



Bulletin de méthodologie sociologique

Bulletin of sociological methodology

97 | 2008

January

Co-Word Text Analysis Applied To Political Science

2006 International Political & "Parapolitical" Headlines

Karl M. van Meter and Mathilde de Saint Léger



Electronic version

URL: <http://journals.openedition.org/bms/308>

ISSN: 2070-2779

Publisher

Association internationale de méthodologie sociologique

Printed version

Date of publication: 1 January 2008

Number of pages: 18-38

ISSN: 0759-1063

Electronic reference

Karl M. van Meter and Mathilde de Saint Léger, « Co-Word Text Analysis Applied To Political Science », *Bulletin de méthodologie sociologique* [Online], 97 | 2008, Online since 01 January 2011, connection on 02 May 2019. URL : <http://journals.openedition.org/bms/308>

This text was automatically generated on 2 May 2019.

© BMS

Co-Word Text Analysis Applied To Political Science

2006 International Political & "Parapolitical" Headlines

Karl M. van Meter and Mathilde de Saint Léger

Introduction

- 1 In terms of methodology, there has been a great deal of exchange over the past few years between sociology and political science, often resulting in joint sessions in international conferences and joint publications in which the *BMS* has played a certain role (Marchand, 2005; Marchand, 2007a). Indeed, Marchand's work in political science recently resulted in the publication of a book that cites his contribution to sociological methodology and to the *BMS*, in particular (Marchand, 2007b).
- 2 Michel Pinault's work on the politics and history of French scientific research (Pinault, 2006) has also cited sociological methodological developments and the *BMS* where the co-word analysis system, Calliope, has often been presented (van Meter, Cibois and de Saint Léger, 2004; de Saint Léger and van Meter, 2005). In this particular area of politics, Pinault utilized Calliope to analyze the texts and discourses of the major social actors, and social network analysis to analyze the roles and relationships between those actors in the development of a coherent national policy for French scientific research.
- 3 In a more general area of political science, many different centers of research and information on political developments, particularly on the international level, produce — particularly at this time of year — annual reports, "bilans", yearbooks or "tops of ...". Examples include the International Institute for Strategic Studies' *Strategic Survey*, the Institut des Relations Internationales et Stratégiques (IRIS) annual *L'Année stratégique*, the Swedish National Defence College's *Strategic Yearbook 2006 - European Security and Defense Policy: A European Challenge*, and the *Statesman's Yearbook: The Politics, Cultures and Economics of the World*, by Barry Turner. In most cases, these works are compilations of a selected number of reports made during a given year, but they are often accompanied by analyses

or opinions attributed to experts in the domain concerned, be it international politics, security, economics, armaments, etc. Some of the better known examples include the United Nations' *Yearbook*, the Stockholm International Peace Research Institute (SIPRI) *SIPRI Yearbook*, the CIA's annual *World Factbook*, the Council of Europe's *European Yearbook*, and the *World Political Almanac*, by Chris Cook and Whitney Walker.

- 4 Methodologically speaking, these works usually suffer from a double weakness: the sampling basis for the information or data used in the analysis; the reliance upon an "expert" or "experts" to analyze and interpret the data in a non-formalized manner. The catastrophic consequences of this double weakness in international politics can be seen in the Bush White House decision to invade Iraq.
- 5 While it is extremely difficult — if not impossible — to develop either an exhaustive sample or a representative sample of international political developments for a given year, there is a tremendous amount of information publicly available in this area as the above-mentioned reports show. Indeed, the problem has more to do with handling and managing the vast quantity of information, and selecting the pertinent and representative information for further analysis. Informal experiments have been run in this domain to show how a report by a major wire service — Associated Press (AP), Reuters, Agence France Presse (AFP), etc. — is "rewritten" into multiple, even hundreds, of media reports which furnish little or no further information than that contained in the original wire service report. The difference often resides in the political, cultural and ideological choice of words to communicate and to interpret the same initial information. This choice of words, this language, does furnish further information, but at another level: that of the analysis of the "rewriting" and those that carry it out.
- 6 Although such analyses are fascinating and of great interest, they are not the subject of this study which is oriented much more toward the initial information provided by wire services and press reports concerning international politics and its less-publicly-known sibling, "parapolitics". If political science is the study of politics and the use of political power as it presents itself on the public scene, parapolitics is the study of the use of political power that is not presented on the public scene. The term seems to have been coined by Prof. Peter Dale Scott of the University of California, Berkeley, in the late 1970s or early 1980s. According to Scott, a former Canadian diplomat, poet and author, parapolitics is the study of the "real" use of political power, as opposed to the publicly avowed use of political power.
- 7 One of Scott's students, Jonathan Marshall, founded the journal, *Parapolitics USA*, in the 1980s. Other colleagues of Scott founded the French "Association pour le Droit à l'Information" (ADI) and the journal, *Parapolitics*, during the same period. The ADI continues to publish books (Calvi and Schmidt, 1988; Schmidt, 2005) and reviews, although the journal *Parapolitics* has evolved and changed its title to become — since 1995 — the current Internet journal, *Intelligence*¹. ADI and *Intelligence* information has served previously in the social network analysis (or "traffic analysis") of the long-term cat-and-mouse/cops-and-robbers dynamic between law enforcement and suspected offenders (van Meter, 2002).

The 2006 *Intelligence* "Timeline" Data Set

- 8 One of *Intelligence*'s objectives is to document international politics and parapolitics through the systematic sampling of the media². For several years, the journal has published in each issue a "Timeline" article consisting of daily entries of media report headlines or titles, along with an indication of the source. For example, the extract below is taken from the article, "Timeline From 7 November To 21 November" in *Intelligence* (n. 508, 26 November 2007, p. 2):

1107 [7 November 2007]

- * Rue89, Sarkozy Says "I Love You" to the American Congress
- * Long Beach Press-Telegram, Antiwar Vets Slam Parade Ban
- * Business Journal of Phoenix, Fort Huachuca Intelligence Center Draws Private Contractors
- * Figaro, UN Struggles to Understand Raid on the Euphrates - Even though it was the victim of aerial bombardment, Syria refuses any visit from the IAEA's nuclear experts

1109 [9 November 2007]

- * UN, Pakistan puts Bhutto under house arrest
- * Truthout, Pelosi Turns to Confront Bush on War Spending
- * LAT, Protest Greets Police Plan to Map Muslim Angelenos [Los Angeles]
- * Alternet, Iraqi Government to UN - "Don't Extend Mandate for Bush's Occupation"
- * AP, Stanford Students Protest Rumsfeld Appointment
- * Olympian, Court-Martial of Watada Might Not Come
- * LAT, LAPD to Build Data on Muslim Areas - Anti-terrorism unit wants to identify sites "at risk" for extremism

- 9 Note that there were no noteworthy reports registered for Thursday, 6 November 2007. The reports come from a wide variety of sources: AP, *Los Angeles Times* (LAT), United Nations (UN), the conservative *Le Figaro* French daily, etc. These entries can be annotated with square brackets when necessary to clarify the information. For example, [*Los Angeles*] was added in the entry about, *Protest Greets Police Plan to Map Muslim Angelenos*.
- 10 It should be noted that not only parapolitical information such as *LAPD to Build Data on Muslim Areas* is included, but also more classic political information such as *Pakistan puts Bhutto under house arrest*. Indeed, without the framework or "scenery" provided by publicly avowed political action, the parapolitical complement of information would have little meaning and would be very difficult to analyze or to interpret. Moreover, including both permits a joint analysis and the possibility of examining interrelations between the two.
- 11 *Intelligence* is published every two to three weeks, except during summer and Christmas vacations, thus accounting for 18 to 20 issues per year³. The "Timeline" article of each issue includes daily entries of media reports since the preceding issue and up to the date of publication. In the issue cited above and dated 26 November 2007, the "Timeline" includes entries from 7 November to 21 November. Each "Timeline" article also includes a synopsis, usually consisting for a few paragraphs. For example, the synopsis for the 26 November 2007 issue including the following:

On the international scene, beside the Iraq war, the violent and unstable situation in Pakistan has become a major topic of news reports. Two new developments we mentioned in our previous issue — armed clashes between Kurds and Turks, and Michael Mukasey's nomination as [US] attorney general — have disappeared from the headlines. However, another topic associated with Mukasey — torture and

waterboarding — has retained its place as a major media topic, along with corruption and scandals associated with the Bush White House and the Republican Party. Blackwater mercenaries in Iraq have also disappeared from the headlines as investigations move toward trials in the US.

...

The change in media reports and the relative importance of specific topics are reflected in the composition of the news articles retained by *Intelligence* over the past fortnight: Bush/GOP corruption (still "stable" at 13K, 17X; previously 10K, 15X), Kidnapping-Disappearances-Torture (10K, 12X; significantly down from 18K, 33X), US "domestic" politics and problems (9K, 20X; just slightly up from 8K, 15X), Asia - mostly Pakistan (stable at 9K, 20X; previously 11K, 12X), Pentagon (8K, 13X; a major increase from "no noteworthy news"), Bush's "snoopgate" (7K, 8X; slightly down from 9K, 18X), Iraq war in Washington (stable at 5K, 10X; previously 6K, 12X), Iran (stable at 5K, 10X; previously 6K, 10X), Eastern Europe (5K, 10X; just slightly down from 6K, 12X), War in Iraq (4K, 7X; slightly up from 3K, 5X), Blackwater and other mercenaries (3K, 6X; significantly down from 10K, 12X), FBI (3K, 5X; significantly up from 1K, 1X), and CIA (stable at 2K, 3X; previously 3X, 3X). Western Europe, Middle East, Americas and Canada are all 1K or less.

- 12 These synopses provide not only a summary view of international political and parapolitical developments covered by a given issue, but also provide a bridge between successive issues and the means of following the developments over a period of time as their relative importance (in number of entries) varies from issue to issue, sometimes disappearing and reappearing later.
- 13 In the analysis which follows, we have taken all the *Intelligence* "Timeline" articles covering events in 2006⁴. That means a total of 20 issues. We then removed the synopsis part of these articles and retained only the daily entries of media headlines and titles, along with any accompanying information (in square brackets) necessary to explain or situate the report.
- 14 Given the volume of data necessary to obtain statistically pertinent results, we divided the initial data set into three chronologically successive sections of four months each, thus providing a winter-spring 2006 section (with seven "Timeline" articles), a summer 2006 section (with six articles) and a fall-winter 2006 section (with seven articles). These three sections, more-or-less similar in volume, were systematically used in the diagrams belows. We found that trying to analyze the 2006 data set with more than three sections did not provide statistically stable or pertinent results, although more sections could have been constructed by using the fuller summary or résumé of each report for which we only used the title or headline. Even though this more extensive information was included in each issue of *Intelligence*, the analysis of such a massive data set would have completely changed the exploratory nature of this research. It is, however, a serious option for future work in the application of text analysis to politics and parapolitics.

Co-Word Analysis & Calliope

- 15 The Calliope system of analysis applied to the 2006 *Intelligence* data set is a form of "co-word analysis" or the co-occurrence of terms in an entry used as an index of similarity to form "clusters" or "components". The more often two terms co-occur in headlines or titles together, the "closer" or more "similar" they are, and thus the more likely they are to be members of the same cluster. Calliope⁵ was developed as the Windows version of the former Lexinet-Leximappe system under DOS. The latter was developed by the

Department of Research of the French INIST (Institut d'Information Scientifique et Technique) in the framework of sampling information flow (Chartron *et al.*, 1990). The objective was to develop a computer program which identified new terms and new associations between existing terms that emerge in scientific abstracts. Since the program uses a procedure that is independent of the specific scientific domain under consideration, it can be used for any set of words, even the official biographies of members of the Central Committee of the Communist Party of the former Soviet Union (van Meter *et al.*, 1991) or the titles and headlines of 2006 media reports. In other words, the algorithms incorporated into the program do not refer to pre-established dictionaries and instead are statistically and combinatorially based. The system shares methodological characteristics with other text analysis and content analysis programs such as with Alceste (Reinert, 1990), used also in the above mentioned analysis of Soviet Central Committee biographies, with the Smart system (Salton and McGill, 1983), with lexicometric studies (Salem, 1987), and more recently Lamalle and Salem (2002), and with Marchand's work (2007b).

- 16 Callope and the former Lexinet-Leximappe system are modular programs with the following sections: segmentation of the text which identifies each word and its context in the corpus; normalization of terms by examining those with similar suffixes; distributional filtering which eliminates the common everyday terms; managing synonyms through the use of a dictionary and the intervention of a human expert; identification and management of composite terms through a statistical study of their distribution in the corpus; management of the monoterms; and indexing of the documents with the selected terms or "key words". Each of these modules or sections of the program is structured in such a way that the human expert is called for to make the final decisions and choices concerning terms that are not automatically treated.
- 17 During this ongoing process, three data sets are continually up-dated: the lexicon of single and composite "key words"; the anti-lexicon of discarded words; and a dictionary of synonyms. To start with, all three data sets can be empty and built up or "learned" as the analysis progresses, which was the case in the analysis of the 2006 *Intelligence* data set. The feed-back process between the automatic procedure and the intervention of the human expert results in the creation of a lexicon of significant terms for the corpus under consideration. The lexicon can be corrected by different interactive procedures. The lexicon can also be "imported" from previous work on a similar corpus, as can the anti-lexicon and the dictionary of synonyms.
- 18 An index can be constructed by distributing or "flaging" the lexicon terms or key words throughout the units. The intervention of the human expert concerning his/her subjective choice of terms to be retained and synonym associations can be controlled for by having several human experts intervene jointly. The stability of the results concerning word sequences obtained in this manner was proven by the quasi identical list of the 17 most frequent word sequences found in the Soviet biographies by Lexinet and Alceste mentioned above. The only differences were in the number of occurrences and non-identical similarity between very similar chains of words. These differences were due to the fact that Alceste counts only exact correspondences while Lexinet regroups plurals and other similar expressions under one type of expression.
- 19 The next step is the application of the mapping program to the results generated by the indexing program. The dynamically produced lexicon and the indexed units are the raw material of the mapping program which structures this material according to the

associations between terms in the corpus under study and constructs the themes of interest as components or combinations of components revealed within this structure. The essential concept in this form of co-word textual analysis is co-occurrence of lexicon words in the same unit of analysis, be it geographical names in official Soviet biographies, key words in an abstract of a scientific article or terms in a media headline or title. The statistical index of proximity used as a measurement of co-occurrence of words was developed by Michelet (1988) and remains stable despite the size of the corpus analyzed. Called the index of equivalence, it is incorporated in the Calliope program and is defined by the following equation:

$$E(i,j) = C(i,j) / C(i) \times C(j) / C(j) = C^2(i,j) / (C(i) \times C(j))$$

where $C(i,j)$ is the number of simultaneous occurrences of words i and j and $C(i)$ and $C(j)$ are, respectively, the total number of occurrences of words i and j in the corpus. $E(i,j)$ measures the reciprocal inclusion of words i and j . It is a product of the probability that word i is included in a document containing j and the probability that word j is in a document containing i . The index E tends towards 1 when $C(i,j)$ tends towards $C(i)$ and $C(j)$. The index is higher the more words i and j are closely associated and is advantageous for associated words whose frequencies are similar. But it does have the advantage of not privileging any specific frequency zone in particular.

- 20 The calculation of $E(i,j)$ with Calliope is possible for all pairs of a list of 600 lexicon words. To manage and analyze this mass of co-occurrence information and extract the most significant aspects, Calliope builds clusters or components from the pairs of words with the highest values of $E(i,j)$. The most central word of a clusters is used as the "label" for that cluster. The former Lexinet-Leximappe program imposed a limit of ten members to a cluster, and when this limit was reached, the cluster was withdrawn and stored. This limit was removed with Calliope and a measure of statistical significance is used to delimit the size of any particular cluster.
- 21 This process of cluster construction continues until no further clusters can be formed. The majority of components depict a precise theme but the program does not always allow one unequivocally to associate a theme with each component or cluster. It is sometime necessary to regroup or eliminate certain collections of words. To chose the important components and to aid in analyzing them, Calliope generates two criteria: a statistical mean of the external ties between components members and key words outside the component; a statistical measure of internal cohesion defined as the mean strength of associations between component members (mean co-occurrence of key words).
- 22 These two measures are used, respectively, as the first axis (called centrality) and the second axis (called density) to form a "strategic diagram" on which each component or cluster can be projected by placing the cluster's "label" or most central key word on the two-dimensional graphic. This means that the more a cluster is to the right in the diagram, the more numerous are the ties of its member key words to other key words outside the cluster: the more it is "central". The vertical axis means that the more towards the top of the diagram a cluster is situated, the more the ties between key words within the cluster are dense: the more the cluster is "dense". These two axes divide the strategic diagram into four quadrants which are the first quadrant (upper right-hand), the second quadrant (upper left-hand), the third quadrant (lower left-hand) and the fourth quadrant (lower right-hand).
- 23 This means, by definition, that the first quadrant has relatively dense and central clusters and we use the nickname "mainstream" to describe it (Vinck, 1991). The second quadrant,

with dense but non-central clusters, is labeled "ivory tower". The third quadrant, with non-dense and non-central clusters, is called "unstructured". The fourth quadrant, with central but non-dense clusters, has been labeled "bandwagon". Although some researchers have suggested that new clusters may be formed in the "unstructured" third quadrant and move toward the other quadrants, our own research tends to show that clusters appear either briefly in the "ivory tower" second quadrant before disappearing, or else in the "bandwagon" fourth quadrant before moving into either the "mainstream" first quadrant or towards the "ivory tower" (van Meter, 1995; van Meter and Turner, 1997).

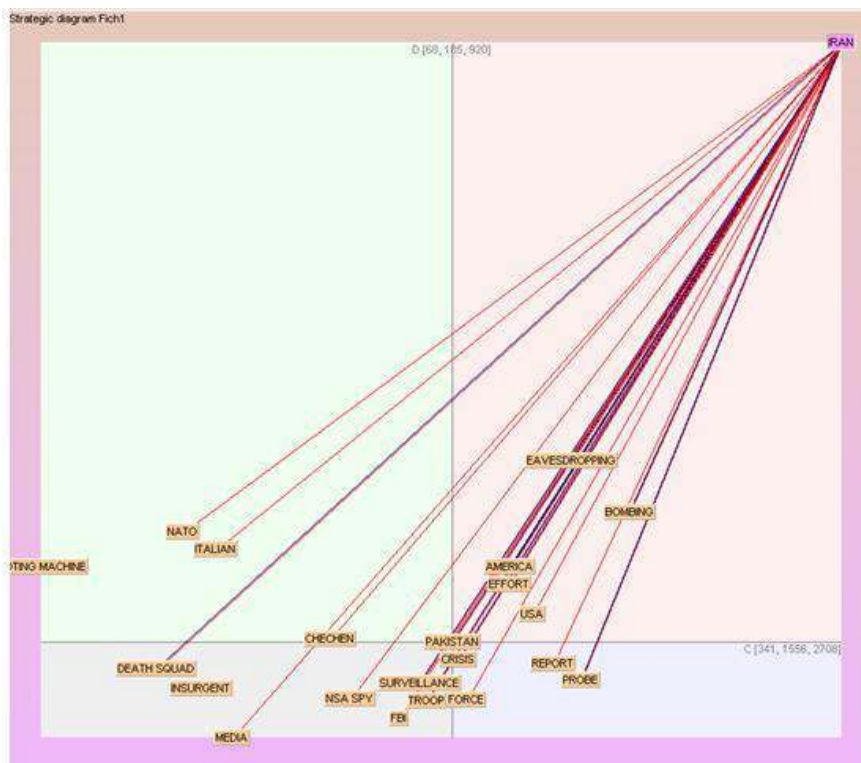
2006 Strategic Diagrams

- 24 With the division of our 2006 data set into three chronologically successive sections, Calliope produced the three strategic diagrams below. Perhaps the first things one notes is that "Iraq" does not dominate in all three diagrams, as would be expected with the massive dominance of the Iraq war in 2006 political and parapolitical developments. The presence of "Iran" in the "mainstream" first quadrant in the first (winter-spring 2006) section shows that the Bush White House intention to attack Iran and Iran's nuclear program played a major political and parapolitical role during the first part of 2006 before diminishing significantly toward the end of the year⁶.
- 25 In the first two diagrams, one should note the important position of "eavesdropping" (first diagram) and "NSA" (second diagram) in exactly the same position. The first diagram also includes "NSA spy" and "surveillance" clusters near the origin, while the second diagram has a "spy" cluster near the origin. These clusters are related to the Bush White House program of warrantless eavesdropping on American citizens by the NSA. However, the third diagram includes only a "surveillance" cluster in unstructured third quadrant. Nonetheless, the attractive power of the term "spy" appears fairly constant over the three periods of 2006 (see further below). With the Iraq war as the overall framework, *Intelligence* stated, in late February: "Although the Bush White House is plagued with numerous scandals and crises, the three current major ones are Iran and its nuclear program, the 'domestic' NSA spying scandal, and the series of reports on US torture/abuse and calls for the closure of Guantanamo." This corresponds with the position of "Iran" and "eavesdropping" in the first diagram, and the continued permanence of the latter throughout the year when "Iran" and "torture" diminished in importance.
- 26 Since a variety of scandals dogged the Bush White House throughout 2006, but would be treated in the headlines with different terms and therefore analyzed differently⁷, it is logical that "Iraq", "Iran" and "eavesdropping" tend to remain important over the entire year. What is of particular interest is situating the 2006 political "yearbook" events on the strategic diagrams and follow their evolution over time.
- 27 The three most notable 2006 "yearbook" events were another invasion of Lebanon by Israel (July), a North Korean nuclear weapon test (October), and a US Democrat Congressional election victory (November). These events are duly situated by the analysis as the respective clusters "Israel" and "Lebanon" (second diagram), "North Korea" (second diagram) and "Congress" (third diagram), but only as passing major events. "Israel", "Lebanon" and "Gaza" figure only in the second diagram and not in the "mainstream" first quadrant and nowhere else during the year. North Korea figures in

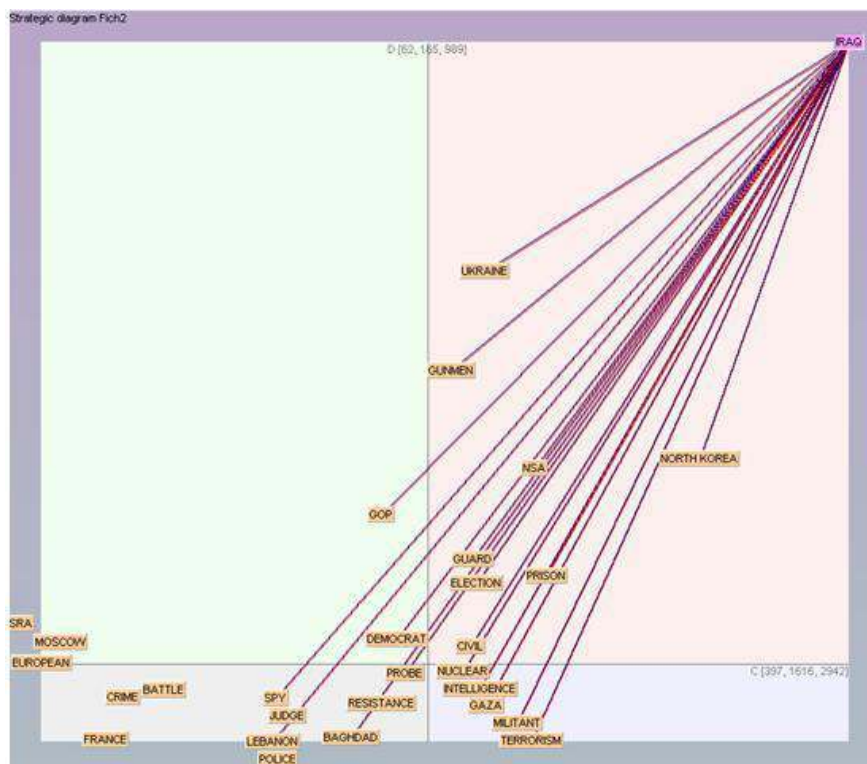
the "mainstream" first quadrant, but only in the second diagram and nowhere else. "Congress" figures only in the third diagram and in the "ivory tower" second quadrant. However, preparations for those important elections were widely commented on during the preceding second period and account in large part for the presence of the clusters "GOP" and "Democrat" in the "ivory tower" second quadrant and "election" in the "mainstream" first quadrant, thus showing the relative importance of the November Congressional elections.

- 28 Other clusters that appear in only one diagram situate important but passing events such as several Chechen war lords captured or killed ("Chechen" between the second and third quadrants in the first diagram), the Italian general elections (April - "Italian" in the "ivory tower" quadrant of the first diagram), a major political crisis in Ukraine (July 2006 - "Ukraine" in the "mainstream" second diagram), and an elected Hamas government official killed by Israelis (November 2006 - "Hamas" near the origin of the third diagram).

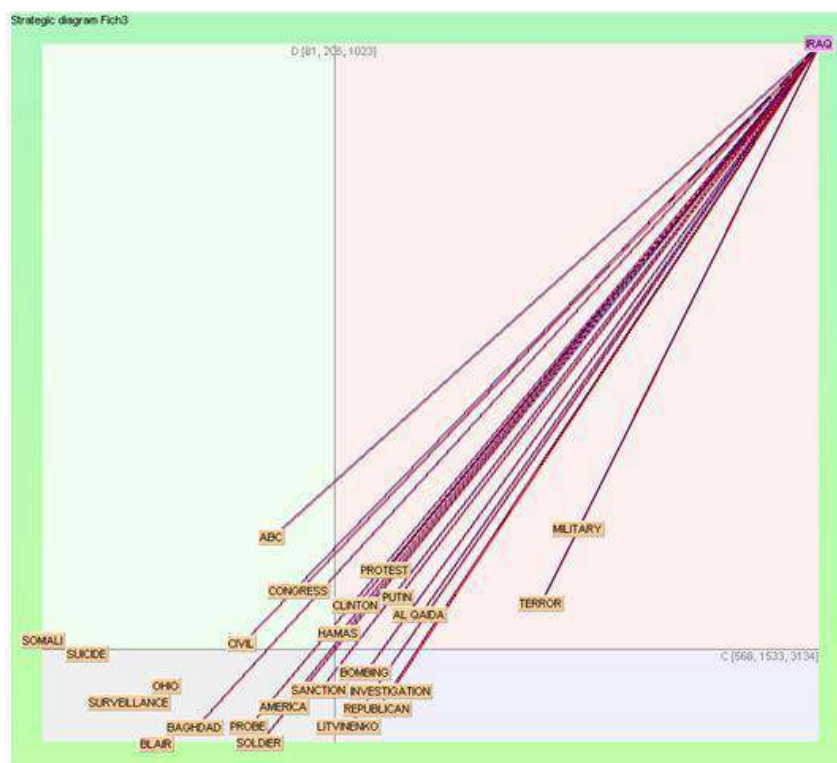
Strategic Diagram 1: First Three Periods, 2006



Strategic Diagram 2: Second Three Periods, 2006



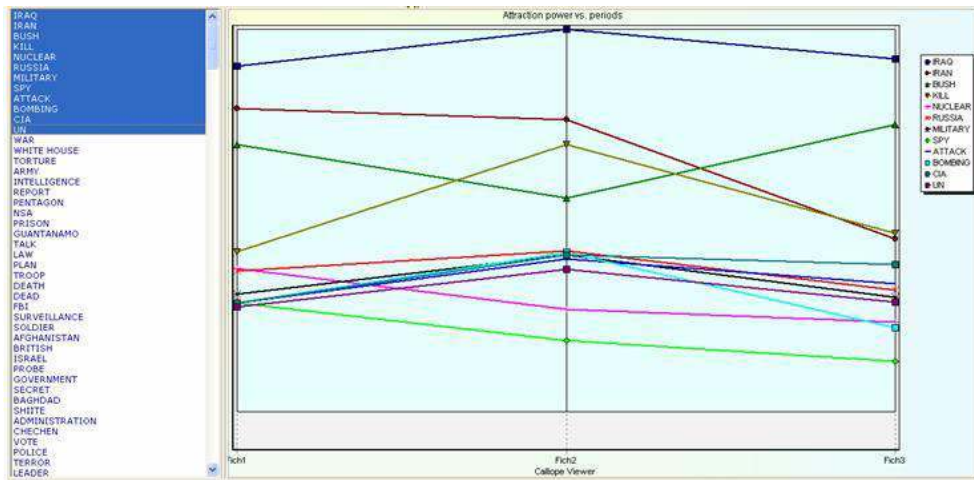
Strategic Diagram 3: Third Three Periods, 2006



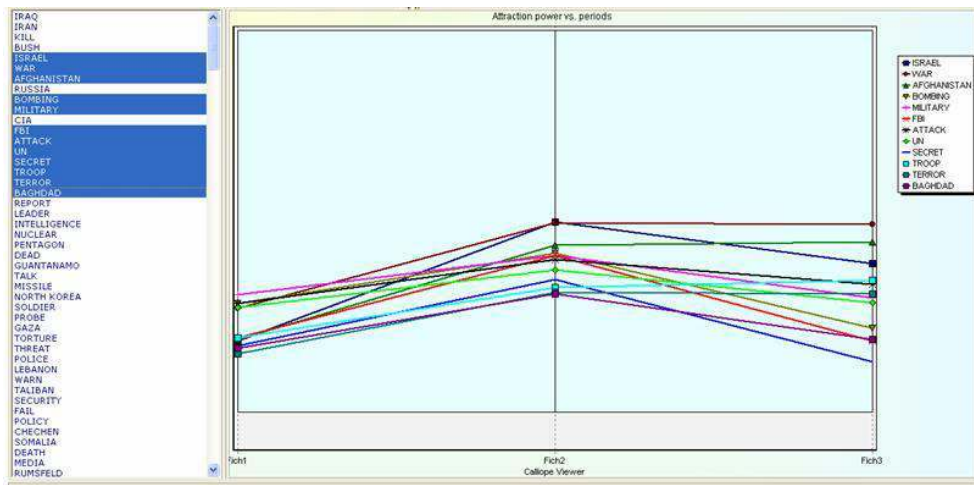
Key Word "Power of Attraction"

- 29 Following the formation of certain clusters backward in time and also forward in time is a necessary form of analysis which is greatly assisted by the following graphs of "power of attraction" of particular key words over our three periods of 2006. The first graph follows the major key words of the first period to show the evolution of their power of attraction over the entire year (that is, over all three periods). The second graph takes the dominate terms of the second period (minus those already appearing in the first graph) and shows their power of attraction over the entire year. And the third graph does the same for the dominant terms of the third period.
- 30 In the first graph, one can see the stable and massive dominance of "Iraq" while that of "Iran" starts strongly decreases slightly and then declines to fourth place. One can easily bet that 2007 will show exactly the opposite for "Iran". One can hypothesize that the high-low-high graph for "Bush" has to do with increasing Bush White House scandals that took attention away from Bush directly until the November election victory of the Democrats which brought Bush directly back into the news. The low-high-low graph for the key work "kill" seems to be associated with a significant second period increase in deaths in Iraq and Afghanistan, in particular. The other terms in the graph show relative stability.
- 31 The graphs for the second and third period show much more stability for the power of attraction of the selected key words.
- 32 The next two graphs represent, respectively, the "emerging" and the "declining" key words of the first period. By emerging, we mean terms that were relatively unimportant in the first period and which became more important later. By declining, we mean exactly the opposite: relatively important terms in the first period which became less important later.
- 33 The emerging term graphic appears fairly coherent except for the key word "abduction". Its increased importance has to do with the opening of CIA abduction trials in Italy and Germany, among other places, and the abduction of an Israeli soldier near Gaza.
- 34 The declining term graphic also appears fairly coherent except for the key words "China" and "Abu Ghraib". The evolution of "China" has to do with its role in international negotiations concerning Iran and North Korea. Once North Korea tested a nuclear weapon, direct negotiations were opened between the United States and North Korea, without China. Similarly, the decrease in importance of Iran as a topic also led to the decrease in importance of China. The decline of "Abu Ghraib" is nearly linear and corresponds with the end of the "press cycle" concerning torture by the US military at the Abu Ghraib prison in Iraq, a scandal that exploded in May 2004 and was in the headlines for more than a year and a half.

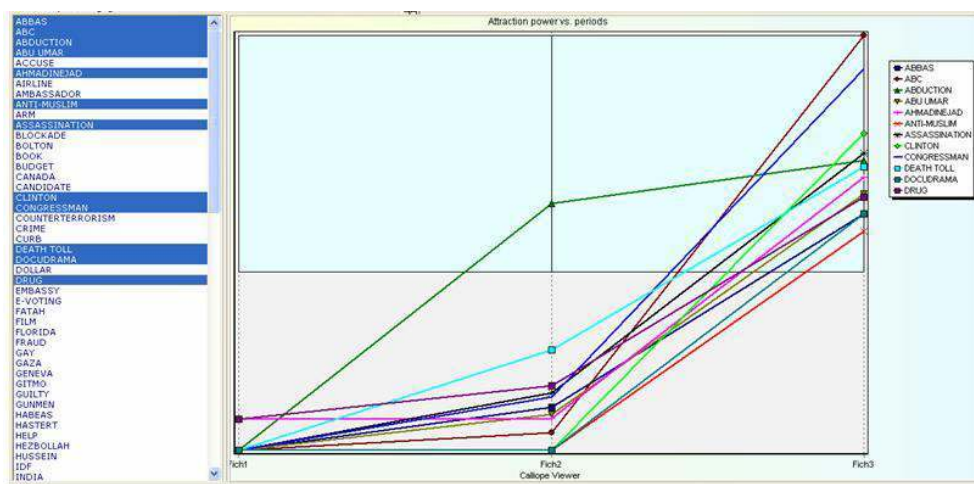
Key Word "Power of Attraction": Major First Period Terms



Key Word "Power of Attraction": Major Second Period Terms



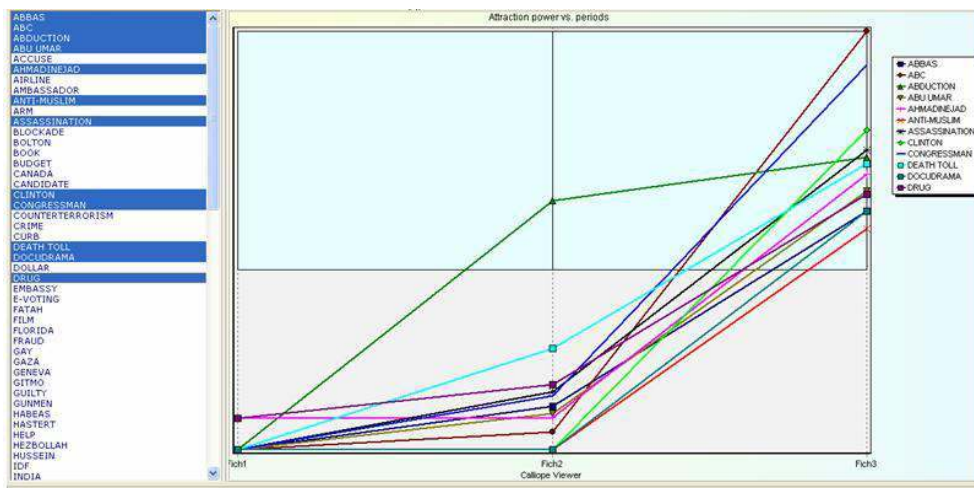
Key Word "Power of Attraction": Major Third Period Terms



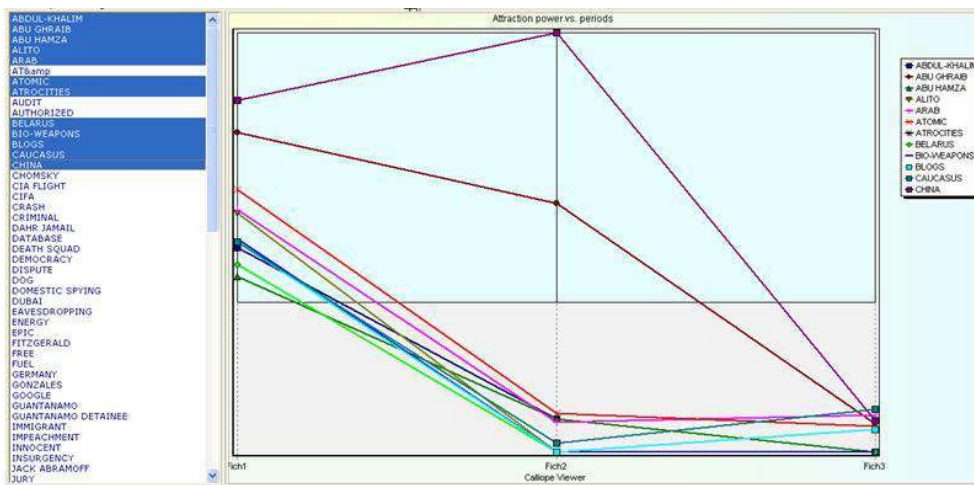
Conclusions, Limits & Weaknesses

- 35 By looking closely at the first quadrant in the first period, one can see some of the weaknesses of text analysis of media reports. The clusters "bombing", "America" and "effort" all figure in the first quadrant. "Bombing" is massively associated with the Iraq war (over 15 times), but also the Bali bombing, bombing in Afghanistan and other divers occurrences of "bombing". But even "bombing" and the Iraq war is not a coherent topic since it includes insurgent bombing and bombing by the US military, both of which were often in the news.
- 36 The cases of "effort" and "America" are even less coherent. "Effort" is associated with the FEMA (Federal Emergency Management Administration), the CIA, the NYPD (New York City Police Department), Iran, the NSA-FISA-eavesdropping program, the FBI, Vice President Cheney and with Bush White House reclassification of secret administration documents. "America" can be part of "Americans" and even part of "Latin America", and is associated with over 14 different topics with little meaning in common.
- 37 The same critique can be leveled against the terms "gunmen", "guard" and "prison" in the first quadrant of the second period, and "protest", "military" and "terror" in the first quadrant of the third period. It is only by following key words or clusters backward and forward in time that coherent political developments can be identified, unless of course it is a question of a major political development. The case of "Hammas" in the third period mentioned above is a good example of both how a political development can be identified and followed with this analysis, and how the international media can bias the coverage of a major political development.
- 38 The case of "Hammas" is quite interesting in its implications for text analysis of political developments. If a "Hammas" cluster appears in the third diagram, it's because the international media concentrated on the killing of an elected Hammas government official by the Israelis in November 2006. But the truly important political development associated with Hammas occurred in January 2006 when Hammas won the elections in Gaza and the Occupied Territories, thus causing a major international change in perspective that should have been widely commented in the press, but was not. Again, in July, Israel detained a large number of Hammas government officials, against the wishes of the larger international community. But, again, Hammas does not appear in the second diagram, and the international media gave only passing attention to this imprisonment of elected Palestinian officials. It was only in November with the death of one such official that the international media finally seemed to pay attention to what was happening.
- 39 Many similar instances of such "uneven" or clearly biased coverage of particular topics or events can be discerned by following the formation of certain clusters back in time and also forward in time. Inversely, if a topic or cluster cannot be followed over time, it is very likely a transient event or not a coherent topic.
- 40 Of course, there is a tremendous advantage in having at hand both the Calliope analysis of the 2006 *Intelligence* data set and the data themselves in digital format. Questions generated in interpreting the Calliope results can be quickly answered by using simple computer search functions of the data set. The ADI has recognized this advantage and is planning to publish the data set and the Calliope results in an integrated fashion to allow interested parties to continue the analysis we have initiated here.

"Emerging" Key Words Over Three Periods



"Declining" Key Words Over Three Periods



BIBLIOGRAPHY

Fabrizio Calvi and Olivier Schmidt (1988), *Intelligence secrètes - annales de l'espionnage*, Hachette, Paris.

Ghislaine Chartron, Alain Degenne, Marie-Odile Lebeaux and Lise Mounier (1990), "Images de la recherche sociologique actuelle", *Bulletin de Méthodologie Sociologique*, n. 28, September 1990.

Mathilde de Saint Léger and Karl M. van Meter (2005), "Cartographie de le premier congrès de l'AFS avec la méthode de mots associés", *Bulletin de Méthodologie Sociologique*, n. 85, January 2005.

- C. Lamalle and A. Salem (2002), "Types généralisés et topographie textuelle dans l'analyse quantitative des corpus textuels", *Actes des 6e journées d'analyse des données textuelles*, Saint Malo.
- Pascal Marchand (2005), "Le grand oral de Dominique de Villepin", *Bulletin de Méthodologie Sociologique*, n. 87, July 2005.
- Pascal Marchand (2007a), "Un vert, ça va; Dix verts, bonjour les débats!", *Bulletin de Méthodologie Sociologique*, n. 94, April 2007.
- Pascal Marchand (2007b), *Le grand oral - les discours de politique générale de la Ve République*, De Boeck, Brussels.
- B. Michelet (1988), "L'analyse des associations", Doctoral Thesis, Université Paris VII.
- Michel Pinault (2006), *La science au Parlement - Les débuts d'une politique des recherches scientifiques en France*, CNRS Editions, Paris.
- Max Reinert (1990), "Alceste - Une méthodologie d'analyse des données textuelles et une application: Aurélia de Gérard de Nerval", *Bulletin de Méthodologie Sociologique*, n. 26, March 1990.
- André Salem (1987), *Le Lexicloud*, Ecole Normale Supérieure, Fontenay-Saint Cloud.
- G. Salton and M. J. McGill (1983), *Introduction to Modern Information Retrieval*, International Student Editions, McGraw Hill, New York.
- Olivier Schmidt (2005), *The Intelligence Files - Today's Secrets, Tomorrow's Scandals*, Clarity Press, Atlanta.
- Karl M. van Meter (1995), "Bounded Chaos Neural Network Model of Scientific Innovation", *Science and Science of Science*, vol. 9, n. 4.
- Karl M. van Meter (2002), "Terrorists/Liberators: Researching and Dealing with Adversary Social Networks", *Connections*, vol. 24, n. 3.
- Karl M. van Meter, Lise Mounier, Ghislaine Chartron and Max Reinert (1991), "Multimethod Analysis: Official Biographies of Members of the Central Committee of the Soviet Union Communist Party", *Bulletin de Méthodologie Sociologique*, n. 33, December 1991.
- Karl M. van Meter and William A. Turner (1997), "Representation and Confrontation of Three Types of Longitudinal Network Data from the Same Data Base on Sociological AIDS Research", *Bulletin de Méthodologie Sociologique*, n. 58, September 1999.
- Karl M. van Meter, Philippe Cibois and Mathilde de Saint Léger (2004), "Correspondence and Co-Word Analysis of Ten Years of BMS Articles 1993-2003", *Bulletin de Méthodologie Sociologique*, n. 81, January 2004.
- Dominique Vinck (editor) (1991), *Gestion de la recherche: Nouveaux problèmes, nouveaux outils*, De Boeck, Brussels.

NOTES

1. See the *Intelligence* Web site at <http://perso.orange.fr/intelligence-adi>
2. Not necessarily English-language media, but any media report that has been translated or summarized in English so that subsequent computerized analysis can be done by a system that handles only one language at a time
3. Sometimes a "Timeline" issue of *Intelligence* is published in the middle of summer vacation.

4. This means the issues dated 2006, which are issues from n. 471 — although it covers the period 9 December to 31 December 2005 — to n. 489 — dated 18 December 2006 — and the n. 490 (8 January 2007) issue whose "Timeline" article covers media reports from 16 December to 31 December 2006.
5. See the Calliope Web site at <http://www.calliope-textmining.com/>
6. But this topic came back as a major theme in middle and late 2007.
7. For examples of Bush White House scandals, see "voting machine" in the first diagram, "Ohio" in the third, "GOP"/"Republican" in the last two, and "probe" in all three.

ABSTRACTS

A system of text analysis by co-occurrence of key words, Calliope, is applied to the 2006 data set of daily headlines and titles of media reports on international political and "parapolitical" events produced by the journal, *Intelligence*. These political (publicly avowed use of political power) and parapolitical (unavowed use of political power) developments are divided into three successive four-month periods, each producing a two-dimensional "strategic digram" with a centrality and a density axis. Graphs over time of "emerging" and "declining" key words are also produced. These results permit an original and intriguing analysis and interpretation of these 2006 events.

Analyse textuelle par mots associés en sciences politiques – les titres 2006 de la politique et de la « parapolitique » internationales : Un système d'analyse textuelle par co-occurrence de mots clés, Calliope, analyse un ensemble de titres de rapports quotidiens des médias de 2006 sur des événements politiques et « parapolitiques » fourni par le journal, *Intelligence*. Ces événements politiques (utilisation publiquement avouée du pouvoir politique) et parapolitiques (utilisation non avouée du pouvoir politique) sont divisés en trois périodes successives de quatre mois, chaque période fournit un « diagramme stratégique » avec un axe de centralité et un axe de densité. Des graphiques longitudinaux de mots clés "émergeants" et "déclinants" sont aussi fournis. Ces résultats permettent une analyse et une interprétation originales des événements de 2006.

INDEX

Mots-clés: Analyse textuelle, Politique internationale, Parapolitique, Calliope, Analyse par mots associés

Keywords: Text Analysis, International Politics, Parapolitics, Co-Word Analysis

AUTHORS

KARL M. VAN METER

(CMH-CNRS-ENS-EHESS, karl.vanmeter@ens.fr)

MATHILDE DE SAINT LÉGER

(CNRS-DIS, mathilde.de-saint-leger@cnrs-dir.fr)