawnee (Wedel 1985).

Dunbar returned to Massachusetts in 1836 to meet with the ABCFM, marry his fiancé and accompany her back to the mission. While there, he published a 74 page booklet about the Pawnee (Wedel 1985). In September of 1839, a site was selected to fulfill the 1833 Treaty between the Pawnee and the United States (Wedel 1985). Unsuccessful attempts to continue their missionary efforts lead to the mission's dissolution in August of 1846. John Dunbar and his family moved to Holt County, Missouri (Wedel 1985). They lived there about ten years, and then moved to Kansas. John Dunbar was appointed the first treasurer of Brown County. His wife died just after they arrived in Kansas, and he died about a year later (Wedel 1985).

Text 1 – Methodology: The Dunbar-Allis Letters on the Pawnee

The range of letters considered for analysis includes only those written during the time that Dunbar and Allis were actively engaged as missionaries (i.e., traveling and living with the Pawnee). Given this requirement, the 1831-1833 letters were omitted, because Dunbar was not an active missionary during that time. Although the goal was to randomly select letters from each relevant year, it was not possible to completely randomize the process. In particular, for three years, only one letter was sent; accordingly, during those years (1841, 1842, and 1849), the selection process could not be randomized because there was only one choice. Similarly, for three years (1837, 1847, and 1848), no letters were sent, so there were no options from which to choose. Therefore, the selection was limited to thirteen letters over fifteen years (see Table 1).

Year:	Dunbar:
1831	1
1832	0
1833	0
1834	6
1835	2
1836	4
1837	0

1838	2
1839	4
1840	4
1841	1
1842	1
1843	2
1844	4
1845	3
1846	3
1847	0
1848	0
1849	1

Table 1. Number of letters written per year.

Dealing with the Data: The Dunbar-Allis Letters on the Pawnee

In order to make the data more manageable, certain criteria were adopted to make the process both easy to understand and replicate. Blue lettering denoted words not used toward the end count, and different colored numbers represent different word counts. Specifically, red numbers indicated the word counts for each sentence; green numbers indicate the total words per paragraph; and purple numbers indicate the number of sentences per paragraph.

Total word counts exclude the following: Salutations/ Greetings, Closing statements, signatures, Headings/addresses, lists (including census figures), sentences with quotations, notes from the transcriber (including insertions, post scripts and edge text). Edge text is defined as text written in the margin of the letter, denoted by the transcriber as written across the page.

Results: The Dunbar-Allis Letters on the Pawnee

After sentence and paragraph counts were calculated for each of the thirteen letters, the resulting data was averaged and used to create scatter plot graphs, to which linear regression trend lines were applied. R squared values were then calculated for each letter, and compared across the thirteen letters.

Initially, the results appeared to show consistency. For example, Letter 1, from 1834, had an R square value of 83 percent, while the final letter (Letter 13) had an R square value of 52 percent. All other letters, however, had a value of less than 35 percent, including five letters that had an R square value of less than eight percent, denoting little to no fit between the average sentence length per paragraph not only within the letters, but also among the letters. Overall, there was no consistency in eleven out of the thirteen, or 85 percent, of the letters (see Table 2).

Letter #	R Squared Value
Letter 1	0.828
Letter 2	0.082
Letter 3	0.004
Letter 4	0.214
Letter 5	0.002
Letter 6	0.0009
Letter 7	0.321
Letter 8	0.332
Letter 9	0.227
Letter 10	0.147
Letter 11	0.015
Letter 12	0.193
Letter 13	0.515

Table 2. R squared values for selected letters.

As noted, evaluating trends within the letters was also an objective of the research. Because Letter 1, from 1834, exhibited the highest overall consistency, it was selected to evaluate internal letter trends. Again, after fitting linear regression trend lines, only one of the five paragraphs had a rate of fit over 50 percent. The remaining four paragraphs had a very low rate of fit (under seven percent). Based on

these results, there is no evidence of predictability of average sentence length, either internally within the letters or across the entire sample (see Table 3).

Letter 1	R squared value
P1	0.033
P2	0.065
P3	0.551
P4	0.021
P5	0.0005

Table 3. R squared values for Letter 1.

Although the results of the data show little to no consistency, this does not necessarily mean that the letters are a bad source of data, or that they were altered by the transcriber. Rather, it may simply mean that an R squared analysis is not a good method of linguistic analysis for historic documents. For example, the resulting lack of apparent trends and little to no predictability within and among the letters could be due to the state of the writer during periods of letter writing, or to any number of other factors not accounted for within existing linguistic methods.

Text 2 – Methodology: History of the American Mission to the Pawnee Indians

The second text to be analyzed, a short book entitled the History of the American Mission to the Pawnee (hereinafter, History), was published in 1838 in Boston, Massachusetts. This is also where the ABCFM and the Seminary, with which Dunbar had close relationships, were located. As noted, Dunbar returned to Boston in 1837 to marry his fiancé and assist her in traveling back with him to the mission. While there, Dunbar published a small pamphlet on the Pawnee language.

Because there appeared to be numerous similarities between the History and the John Dunbar Letters and Journal entries, up until 1837, a test comparison of the two texts was conducted. When the comparison was completed, the similarities were striking: namely, 54 percent of the text on page twenty of the History is a direct quote from the Dunbar letter dated on October 22, 1835. Moreover, 85 percent of the text in the same chapter is paraphrased from the Dunbar letters. After reviewing this excerpt, it seemed appropriate to evaluate the

entire booklet, to identify potential consistencies between it and the Dunbar - Allis Letters and Journal entries in an attempt to establish authorship, as well to evaluate whether the Dunbar - Allis Letters are a true (unaltered) version of the original correspondence.

First, the History was meticulously cross referenced to the 1985 version of the Dunbar - Allis letters and journal entries. Second, as quoted passages were located, the words were highlighted in red. Although the History does not cite or reference any other document, certain passages are direct (undenoted) quotes from the Dunbar Letters. Often, large passages were directly, or nearly directly, quoted, with one or more synonyms inserted to replace words (hereinafter, "paraphrased passages"). Paraphrased passages were not indicated in the document by a different color scheme; instead, because it was fairly easy to recognize large areas of red lettering, these areas were counted as paraphrased. Word counts were obtained for each of these sections.

Results: History of the American Mission to the Pawnee Indians

In Chapter 1 of the History, 54 percent (1422/2759) of the text is either a direct quote or paraphrased with one to two words per sentence changed from Dunbar's Letters and Journal entries. The first 679 words in Chapter 1 of the History are, according to the narrator Clifford, attributed to the Prudential Committee's last report. If this section (25 percent of the chapter) is eliminated, 68 percent of Chapter 1 directly quotes or paraphrases Dunbar's writings.

In Chapter 2 of the History, there are 3323 words total. Of these, 1464 (44 percent) are direct quotes from Dunbar's letters and journals; another 83 (two percent) are from Allis' letters. When all paraphrased and quoted portions of the chapter are added together, it totals 2441 words, or approximately 73 percent of the entire chapter. The remainder of Chapter 2 is comprised of a geographic description of the Platte River (10 percent); a description of the characteristics necessary to conduct missionary work in the area (three percent); a description of the chapter contents (one percent), and, collectively (13 percent), the names of the people engaged in the conversation, the questions they are asking, and the words/sentences used to connect the quotations/paraphrasing. If the geographic description, missionary characteristics, and the chapter contents are removed from the total word count for the chapter (475 words), then direct quotes and paraphrasing account for 83 percent of its content.

Chapter 3 is comprised of 58 percent direct quotes from the Dunbar Letters and Journal entries. When paraphrasing is added with the quotations, 77 percent (1819/2353) of the content can be directly

attributed to Dunbar. There is a commentary between the people having the "conversation" that accounts for 17 percent of the chapter (394/2353); and the chapter description is another one percent (19/2353) of the chapter; when these are removed from the total, 94 percent of the text in Chapter 3 is either quoted or paraphrased directly from the Dunbar Letters and Journal entries.

Chapter 4 of the History consists of 2123 words total, with 1070 (50 percent) being direct quotes from Dunbar's Letters and Journals. When paraphrased and quoted portions are combined, 73 percent of Chapter 4 is taken from the works of Dunbar. Approximately two percent of the words counted are the chapter description. Approximately 21 percent of the chapter refers to events surrounding Dr. Satterlee's reception by the Pawnee, of which only a few lines can be located in the Dunbar and Allis Letters. Of further interest is that Dr. Satterlee was killed in the late spring/early summer of 1838, and the letter dated July 27, 1898 (Wedel 1985: 633 0 635), which contained that information, is not referred to at all. This could be due to the lateness of the letter (i.e., it was not received in time to contribute to the publication), or that it was purposefully omitted. Given the prevalence of undenoted quoted/paraphrased passages within the History, it further seems plausible that information in the publication not originating from the letters was gained via an interview with Dunbar while he was in Boston in late 1836 or early 1837.

Chapter 5 of the History uses very little of the content from either the Dunbar or the Allis letters and journal entries. A mere six percent of Chapter 5 is either a quote or a paraphrase. Only 18 percent of the final chapter deals with topics related to the missionaries Dunbar and Allis (e.g., discussing the trip to Massachusetts by Dunbar, his fiancé/ marriage, future of the mission, et cetera). The remainder of the chapter deals with Dr. B in Syria (55 percent); suggestions for missionary experience and character (eight percent); and a hymn (four percent). The remaining 15 percent contains the chapter description, names of the people engaged in the conversation and connecting words (often in question form) to tie together the different sections.

When all of the words of the History are added together (12597), and the totals for quoted and paraphrased content from the Dunbar - Allis Letters are calculated (7354), 58 percent of the booklet can be directly connected to the Letters and Journal entries. When the extraneous information is removed (chapter descriptions, portions that deal with the mission - as discussed above), 81 percent (7354/9083) can be directly referenced to the Letters and Journal entries. The remaining 19 percent of the History are connecting words, sentences and conversational questions. A careful review indicates that the last

traceable letter used by the History (published in 1838) is the May 5, 1836 entry by Dunbar.

Conclusion

Current forensic linguistic methods using R squared analyses appear to be incapable of predicting fit within or among the Dunbar – Allis Letters and Journal entries. While every effort was made to eliminate data that would create outliers, the letters are not only inconsistent when compared to each other, but they also exhibit internal inconsistencies within individual letters. In short, an R squared analysis revealed no temporal trends and little or no consistency among letters, as would be expected if the linguistic method of statistical analyses were robustly applicable to these historical texts.

While R squared values failed to indicate any sort of sameness, it appears fairly likely that the transcriptions of the letters, especially those between 1834 – 1836, were transcribed as written by Dunbar. This is demonstrated by the overwhelming similarity between the Dunbar letters published in 1915-1918 (republished in 1985) and the History of the American Mission to the Pawnee Indians, published in 1838. The data within the Dunbar Letters and Journal entries also appears to be highly consistent when compared with other documents from the same time period (such as Allis' letters, Treaties, and other corroborating historic sources not included in this study).

While it can be safely stated that the Dunbar letters (especially those written between 1834 - 1836) have been transcribed as they were originally written, caution remains warranted when using these texts as a source of data. Though it seems unlikely that Dunbar would intentionally record false statements, and many of his accounts/dates can be compared with other documents (such as the 1833 Treaty, the records of the Indian Agents to the Pawnee, the records of the Dragoon Campaign in 1844, and census records) Dunbar's accounts are still often, despite his affection for the Pawnee, emotionally charged, and his biases are easily recognized. While some of the terminology may be off putting by today's politically correct standards, Dunbar's long years of service to the Pawnee, and his numerous clashes with government officials appear to indicate his passion for not only his work, but his deep fondness for the tribe. While there are, of course, numerous personal and sentimental relationships exhibited within the writings, it would appear that, overall, these accounts can be taken as accurate depictions, albeit somewhat biased by personal affection.

Recommendations for Future Research

While this lack of similarity among the Dunbar – Allis Letters and Journal entries could be indicative of some sort of tampering (for example, by the transcriber) or an attempt at deceit, it seems more likely in this instance that it is due to a more benign factor, such as inconsistencies inherent in the act of letter writing, and the author John Dunbar's apparently active and busy lifestyle. For example, many letters (and journal entries) indicate that they were written over extended periods of time, such as Letter 1, dated July 29, 1834, which is comprised of several dates within the letter (June 5, June 7, June 14); in other letters (such as March 18, 1839 or Letter 5), he states that their child was quite ill and that another son had been born. In the latter case, it seems likely that Dunbar had limited time to formulate his letter, as is potentially indicated by this letter possessing one of the lowest R squared values at 0.0022. It may prove useful to correlate low R squared values with the content of the letters to determine if letters that indicate illness, as well as personal duties may have prevented Dunbar from writing the letters in one session without interruption.

References Cited

Adams, S. H.

1996 Statement Analysis: What Do Suspects' Words Really Reveal?
Polygraph (25) (4):266-278.

Andrew, John A.

1992 *From Revivals to Removal: Jeremiah Evarts, the Cherokee Nation, and the Search for the Soul of America*. University of Georgia Press, Georgia.

Golossanov, Rebecca

2006 (Spring) "Did You Know?" *Christian History & Biography* (90): 3.

Kansas State Historical Society

1915 - 1918 *Collections of the Kansas State Historical Society* (14).
Topeka, Kansas.

Massachusetts Sabbath School Society

1838 *History of the American Mission to the Pawnee Indians*.
Revised and printed by the Committee of Publication
Massachusetts Sabbath School Society.

Maxfield, Charles A.

- 1995 "The Formation and Early History of the American Board of Commissioners For Foreign Missions. *The 'Reflex Influence' of Missions: The Domestic Operations of the American Board of Commissioners for Foreign Missions, 1810-1850.*

Olsson, John

- 2004 *Forensic Linguistics: An Introduction to Language, Crime and the Law.* Continuum, New York, New York.

Wedel, Waldo R. (editor)

- 1985 *The Dunbar-Allis Letters on the Pawnee.* Garland Publishing, Inc. New York & London.