1914 Revisited:

An Exploration in Automated Content Analysis of the Effect of Communications from European Capitals on British Decision Makers

One of the most prominent early efforts in the quantitative study of international politics was the 1914 Project (sometimes referred to as the Stanford Studies in Conflict and Integration [North 1976]). The initial aim of this research was to try to understand the outbreak of World War I through an analysis of the principal decision makers in the five protagonist nations through a data generation process based on content analysis, which can be briefly described as the analysis of communication to identify who (author identification) said what (sentiment analysis) to whom (audience analysis) with what effect (impact analysis).

The output of this project was substantial and generated fundamental insights into the manner in which decision makers dealt with an unexpected situation. Zinnes (1968), drawing upon the perceptual data associated with the leaders of major powers who initially became involved in the war, investigated the extent which a state’s hostile behavior is a function of its leader’s perceptions of hostility and the corollary that the more hostility the leaders perceive the more hostility the state will express. At the same time, in research closely related to Zinnes, North and his colleagues (North, Holsti and Brody 1968) investigated the utility of a stimulus response model in which a state’s acts were mediated through the perceptions of hostility by the leaders.

All of this and other research relied on laborious hand coding of documents, the process of which was described by North (1976) as a “task of Herculean proportions” where the “requirements of keen insight, precision, and seemingly endless drudgery were an invitation to nervous breakdowns” (p. 74). Because of these impediments Holsti embarked on the development of a machine based content analysis program, derived in part from Stone’s General Inquirer (1966). At the heart of the program was the Stanford Political Dictionary (here after SPD) measuring more than 3,300 politically relevant words on the degree to which each was positive or negative, strong or weak, or active or passive, using Osgood, Suci and Tannenbaum’s (1957) method of the semantic differential as a way of capturing linguistic expression.[[1]](#footnote-2) The shift to computer analysis, however, did not obviate the need for human coding, as the analysis was sensitive to the syntax involved in analyzing the relationship between actors, their attitudes and the target of their attention. Holsti describes the coding process as “easily learned” (1969, 380), but is nevertheless “time consuming” and “the highest cost factor in any research using this program.”

Empirical research based upon data generated through this program did emerge. It included papers on the Cuban Missile Crisis (Holsti, Brody and North 1964), alliance cohesion in the Warsaw pact and NATO in international crises (Hopmann 1967), Sino-Soviet relations (Holsti 1966) and conflict in the Middle East during the Suez Crisis (Siverson 1972). However, by the mid-1970s research using the program more or less ceased, probably because the amount of effort required for the coding was excessive relative to the amount of data that could be collected, a particularly acute issue if the research question involved data from more than one case.

What follows is a report on an attempt to assess the utility of the SPD in a system that involves no coding, and only minimal document preparation before processing. It is, in essence, a test to see if the SPD can be used to generate data on perceptions in a fast and reliable way that maps to previous results. In what follows we first describe the method used. We then apply it to a comprehensive body of documents authored by senior British foreign policy officials during the 1914 Crisis; these same documents were part of those used in the content analysis of the 1914 Project. After presenting the results of our procedure, we compare them to the results yielded by the earlier analysis. We close with a tentatively optimistic assessment of the utility of our procedure.

**A New Approach to Content Analysis of Political Documents**

The approach we have taken might be broadly described as key words in context. The process is extremely simple. We processed each document in several stages. First, the raw text was loaded into the R programming environment as a “corpus”, or body of textual data. Each document was then cleaned and prepared for analysis, as described immediately below. One key challenge in natural language processing is that there are, simply put, lots of words; the English language contains over one million unique words (Michel *et al.* 2010). This count includes thousands of homonyms, synonyms, combinations of tenses and cases, and so on. This makes automated analysis extremely difficult, as any parser has to assign some semantic meaning to each individual word in the corpus being analyzed. The greater the number of words, the more difficult this task will be. To deal with this apparent complexity, we used a set of procedures and tools standard in the literature to reduce the size of the vocabulary in these cables, telegrams and letters. Our goal was to strip away as much uninformative linguistic ephemera as possible, while maintaining the core information contained in each cable.

To accomplish this punctuation, whitespace, and numbers were removed, and all text was unified to lower case. Next, all words in the document were stemmed, a process that removes suffixes and tenses to reduce words to their common roots. For example, “accuser”, “accused”, “accusation”, “accusatory” and so on are all forms of the root word “accuse”. The stemming process reduces all variants of a word to their root, greatly reducing the size of a given vocabulary and simplifying the task of language analysis (Krovetz 1993). This process of stemming is essential for natural language analysis, as it unifies word variants so that the core meaning can be included in sentiment analysis models. It is worth noting that stemming algorithms make the key assumption that all variants of the root word have exactly the same meaning. As such, a core challenge of stemming is ensuring that words with identical meanings are reduced to their common roots, while words that *appear* similar but have different meanings are preserved (for example, “executor” and “executioner”). This is done by referencing human-created dictionaries that identify words that are morphologically but not meaningfully related to one another (Porter 2001). To carry out this stemming process, we implemented Porter’s (2001) Snowball algorithm, a rules-based “stemmer” that iteratively filters English-language words based on suffix to accurately identify root words.

Next, we further reduced the size of the vocabulary corpus by removing stopwords. These are words in the English language that appear commonly in written or verbal communication, but themselves contain little information. For example, pronouns like “I”, “she”, “him”, or prepositions like “on”, “for”, “at” are all useful for communication, but are not independently informative enough to be of use in sentiment analysis. As such, they were dropped from each document as well. Processing the documents significantly reduced their size and complexity. Finally, we identified all stemmed words in the processed document that matched entries in the Stanford Political Dictionary, and assigned each word its corresponding positive/negative, active/passive, and strong/weak scores. This allowed us to not only gauge the overall sentiment of each cable or days’ worth of cables, but also gauge the sentiment associated with sentences, paragraphs, or other verbal structures within a given cable.

Following the document preparation, we identified the relevant entities who were likely to be important elements in the documents. We are primarily interested in British communications regarding the military and political situation on the continent, so we searched for all permutations of the words “Germany”, “Austria”, “Hungary”, and “Dual Alliance”. Our goal is to identify the sentiment, as measured by the SPD, of language used to discuss these key terms. Our initial approach is a relatively simple one: after locating each instance of our key words in a given cable document, we analyzed the sentiment scores of all words within distance *K* on each side of each key word. So for example, if the text “We are worried that Austria-Hungaryis preparing to mobilize ...” occurs, our program will identify the *K* words preceding and following the key term “Austria-Hungary”. For the value *K* = 6, our program will identify and look up the root words *worry*, *prepare*, and *mobilize* in the SPD, and record the scores of each word. The program will then move on to the next appearance of a key term and so on until the end of the document, at which point the values would be summed and recorded. The result is presented as the mean positive/negative, strong/weak, and active/passive scores of all such words within a distance of *K* of any key terms in the document. We understand this statistic to represent the tone in which the key term is being discussed or regarded within the document being analyzed. This approach is highly flexible, so that any number of relevant actors can be specified and the number of words on each side of the actor to be scanned can be set to any value desired.

The documents we analyze are drawn from the electronic version of the final volume of British Documents on the Origin of the War, 1898-1914 (Gooch and Temperley 1926, accessed at <http://net.lib.byu.edu/estu/wwi/1914m/gooch/firstpps.htm>). These are documents circulated within the British Foreign Office as well as those sent to British officials in continental Europe. The primary authors of these were the foreign minister, Sir Edward Grey, and the permanent under-secretary, Sir Arthur Nicholson. As noted earlier, these documents were the source of the British data reported in the 1914 Project. Table 1 reports the number of documents by day over the period from the assassination of the Archduke to August 4, 1914, the day the UK declared war. In early July there are some days with few or no documents, but as the days passed the volume of communications increased sharply.

*Table 1: Frequency of cables sent by London, 6/19/14 – 8/3/14*

|  |  |  |  |
| --- | --- | --- | --- |
| **6/19/1914** | **0** | **7/13/1914** | **0** |
| **6/20/1914** | 0 | **7/14/1914** | 0 |
| **6/21/1914** | 0 | **7/15/1914** | 0 |
| **6/22/1914** | 0 | **7/16/1914** | 0 |
| **6/23/1914** | 0 | **7/17/1914** | 0 |
| **6/24/1914** | 1 | **7/18/1914** | 0 |
| **6/25/1914** | 1 | **7/19/1914** | 0 |
| **6/26/1914** | 0 | **7/20/1914** | 2 |
| **6/27/1914** | 0 | **7/21/1914** | 1 |
| **6/28/1914** | 0 | **7/22/1914** | 1 |
| **6/29/1914** | 2 | **7/23/1914** | 1 |
| **6/30/1914** | 1 | **7/24/1914** | 4 |
| **7/1/1914** | 0 | **7/25/1914** | 8 |
| **7/2/1914** | 1 | **7/26/1914** | 3 |
| **7/3/1914** | 0 | **7/27/1914** | 4 |
| **7/4/1914** | 0 | **7/28/1914** | 10 |
| **7/5/1914** | 0 | **7/29/1914** | 11 |
| **7/6/1914** | 2 | **7/30/1914** | 6 |
| **7/7/1914** | 0 | **7/31/1914** | 11 |
| **7/8/1914** | 2 | **8/1/1914** | 15 |
| **7/9/1914** | 1 | **8/2/1914** | 9 |
| **7/10/1914** | 0 | **8/3/1914** | 8 |
| **7/11/1914** | 0 | **8/4/1914** | 25 |
| **7/12/1914** | 0 |  |  |

Because the program is so flexible, we chose to process the documents by gathering the scores on words within a three-, six-, and nine-word distance on both sides of each key noun (*Germany, Austria, Hungary, Dual Alliance*). The next step we took was to see if there were palpable differences in the scores across the documents in these three modes of processing. We examined the correlations among each of the dimensions and, perhaps not surprisingly, decided that the six word span was best because, pragmatically, it offered the best overall correlations across all three of the dimensions and we are interested in a tool that is useful in retrieving and summarizing as much usable data as possible.

The summary of the documents in Table 1 shows several gaps in the London documents, where data are present for only twenty-four days. To put it differently, for sixteen of the thirty-seven days between the assassination of the archduke and duchess on June 28th and the outbreak of the war on August 4th, the decision makers in London have little or nothing to say about the situation on the continent. Given the effort expended on the collection of the British documents it is unlikely many or any were missed. There are at least two reasons for this gap. First, from his intimate letters to Venetia Stanley, we know that Prime Minister Asquith was completely absorbed in the “Irish Problem”, particularly as it related to Ulster (Brock and Brock, 1985). From these letters we also know it was not until July 24, four days before the Austro-Hungarian declaration of war against Serbia, that Asquith showed any concern the conflict could become dangerous, but discussions of Ireland and other political matters were a prominent continuing theme in the letters. Second, from the record of parliamentary deliberation in Hansard (1914) there appears to have been a sense of detachment from the events unfolding in Europe. Discussion in the House of Commons was scant, and when the conflict was mentioned it was in the context of ministers offering brief comments that the matter would be resolved peacefully and that British relations with Germany had improved. Perhaps this optimism was based on the successful resolution of several relatively recent crises, such as Agadir and Morocco, as well as the containment of two Balkan wars between 1912 and 1913 to the minor powers. To British decision makers at the time, the crisis between Austria-Hungary and Serbia may have seemed like another regional flare-up that would be contained before it affected the other powers on the continent.

**Results**

The 1914 Project focused much of its coding on perceptions of hostility. The SPD, of course, does not have an explicit category for such perceptions, but it does record positive and negative sentiments, and it is not unreasonable to interpret the extent to which perceptions that are negative outweigh those that are positive as hostility. Similarly, it is possible to use the sum of the valences that are on balance strong, active and negative as an indicator of threat.



*Figure 1: Lowess correlations by day of transformed positive negative UK perceptions of the Dual Alliance and its members.*



*Figure 2: Lowess correlations by day of transformed UK perceptions of “threat” by the Dual Allaince and its members.*

Figures 1 and 2 present displays of the lowess correlations by day of the crisis for (1) our measure of positive-negative sentiment, and (2) the sum of the measures of values, respectively. Figure 1 shows some heterogeneity in the days immediately after the assassinations, but after July 20 they turn sharply negative, peaking, nor surprisingly, just before the UK’s declaration of war. The early data on threat perception displayed in Figure 2 are less heterogeneous than those in Figure 1, but they do display the same sharp increase after July 20. These data also have smoother appearance. In terms of what we know historically about the crisis, these data have strong face validity. However, how do they compare with the human coding reported in the 1914 Project?

The early paper by North et al. (1968) aggregates perceptual data by alliance, so measuring perception of hostility in the UK from these is not possible. However, in a later study Holsti (1972, 243) breaks out perceptions of hostility by UK decision makers. We can compare those data with the data in Figures 1 and 2. The lowess plot of British perceptions of hostility by day of the crisis is shown in Figure 3.



*Figure 3: Lowess correlations by day of UK perceptions of hostility by the Dual Alliance and its members.*

The data are sparse in the first days of the conflict, but show the same dramatic increase after July 24th. To be sure, these data are more tightly clustered than the others, but it is hard to avoid the conclusion that the use of the SPD captures the rise in the negative and threatening perception of the Dual Alliance as well as does the human coding of the 1914 Project. In addition, the data were generated in a process requiring little in the way of human intervention or judgment; there is no test for inter-coder reliability, because there are no coders. There is also no post hoc information bias that might enter into the process of coding text.

These is at least one further topic that needs consideration. If our employment of the SPD can capture perceptions of hostility and threat, can it also distinguish between enemies and allies? That is, perhaps the procedure we use is simply tapping in to a generalized sense of threat and hostility by the British. If so, then perceptions of France, a British ally, might display the same patterns. To test for this issue, we re-ran the analysis program focusing on France as the key term in the British cables.



*Figure 4: Lowess correlations by day of UK positive-negative perceptions of France.*

The results for the positive-negative dimension are shown in Figure 4. Here the positive-negative values are not multiplied by -1, so the movement upward shows an increasingly positive view of France. In short, the SPD does distinguish between friend and foe: it appears that as Great Britain’s estimation of threat and hostility from its rivals Germany and Austria-Hungary increased, this was accompanied by an increase in positive sentiment towards its more reliable ally France.

**Discussion**

The material above is not a proof of concept but rather a test of concept. The results of the test lead us to have a measured sense of optimism about the feasibility of using the SPD in other contexts for the rapid, minimal human intervention processing of political texts for sentiment measurement. There are several contexts in which the use of the SPD in this way might be highly useful. First, it could serve as a useful adjunct to the collection of event data (Schrodt 2012). Event data is in itself a form of content analysis in which events gathered from a news feed are placed into categories for analysis; these categorizations reveal who is doing what to whom, but they are likely to miss the sentiments that attend the acts. It is plausible to suggest that the SPD might be a useful adjunct to this type of event data. In addition, the SPD in itself might be useful in forecasting how states will behave toward other states. For example, Hunt’s (1997) consideration of how the content of newspapers of one state —particularly those subject to government control—can be used to anticipate aggressive behavior toward other states. The data analysis used in this study required hand coding of news stories, something that can be done through the SPD quickly and easily.

While the above makes a palpable case for the efficacy of the SPD in deriving measures of hostility and threat, it would be presumptuous to think that it could not be improved. Perhaps the most notable improvement would be an effective way of managing syntax so that a more fine-grained understanding of the views of actors toward others could be obtained. For example, a more in-depth approach of syntax mapping would increase our confidence that sentiment-loaded words are actually being used to describe or address the key actors in our analysis. While we believe that proximal words often relate directly to one another, simply noting that “attack” appears near “Austria-Hungary” does not prove that Austria-Hungary is actually the object (or indeed, related at all) to any sort of attack. Semantic processing tools such as Chen & Manning’s (2014) neural network-based sentence parser allow for automated mapping of natural language structures such as sentences and paragraphs, allowing for highly accurate identification of (for example) subject/object relationships in unstructured text.

Given the strength of our results using this this comparatively crude approach, we believe that using these more sophisticated machine-learning tools would strengthen our results and add further nuance, rather than fundamentally change them. Overall, we believe this approach of pairing automated text analysis tools with hand-coded sentiment categorization through the SPD shows considerable promise in understanding how decision-makers assess, discuss, and respond to changes in their political environment, and as such, is an avenue worth pursuing further.

Works Cited

Brock, Michael and Eleanor Brock. 1985. Editors. *H. H. Asquith Letters to Venetia Stanley*. Oxford: Oxford University Press

Chen, D. and Manning, C.D., 2014, October. “A Fast and Accurate Dependency Parser using Neural Networks.” *EMNLP:* 740-750.

Gooch, G. P. and Harold Termperley. 1926. *British Documents on the Origins of the War, Volume 11*. London: His Majesty’s Stationary Office.

Hansard Digital Archive of Parliamentary Debates. *House of commons Debate (5th series) 19 July 1914 volume 8*. At <http://hansard-archive.parliament.uk/>. Accessed August 17, 2016.

Holsti, Ole R. 1972. *Crisis Escalation War*. Montreal: McGill-Queen’s University Press

Holsti, Ole R. 1966. “External Conflict and Internal Consensus: The Sino-Soviet Case” in *The General Inquirer*, edited by Philip J. Stone , Dexter Dunphy, Marshall Smnith and Daniel M. Ogilive, 343-58. Cambridge: MIT Press.

Holsti, Ole R. 1969. “A Computer Content Analysis Program for Analyzing Attitudes: The Measurement of Qualities and Performance”, in *The Analysis of Communication Content,* edited by Groege Gerbner, Ole R. Holsti, Klaus Kripendorff, William J. Paisley and Philip J. Stone, 355-380. New York: Wiley.

Hopmann, P. Terrance. 1967. “International Conflict and Cohesion in the Communist System.” *International Studies Quarterly* 11 (September): 212-37.

Hunt, Benjamin. 1999. *Getting to War*. Ann Arbor: University of Michigan Press.

Krovetz, Robert. 1993. "Viewing morphology as an inference process." *Proceedings of the 16th annual international ACM SIGIR conference on research and development in information retrieval*, 191-202.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, Joseph P. Pickett, Dale Hoiberg et al. 2011. "Quantitative analysis of culture using millions of digitized books." S*cience* 331 (6014), 176-182.

North, Robert C. 1976. “The Stanford Studies in International Conflict and Integration.” In *In Search of Global Patterns*, edited by James N. Rosenau, 73-80. New York: Free Press.

North, Robet C., Richard A. Brody and Ole R. Holsti. 1968. “Perception and Action in the 1914 Crisis” in *Quantitative International Politics,* edited by J. David Singer, 103-122. New York: Free Press.

Osgood, Charles E, Geoege J. Suci and Percy H. Tannenbaum. 1955. *The Measurement of Meaning*. Urbana, Ill. University of Illinois Press.

Porter, Martin F. "Snowball: A language for stemming algorithms." 2001. At <http://snowball.tartarus.org/texts/introduction.html>. Accessed August 17, 2016.

Randolph M. Siverson. 1972. "The Evaluation of Self, Allies, and Enemies in the 1956 Suez Crisis." *Journal of Conflict Resolution*, 16 (June), 203-210.

Schrodt, Philip A. 2012. "Precedents, progress, and prospects in political event data." *International Interactions* 38 (4): 546-569.

Stone, Philip J., Dexter Dunphy, Marshall Smnith and Daniel M. Ogilive. 1966. Editors. *The General Inquirer*. Cambridge: MIT Press.

1. An extensive discussion of the development of the dictionary, the coding and the resulting output is given in Holsti (1966). [↑](#footnote-ref-2)