10-2008

The Words You Brought Us: Digital Lexicon of the Recognized American Indian Treaties [Article]

Charles D. Bernholz
University of Nebraska-Lincoln, cbernholz2@unl.edu

Brian Pytlik Zillig
bzillig1@unl.edu

Cokie G. Anderson
cokie.anderson@okstate.edu

Follow this and additional works at: http://digitalcommons.unl.edu/libraryscience

Part of the Library and Information Science Commons

http://digitalcommons.unl.edu/libraryscience/174

This Article is brought to you for free and open access by the Libraries at University of Nebraska-Lincoln at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, UNL Libraries by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
The Words You Brought Us: Digital Lexicon of the Recognized American Indian Treaties

Charles D. Bernholz  
Love Memorial Library, University of Nebraska–Lincoln, Lincoln, NE 68588, USA  
Corresponding author: cbernholz2@unl.edu, fax 402 472-5131

Brian L. Pytlik Zillig  
Center for Digital Research in the Humanities, Love Memorial Library, University of Nebraska–Lincoln, Lincoln, NE 68588, USA; bzillig1@unl.edu, fax 402 472-5131

Cokie G. Anderson  
Electronic Publishing Center, Library Annex, Oklahoma State University, Stillwater, OK 74078, USA; cokie@okstate.edu, fax 405 744-4515

1. Introduction

Colonel Richard Irving Dodge (1827–1895) was a unique military man. Although he had served well during the Civil War, he is remembered best for his journals and writings on American Indians [1]. Not only did he report on United States Army operations against the tribes [2], but he also demonstrated sensitivity to their plight, acquired during a number of excursions in the West [3].

In his *Hunting Grounds of the Great West*, Dodge concluded that “[t]he three principal causes of wars with the Indians are: First. Nonfulfilment of treaties by the United States. Second. Frauds by the Indian agents, and Third. Encroachments by the whites.” He further declared, in the late 1870s, that “it is only necessary to observe that it would be extremely difficult to find any treaty entered into by the Government with the Indians during the last twenty years, which has been strictly and honourably fulfilled” [4]. These shortcomings were echoed in the words of the Brulé Sioux chief Spotted Tail (1823–1881; Fig. 1) at the 1876 proceedings that yielded additional Indian cessions beyond those already conveyed: “We have come here to meet you, my friends. We have considered the words you brought us from the Great Father…. These promises have not been kept” [5].

Charles J. Kappler, in his multi-volume *Indian Affairs: Laws and Treaties* [6], collated the words that Spotted Tail bemoaned, as well as those from similar negotiations conducted with other tribes. These legal materials have been digitized by the Oklahoma State University Library Electronic Publishing Center (OSU) [7], and the suite’s format and applicability have been described [8]. The few documents absent from Kappler’s second volume of recognized treaty texts include instruments created by the tribes and the British government during its presence before Independence. In total, nine contracts were gathered in a complementary digital collection [9], so that all recognized treaties with the America Indians may now be accessed on the World Wide Web [10].

2. Forming the digital lexicon

Text analysis has become a major area of investigation for many academic disciplines [11] and has instigated the development of useful tools [12] to propel these studies. TokenX is a web application developed by the second author with the support of the University of Nebraska–Lincoln’s (UNL) Center for Digital Research in the Humanities (CDRH). It was designed for analysis and visualization of Extensible Markup Language (XML) and Text Encoding Initiative-compatible (TEI) documents [13]. The application’s inherent flexibility means that virtually any digital text may be tokenized.
Briefly, the treaty files from Kappler’s collation were obtained from OSU in Hypertext Markup Language (HTML) format and converted for text analysis. The remaining nine instruments were taken from UNL’s Early Recognized Treaties with American Indian Nations web site and processed in a similar, four step manner. First, files were transferred into Extensible HyperText Markup Language (XHTML) by using HTML Tidy, an open source program and library for checking and generating HTML/XHTML. Next, the XHTML files were edited to eliminate any navigational or unnecessary links. A series of files was then created and verified against standards for XML. UNIX shell scripts and Extensible Stylesheet Language for Transformations (XSLT) conversions were subsequently employed to convert the XML files into TEI [14]. These files were divided into the two categories of American Indian treaties – the British and the American suites. Finally, the application processed each corpus and created an XML file that contained a word frequency data table. The two data tables were each converted into a Structured Query Language relational database and these were queried using the scripting language PHP: Hypertext Preprocessor.

3. Research applications and conclusions

A full explanation of this study, entitled You say cranberry and I say cramberry, let’s call the whole thing off: A lexicon of the recognized American Indian treaties, is available at [http://cdrh.unl.edu/treatylexicon/][15]; the lexicon files for each of the British and American treaty arrays are linked to this site. The goal of the description is to expose the power of TokenX, as well as the diversity of the terms found in these 375 treaties. As part of this exercise, examples of errors in Kappler’s collations, in the Statutes at Large, and in the OSU digital conversions were found and reported. These faults reveal various interesting aspects of the perspectives and/or the misunderstandings of the treaty participants, or of the collators of these documents. John Burrows has perceived that “the real value of studying the common words rests on the fact that they constitute the underlying fabric of a text, a barely visible web that gives shape to whatever is being said” [16]. This initial surface, now uncovered for American Indian treaties, demands further disclosure.

At a bare minimum, the observed frequencies recorded in the lexicon files offer a prospect upon legal thought of the times, and upon the vocabulary used to express the goals, purposes, and deeds of the negotiators. The capability of TokenX to disassemble these materials—or any other digital documents of interest—makes possible for an unlimited number of disciplines further text analysis exploration. Investigating recognized American Indian treaties is just one such exercise, expedited by a well-defined document corpus, that may be thought of as prototypic of potential studies in the future. As the repository for such resources within academic settings, current and future libraries will need to become more involved in text analysis undertakings through the creation of stronger bonds with groups and departments that fashion these digital portals [17].
References

1. Kime, W. R., Colonel Richard Irving Dodge: The Life and Times of a Career Army Officer (Norman, OK: University of Oklahoma Press, 2006). This six hundred page biography presents every phase of Dodge’s life, and concludes with the statement that his multifaceted service in the West was “the chief legacy of this humane warrior, this talented soldier-author, this engaging personality” (p. 513).


5. Dodge, The Hunting Grounds of the Great West, xlvi, emphasis original. The council concluded with an agreement; treaty negotiations with the tribes had ceased in 1871 (16 Stat. 544, 566). An act to ratify an agreement with certain bands of the Sioux Nation of Indians and also with the Northern Arapaho and Cheyenne Indians, at 19 Stat. 254, provides the text of this document. Spotted Tail’s speech was also described in a New York Times column entitled The Indian war: A new treaty with the Sioux (27 September 1876). The Treaty with the Sioux - Brulé, Oglala, Miniconjou, Yanktonai, Hunkpapa, Blackfoot, Cuthead, Two Kettle, Sans Arc, and Santee - and Arapaho, 1868 was the transaction modified by this agreement. See the OSU site for this document: http://digital.library.okstate.edu/kappler/Vol2/treaties/sio0998.htm. We thank Paul Burke, at http://firstpeople.us/, for permission to use that site’s image of Spotted Tail.


10. Oklahoma State University has added access to these additional items from their main Kappler page through the link entitled “Early Indian Treaties from the University of Nebraska–Lincoln Electronic Text Center.”

11. John Unsworth’s presentation “Forms of Attention: Digital Humanities Beyond Representation” (http://www3.isrl.uiuc.edu/~unsworth/FOA/), before the 2004 Canadian Symposium on Text Analysis, provides a strong historical summary of this digital realm. His bold statement—“Statistics are all about measuring the measurable, whereas the humanities are all about effing the ineffable”—expresses a major goal for text analysis investigators.


13. See http://tokenx.unl.edu/ for more on TokenX.

14. HTML, XHTML, and XSLT are World Wide Web Consortium (http://www.w3.org/) standards for web pages. TEI is a standard for the representation of texts in digital form. HTML Tidy may be acquired through the library at http://tidy.sourceforge.net/.

15. As an indication of the magnitude of this undertaking, the lexicon for the seven British documents contains 78,348 words, with 5,566 unique terms, while the list for the 368 American instruments holds 23,244 distinct items from more than 640,000 words.

16. See this remark on page 323 of his Textual Analysis, in A Companion to Digital Humanities (pp. 323–347).

17. Howard Besser, The Past, Present, and Future of Digital Libraries, in A Companion to Digital Humanities, pp. 557-575, offers an answer to the question “What is a Library?” by declaring that “[t]he digital collections we build will not truly be digital libraries until they incorporate a significant number of [the] services and ethical traditions” (p. 560) that all libraries have sustained.

The OCLC accession numbers for the Kappler, early treaties, and lexicon collections are 44288939, 70172136, and 231397265, respectively.