Monitoring conflict using automated coding of newswire reports: a comparison of five geographical regions

Philip A. Schrodt, Erin M. Simpson, and Deborah J. Gerner

Department of Political Science
University of Kansas
Lawrence, KS 66045 USA
email contact: schrodt@ku.edu
project web site: http://www.ku.edu/~keds
phone: +1.785.864.9024
fax: +1.785.864.5700

Paper presented at the PRIO/Uppsala University/DECRG High-Level Scientific Conference on Identifying Wars: Systematic Conflict Research and Its Utility in Conflict Resolution and Prevention, Uppsala, Sweden 8-9 June 2001

This research was supported in part by the U.S. National Science Foundation grant SES-0096086 and by the General Research Fund of the University of Kansas

Abstract:

This paper discusses the experience of the Kansas Event Data System (KEDS) project in developing event data sets for monitoring conflict levels in five geographical areas: the Levant (Arab-Israeli conflict), Persian Gulf, former Yugoslavia, Central Asia (Afghanistan, Armenia-Azerbijan, former Soviet republics), and West Africa (Liberia, Sierra Leone). These data sets were coded from commercial news sources using the KEDS and TABARI automated coding systems. The paper discusses our experience in developing the dictionaries required for this coding, the problems with the number of reported events in the various areas, and provides examples of the statistical summaries that can be produced from event data. We also compare the coverage of the Reuters and *Agence France Presse* news services for selected years in the Levant and former Yugoslavia. We conclude with suggestions for four topics where additional efforts that could be usefully undertaken by multiple research projects. The coding software, coding dictionaries and data are available at the KEDS web site, http://www.ku.edu/~keds.

1. Introduction

This paper discusses the experience of the Kansas Event Data System (KEDS) project in developing event data sets for monitoring conflict levels in five geographical areas: the Levant, Persian Gulf, former Yugoslavia, Central Asia, and West Africa. Event data—nominal or ordinal codes recording the interactions between international actors as reported in the open press—break down a sequence of political activities into a sequence of basic building blocks (e.g., comments, visits, rewards, protests, demands, threats, military engagements) reported on a daily basis. The paper discusses our experience in developing the dictionaries required for coding events from the commercial newswire services Reuters and *Agence France Presse* (AFP) using the KEDS and TABARI automated coding systems, and provides a number of examples of the statistical summaries that can be produced with event data. We also discuss the problems with the number of reported events in the various areas and compare the coverage of Reuters and AFP news services for selected years in the Levant and former Yugoslavia. We conclude with suggestions for four topics where additional efforts that could be usefully undertaken by multiple research projects.

Over the past forty years, a large variety of research projects in international and comparative politics have made use of event data (see Azar, Brody & McClelland (1972); Burgess & Lawton (1972); Sigler, Field, & Adelman (1972); Azar & Ben-Dak (1975); and Merritt, Muncaster, & Zinnes (1993)). We have defined an event as:

> …an interaction, associated with a specific point in time, that can be described in a natural language sentence that has as its subject and object an element of a set of *actors* and as its verb an element of a set of *actions*, the contents of which are transitive verbs. (Gerner, Schrodt, Francisco, and Weddle 1994, 95).

Groups of events can then be aggregated by week, month, quarter or year for a summary measure of political activity; this summary is usually in terms of net cooperation or conflict between a pair ("dyad") of actors.

Event data research has recently experienced a renaissance. In the 1960s and 1970s event data were among the most commonly used tools in the study of international relations, yielding Charles McClelland's (1976) World Event Interaction Survey (WEIS) dataset and Edward Azar's

(1982) Conflict and Peace Data Bank (COPDAB), among others. During this period much attention was directed toward the possibility of developing early warning programs, such as the Early Warning and Monitoring System (EWAMS) funded by the US Defense Advanced Research Projects Agency (DARPA) (see Laurance 1990). However, the failure of these systems to meet expectations contributed to the decline of event data research. A second factor was criticism from the academic community itself (e.g., Vincent 1983) related to the face validity of event data. Plagued by a decline of funding and a lack of credibility within the academic community, event data research waned significantly during the 1980s.

The 1990s, however, brought new advances in event data research, particularly with the support of the U.S. National Science Foundation's Data Development in International Relations (Merritt, Muncaster, & Zinnes 1993). While the original WEIS and COPDAB datasets were hand-coded by teams of university students pouring over newspaper microfilms, by the 1990s, personal computers and machine-readable text allowed for automated coding of event data. Using KEDS and Reuters, completed a number of different tests to establish the face validity of this data with events in the Middle East (Schrodt and Gerner 1994). We concluded that this denser, automated coded data did, in fact, accurately reflect events on the ground and generally correlated well with human-coded WEIS datasets for similar dyads. To date, machine-based event data and time-series analysis have been used to evaluate triangulation and reciprocity in the Balkans (Goldstein & Pevehouse 1997) and Middle East (Goldstein, Pevehouse, Gerner, & Telhami forthcoming), analyze foreign policy decision making (Gerner 1993), develop early warning systems of political instability (Schrodt & Gerner 1997, 2000, Schrodt 2000, 2000a) and study intrastate civil conflicts (Bond, Jenkins, Taylor, & Schock 1997, Huxtable 1997, Thomas 1999).

1.1. The KEDS Project

For the past decade, Gerner and Schrodt, along with a number of collaborators, have been working on the development of the **K**ansas **E**vent **D**ata **S**ystem (KEDS), a computer program that creates event data from machine-readable text (Schrodt, Davis, & Weddle 1994). KEDS is a pattern-matching system that uses a computational method called "sparse parsing." Instead of trying to decipher a sentence fully, KEDS determines only the parts required for event coding—for instance, political actors, compound nouns and compound verb phrases, and the references of pronouns—and then employs a large set of verb patterns to determine the

appropriate event code. Unlike more complex full parsing, sparse parsing techniques can be used successfully on unedited news wire text. We have experimented with coding a variety of texts, including specialized regional sources in English and German (Gerner et al. 1994). Most of our work, however, has been with Reuters News Service lead sentences. The lead is usually a simple declarative sentence that summarizes the article, e.g., "The United Arab Emirates welcomed a resumption of formal diplomatic ties between Egypt and Syria after a 12-year rift." For closely reported crisis areas such as the Middle East and the Balkans, lead sentence coding provides thorough coverage of political events, but KEDS has also been successfully used to code complete stories in regions that are less well reported, such as West Africa (Huxtable 1997). The coding software, coding dictionaries and data developed by the project are available at the KEDS web site, `http://www.ku.edu/~keds`.

We have validated KEDS against both the textual record and human-coded events and found no systematic biases in machine coding (Gerner et al. 1994; Schrodt & Gerner 1994). Thomas (1999) found similar results in an independent validation for his KEDS-coded event data on Northern Ireland. When the human and machine-coded data are used in statistical tests, the results are almost indistinguishable except for differences due to the higher number of events in the machine-coded data. An independent test of KEDS by the **P**rotocol for the **A**nalysis of **N**onviolent **D**irect **A**ction (PANDA) project of the Program on Nonviolent Sanctions and Cultural Survival at Harvard's Center for International Affairs found that when coding dictionaries have been optimized for a set of data, the sparse-parsing methods of KEDS can assign event codes to Reuters leads with a 91% accuracy. Typical accuracy is in the range of 75% to 85%, similar to the accuracy of human-coded data (Burgess & Lawton 1972).

In the spring of 2000, Schrodt produced a new automated coding system named TABARI— **T**extual **A**nalysis **B**y **A**ugmented **R**eplacement **I**nstructions—that is based on the same sparse-parsing principles as KEDS (and hence can use dictionaries developed for KEDS) but is far faster and more flexible. KEDS was written in Pascal and worked only on the Macintosh operating system; TABARI is written as "open-source" code in ANSI C++ and is available on the Linux, Macintosh, and Windows operating systems. TABARI eliminates some deep-seated idiosyncrasies of KEDS and is substantially faster, reducing the time required to recode a data set from hours to minutes or even seconds.

We originally became involved with machine coding because, after initial start-up costs, it is dramatically faster and less expensive than human coding. Once a researcher has established vocabulary lists of actors and verb phrases, the only significant expense involved in generating event data is the acquisition of machine-readable news reports. Furthermore, a coding system developed at one institution can be used by other researchers through the sharing of vocabulary lists and coding software.

In working with KEDS, we discovered an additional advantage to machine coding: It is free of non-reproducible coding biases and is therefore both reliable and transparent. Human coding is subject to systematic biases because of unconscious assumptions made by the coders. For example, Laurance (1990) notes that even expert coders in the military tended to over-estimate the military capability of China in the 1980s because they knew China to be a large Communist country. When event coding is done part-time by students, coder biases are even more unpredictable and difficult to control. In contrast, with machine-coding the words describing an activity will receive the same code irrespective of the actors or time period involved. Any biases embedded in the machine coding system are preserved explicitly in its vocabulary and can be modified by the researcher; there is no such record in human coding and thus no ability to address this potential problem.

When the KEDS project began in the late 1980s, accurate machine coding was regarded as impractical. As late as 1998, an article on early warning dismissed automated coding as something beyond "our current (or foreseeable) knowledge" (Davies & Harff 1998:81).[1] These pessimistic assessments, however, did not take into account "Moore's Law"—the doubling of computer capacity every 18 months—which has made a desktop computer in 2000 roughly 250-times more powerful than a computer in 1988, when the NSF DDIR began.

With high-capacity computers, automated coding has proven to be an imminently tractable problem. During the past five years, machine-coded data have become completely accepted in the political science community as a viable—and in most cases, preferable—alternative to traditional human coding:

- ❖ Articles using KEDS-coded data have been published in the top peer-reviewed journals in political science, including the *American Political Science Review* (Goldstein & Pevehouse 1997; Edwards & Wood 1998; Wood & Peak 1999; Schrodt & Gerner 2000), *American Journal of Political Science* (Schrodt & Gerner 1994), *Journal of Conflict Resolution* (Bond et al. 1997; Schrodt & Gerner 1997; Pevehouse & Goldstein 1999), and *International Studies Quarterly* (Gerner et al. 1994).

- ❖ At least two dissertations have been completed that develop new event data sets using the KEDS system—Huxtable (1997) and Thomas (1999), and we know of additional dissertations in progress at the University of Michigan, The Ohio State University, Texas A&M University, and the European University Institute (Florence).

- ❖ Several government projects in the United States and Europe, as well as a United Nations project, have employed machine coding systems (KEDS, TABARI and the commercial VRA coder now used by the PANDA project) in crisis early warning systems.

2. Adapting automated coding to multiple geographical regions

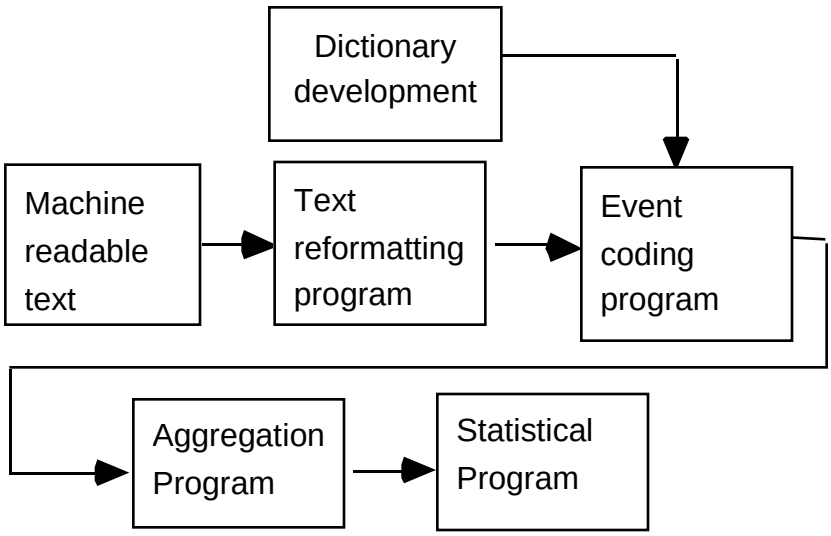
The automated coding process is illustrated schematically in Figure 2.1.[2] From the perspective of the researcher trying to use machine coding to create data for a specific region, the key bottleneck in this process is dictionary development. KEDS and TABARI use two dictionaries for coding, a .verbs dictionary containing verb phrases and an .actor dictionary containing proper nouns (e.g. `GEORGE BUSH`) and noun phrases (e.g. `SERBIAN MILITIA`). Dictionary development involves adding phrases to these dictionaries so that news reports for each region are correctly coded.

---

[1] Ironically, in December 2000, the University of Maryland's "Global Event Data System"—the project from which this quote derived—shut down its human coding operation and is now moving to the TABARI machine coding system.

[2] Additional discussions of machine coding can be found in Bond et al 1997, Gerner et al 1994, Schrodt & Gerner 1994, Huxtable & Pevehouse 1996, and Schrodt, Weddle & Davis 1994.

Figure 2.1. Developing machine-coded event data



Over the past two years we have developed a "standard" verbs dictionary that is a composite of several regional dictionaries that were developed by the KEDS project over the past eight years, including our original Levant dictionaries, the PANDA dictionary, Huxtable's West Africa dictionary (Huxtable 1997) and the Pevehouse's Balkans dictionaries (Goldstein and Pevehouse 1996).[3] After merging these dictionaries, we eliminated most verb phrases containing more than a half-dozen words on the grounds that such phrases would be repeated only rarely. The resulting dictionary contains most of the verb phrases used by Reuters to describe *international* political events, as well as discard codes for a wide variety of athletic events, natural disasters, and fatal mishaps involving various modes of transportation.

Dictionaries for the coding of *internal* political events present additional challenges and it is not clear whether it will be possible to develop a general-purpose .verbs dictionary for this task. When developing .verbs dictionaries to code internal events in a disparate set of states—Russia, China, Albania, Colombia, Mexico, Syria, Algeria, Pakistan and Nigeria—we found that the vocabulary referring to domestic events could vary significantly across regions. First, internal events involve a much larger set of verb phrases than international events and many of these

[3] These dictionaries were not developed independently so they contain very substantial amounts of overlap.

phrases are idiosyncratic to specific states. For example, reports on Colombia and Mexico reflect a great deal of large-scale, quasi-political criminal activity involving the trade in illegal drugs; this type of activity is rare in reports involving the Middle East, Africa or Europe. Africa, in contrast, involves quasi-political criminal activity involving the smuggling of diamonds that is not found elsewhere in the world. Islamic politics is important in the Middle East and parts of Africa; it is irrelevant to Latin America. When we coded Albania for the 1996-97 period, we encountered a series of events involving the collapse of pyramid investment schemes, followed by an almost complete breakdown of political order, followed by an international intervention that very quickly restored order. That sequence was quite distinct from the civil disorder we coded for Lebanon.

Second, despite the generally consistent style found in Reuters reports, reporters and editors in each geographical region employ certain distinct verb phrases. Each set of regional reports has a few idiosyncratic turns of phrase that we have not encountered earlier.[4] Because these phrases are common, they are discovered very quickly when spot-checking the dictionaries and actually simplify coding when events are reported using a small set of routine sentence structures. Nonetheless, the presence of idiosyncratic phrases means that a dictionary developed on one region will miss some important events if it is used, unaltered, in another region. These errors are almost exclusively false negatives—an idiosyncratic phrase used in one region will almost never correspond to a distinctly different behavior in another region.

Our conclusion from these projects is that anyone attempting to code internal events should invest time in customizing the standard dictionaries. If a sequence of very unusual events has occurred—the collapse of the Albanian financial system, for instance—it may be advisable to develop dictionaries specifically to code that period. Routine international behavior, on the other hand, can probably be coded reliably using standard dictionaries with some spot-checking for distinct phrasing and regionally-specific forms of political activity.

While the verb phrases used to describe international political events differ little across time and geographical regions, there is substantial variation in the political actors, particularly if one is coding sub-state and non-state actors. Consequently, any project focusing on a new geographical

region needs to supplement the .actors file. These files also need to be periodically updated with the names of new political leaders and, in the post-Cold War period, new states. If internal events are being coded, the required modifications can be quite extensive.

(An actor list that contains most of the frequently-occurring names is necessary for the program to be able to correctly analyze sentences using the sparse-parsing technique. A missing actor will be treated as an unknown word. If it occurs before the verb as the subject, then the sentence will not be coded and the event will be missed.[5] If the missing actor occurs *after* the verb as an object, the parser will continue to search for an actor it recognizes, and may erroneously code that actor as the target of the event. A positive side-effect of this, however, is that once a suitable list of regional actors has been developed in insure correct event coding, it is very simple to also assign these actor individual codes so that opposition leaders, branches of government and other sub-state actors can be coded.)

We have created a standard .actors dictionary that lists all significant states and international organizations in the international system, as well as major-power political leaders such as U.S. presidents, European prime ministers, and heads of UN organizations. This can be supplemented with a set of standard noun phrases of the form

```
<NATION'S>_LABOUR_MOVEMENT [NNNLAB]
<NATION'S>_LEADER [NNNGOV]
<NATION'S>_LEGISLAT [NNNGOV]
<NATION'S>_MILITARY [NNNMIL]
<NATIONAL>_ACTIVIST [NNNPOL]
<NATIONAL>_AGENT [NNNGOV]
<NATIONAL>_AIR_FORCE [NNNMIL]
<NATIONAL>_ARMY [NNNMIL]
```

By using a global search-and-replace on the placeholders `<NATION'S>` and `<NATIONAL>`, and the code placeholder `NNN`, state-specific phrases can be created and added to the .actors dictionary.

While it is possible to detect new actors by going through the source texts manually, that process is quite labor intensive because an opposition leader or group may achieve “fifteen

---

[4]When developing a dictionary, there are times when one has a sense of looking over the collective shoulder of Reuters: For example an indicator that Reuters is using inexperienced reporters (or over-worked editors) to cover a crisis are reports containing two consecutive apostrophes (' ') instead of a quotation mark (").

[5] Actually, skipping sentences where no actor is found prior to a codeable verb is an option in KEDS and TABARI, but we have used this in all of the data sets we have coded.

minutes of fame" somewhere in the middle of the data set. To deal with this problem, we have partially automated the process of identifying new actors by using a computer program called Actor_Filter (available in both C and Java versions). This software tabulates phrases that possibly refer to new political actors based on patterns of consecutive capitalized words. The output of this program is a "keyword-in-context" (KWIC) file of the actors that cannot be found in an existing KEDS dictionary, listed in order of frequency.

Table 2.1 shows an example of the first records of this file from an Actor_Filter index on events dealing with Algeria. In the KWIC format, the actors are highlighted with <<...>>. The first two sets of records show two common actors that were not already in the dictionary—the Islamic Salvation Front (209 occurrences) and President Liamine Zeroual (182 occurrences)—and also highlights other proper nouns such as Air Algerie and Ali Belhadj.[6] The third set of records identifies a non-political name that is apparently very common in the texts—the *El Watan* newspaper, which Reuters frequently uses as a source.

Reuters will refer to an actor using a variety of different phrases. For example, Algerian President Liamine Zeroual might be referred to using any of the following formulations:

```
Algerian President Liamine Zeroual
Algerian President Zeroual
President Liamine Zeroual
President Zeroual
Liamine Zeroual
Zeroual
```

Because of this, most major individual politicians such as chief executives require multiple entries. These are almost invariably subsets of the general structure

<nation name> <title> <first name> <last name>

[6] These records also show some problems with the consecutive-capitalization rule, particularly when dealing with languages other than English. For example, Reuters does not capitalize the Arabic article "al-", so "Bashir al-Assad" is not seen as a single phrase. The French proper noun "L'Authentique" fails the test because of the contraction; "Liberte" because it is only one word. More sophisticated rules could be developed to deal with these cases, but at the expense of a greater number of false positives. One still has to deal with idiosyncrasies such as the transliteration "El Watan" rather than "al-Watan." ("The Nation")

Table 2.1. Actor_Filter KWIC File Output

```
Islamic Salvation Front  (209)

960605  REUT-0004-04
The banned <<Islamic Salvation Front>> (FIS), which has been battling to topple
the Algerian government since 1992, has increased contacts with Libyan
authorities seeking information about the missing men, al-Hayat said

960611  REUT-0002-01
Lawyers defending <<Ali Belhadj >> a leader of the banned <<Islamic Salvation
Front>> (FIS), said on Tuesday that Algeria had arbitrarily detained their
client for the past year in a secret location

960612  REUT-0007-07
France has consistently refused to have any official contact with the <<Islamic
Salvation Front>> (FIS), deprived of almost certain victory in 1992 general
elections when the army intervened to cancel the second round of voting

==========================================

President Liamine Zeroual  (182)

960606  REUT-0003-22
<<President Liamine Zeroual>> has started a dialogue with political party
leaders and other prominent figures, excluding radical Islamists, in his quest
for a peaceful settlement after winning the presidential elections in 1995

960610  REUT-0001-02
It said <<President Liamine Zeroual>> accepted the resignation last week of
Mohamed Benchercheli appointed in 1994

960611  REUT-0002-03
They said they had appealed to the authorities several times, writing to
Algeria's justice minister and to <<President Liamine Zeroual>> among others
but to no avail

==========================================

El Watan  (137)

960609  REUT-0004-03
The newspapers, including the best known dailies <<El Watan >> al-Khabar,
L'Authentique and Liberte, did not appear on Sunday for the seventh day

960610  REUT-0001-01
The director general of <<Air Algerie>> resigned from the troubled state-run
airline, the Algerian newspaper <<El Watan>> reported on Monday

960610  REUT-0001-08
<<El Watan>> said there was also a sharp drop in revenue when <<Air Algerie>>
suspended direct flights to Paris last July after French aviation authorities
insisted it switch from Paris Orly airport to Roissy
```

and one could presumably use this regularity to further automate the dictionary development process. For example, if the system recognized "Hobbit Liberation Front" as a category [HLF] and encountered the unknown proper noun "Frodo Baggins" in the context, "Frodo Baggins, a leader in the Hobbit Liberation Front", it would tentatively assign "Frodo Baggins" to the category [HLF].

The Reuters editorial style appears to specify that whenever a relatively unknown actor is introduced in a story, he or she is identified by nationality. This characteristic makes the coding of international events relatively easy, because one knows the nationality of an actor even if the individual actor is not in the dictionary. It is less helpful in the coding of internal events, where the actor's national identity can generally be assumed from the context of the story. For example in the lead:

```
Moslem guerrillas killed 14 people overnight in Tabainat village in Blida
province, 50 km (30 miles) south of Algiers, Algerian security forces said
on Monday.
```

both the guerrillas and the people killed are Algerian, but this is not stated explicitly.[7] The KEDS program contains some specialized routines (developed for the PANDA project) that identify geographical location, but we have not used these systematically.

The rule of thumb that we've been using is to include actors in a regional dictionary if they occur in more than 0.01% of the lead sentences of a data set (e.g. in developing our Gulf data set, which contained approximately 80,000 leads, we added any actors that occurred eight or more times). Because there are a large number of low-frequency actors, the proportion of source texts containing uncoded actors is substantially larger than 0.01%, but the addition of any single actor would change the coding no more than one out of every 10,000 leads. This actually over-states the error, because rare proper names usually contain other identifying information — "Afghan Minister of State for Foreign Affairs Najibullah Lafraie" or "leader of the Tajik armed opposition, Sayed Abdullo Noori"—that permit correct identification. Most of the proper nouns that occur in fewer than 0.01% of the leads involve individuals who are briefly newsworthy, such as hostages or victims of terrorism, election candidates from small political parties and the like.

Once the dictionaries have been suitably refined, the data should be recoded in fully automatic mode—a process we call “autocoding”—to ensure that the coding rules are consistently applied across the entire data set. If only part of the data set is machine-coded, with occasional records manually "corrected", then inconsistencies will be introduced into the time series that might show up as statistical artifacts later in the analysis. Autocoding also insures that the coding can be replicated by later researchers and can be updated.

The speed of autocoding clearly depends on the speed of the computer, but the following statistics will give some indication of the time involved. During most of the project we were using computers in the 100 Mhz range; these coded around 11 events per second, so a 80,000 event data set such our Levant and Gulf cases required around two hours to recode using KEDS. (This processing time includes the evaluation of texts that do not produce events.) On machines in the 300 Mhz range, the speed increased proportionately to around 45 events per second.

While KEDS seemed quite fast compared to human coding, TABARI proved to be even faster by almost two orders of magnitude, coding 2000 events per second on a 350Mhz Macintosh G3. This is roughly 70 times faster than KEDS on the same machine.[8] As a rule of thumb, human coders can reliably produce 40 events per day, so TABARI running on a G3 does in a second what a human coder does in about three months. This is a wall-clock speedup of around a factor of 7.8-million. This speed means that we are close being able to actively experiment with the implications of changing dictionaries, i.e. a researcher could change a dictionary entry and then plot the results for a ten-year series within a few seconds, rather than looking only at a single text.

---

[7] In Gerner et al (1994) we describe a similar problem we encountered when trying to code chronologies from the *Journal of Palestine Studies*. *JPS* assumed certain nationalities were known, so that the statement "Israeli police beat Palestinian demonstrators" was always rendered as simply "Police beat demonstrators."

[8] On a 650Mhz Dell Pentium III running under Linux, the speed is around 3000 events per second. Timing tests were done using 26,000 Reuters leads for the Levant for 1987-1990, and using the default ("None") autocoding mode that provides no screen feedback.
The increase in speed is probably due to two factors. First, the current version of KEDS was compiled using a ten year old Pascal compiler that could not take advantage of speed-enhancing features of contemporary machines such as pipes and caches. In fact KEDS was running in 680x0 emulation rather than in native code; and KEDS minimal screen feedback also may have slowed the program substantially. Second, C generally is written closer to the machine than Pascal, and both Metrowerks CodeWarrior for the Macintosh and GNU g++ for Linux. are known for producing very fast code. While nominally written in C++, virtually all of TABARI is actually in C, and makes almost no use of time-consuming operations such as object instantiation and destruction. While the code in TABARI is a lot cleaner than KEDS, generally the algorithms haven't changed much.

3. Event Data in Five Conflict Regions

This section will discuss some of the features of the regional data sets that we have developed using the KEDS and TABARI programs. All of the data sets, and the dictionaries used to code them, are available from the KEDS web site. For simplicity of display, most of the dyadic data are shown as monthly aggregations using the Goldstein (1992) scale, but it should be kept in mind that the data themselves consist of individual daily events. All of these data sets are based on Reuters stories downloaded first from NEXIS, and later from the Reuters Business Briefing service. Unless otherwise noted, the series are based on coding lead sentences only.

The content of the various data sets has been defined by the stories (or lead-sentences) that were *downloaded* for coding. For example, the "Levant" data set looks for stories that involve Egypt, Israel, Jordan, Lebanon, the Palestinians, or Syria. However, within these stories we code for a much larger set of political actors. For example all of our dictionaries include the names of almost all nation-states and major international organizations. Consequently if the prime minister of Malaysia visits Lebanon, this will be picked up as a Malaysia-Lebanon event, even though Malaysia is not part of that region. A number of events will involve none of the actors in the region, as in the case of a discussion between the United States and France about the Palestinians. The sensitivity of the data to the search criteria will be discussed in more detail in the next section.

3.1. Levant

Although we have used KEDS to create several data sets, our most detailed work has dealt with the Middle East. The Middle East exhibits some of the most complicated political behavior in the world, with a variety of state and non-state actors vying for influence in the context of the ongoing Arab-Israeli conflict and, until 1990, US-Soviet competition. This extensive foreign policy activity presents a realistic challenge to any system of automated coding because of the quantity and variety of material, and it has been the benchmark that we used to develop the system.

Figure 3.1. Israel-Palestinian Cooperation and mediation

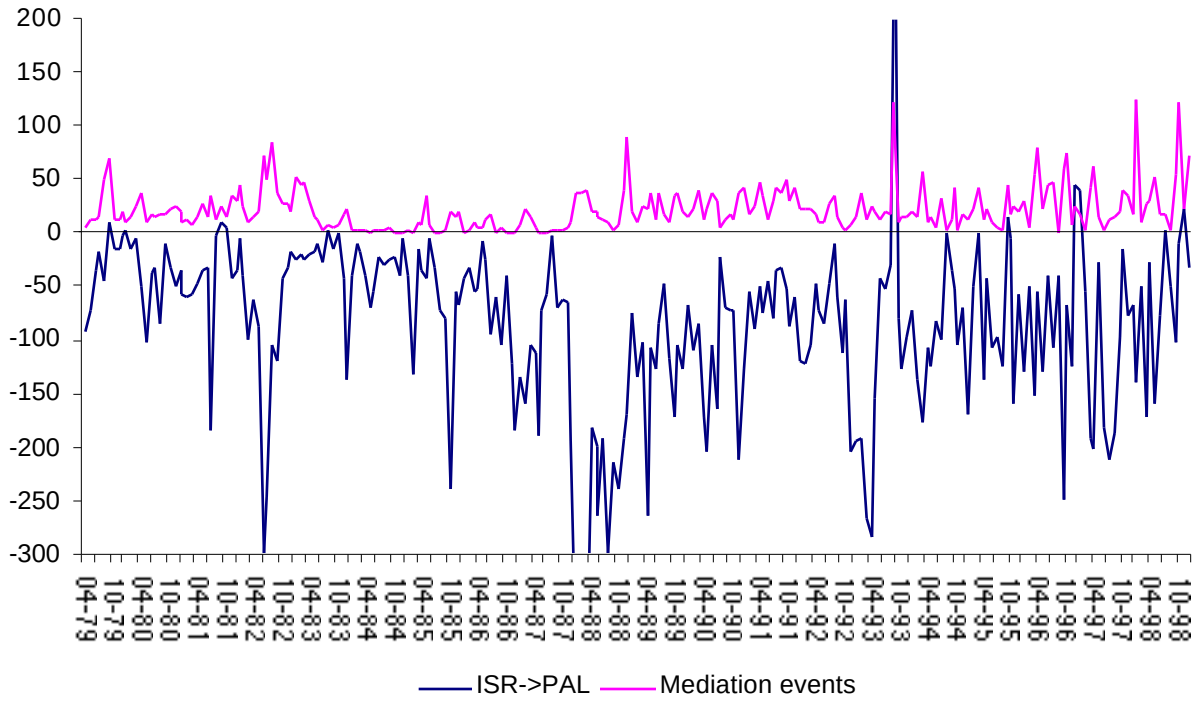
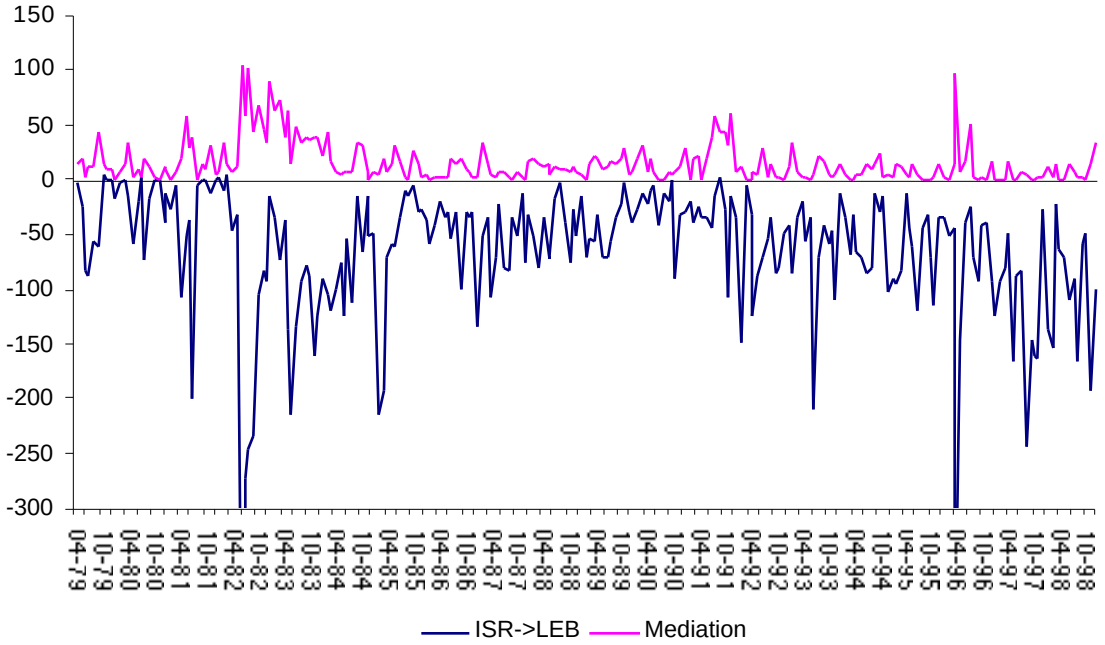


Figure 3.2. Israel-Lebanon Cooperation and mediation



An extensive discussion of the validity of this data, including a comparison with a human-coded data set, is found in Schrodt and Gerner (1994). For purposes of illustration, Figures 3.1 and 3.2 show both the Goldstein-scaled monthly aggregations of events of Israel to the Palestinians and Israel to Lebanon (these are the two most actively reported dyads in this region), as well as two new series of "mediation events" that we coded for our current research project on the effects of third-party mediation (Gerner and Schrodt 2001).

The Goldstein scale produces negative numbers of net conflict and positive numbers for net cooperation. In Figure 3.1 the major events of the Israeli-Palestinian conflict are conspicuous features of the series, for example Israel's 1982 operations in Lebanon against the PLO, the first Palestinian *intifada* of 1988-1991, and the negotiations of the Oslo process.[9] In addition to these major events, the series shows considerable detail in the ebbs and flows of conflict, such as the contrast between the relative lull in conflict in the period between Israel's withdrawal from most of Lebanon in 1985 until the outbreak of the first *intifada*, as contrasted to the erratic mix of conflict and cooperation that has followed the Oslo agreement. The Israel-Lebanon dyad in Figure 3.2 shows a similar level of detail, both during the military occupation of 1982-1985, and in the gradual escalation of violence in southern Lebanon starting around 1994.

The "mediation" measure is not Goldstein-scaled but instead is the frequency of a specific *pattern* of events: meetings (WEIS cue category 03) between a designated mediator (the USA, UN or EU) and both parties in the dyad that occurring within a period of seven days. As with the Goldstein-scaled time series, our measure of mediation activity tracks the historical record fairly well. The Israel-Palestinian dyad receives mediation efforts almost continuously except during the 1983-1988 period, with conspicuous spikes corresponding to events such as the 1982 invasion, US resumption of formal negotiations with the PLO in 1988 (see Gerner and Wilbur 2000), and various agreements in the Oslo process. In contrast, mediation in Lebanon tends—necessarily but not sufficiently—to coincide with periods of violence.

[9] The most negative—that is, most conflictual—months in the graph have been truncated so that the variation in the low-conflict periods will be more visible.

3.2. Persian Gulf

Schrodt and Gerner (1998a) report on an experiment where we coded stories from the Persian Gulf data using dictionaries that had not been customized. The purpose of this experiment was to determine how much detail one would get by simply downloading stories and coding them without additional (labor-intensive) dictionary development. We also did some systematic comparisons of lead-sentence and full-story coding in this region. The examples given in this section are based on that experiment, though we subsequently developed customized dictionaries for this region.

Figure 3.3 Goldstein-scaled series for Iran → Iraq, 1979-97

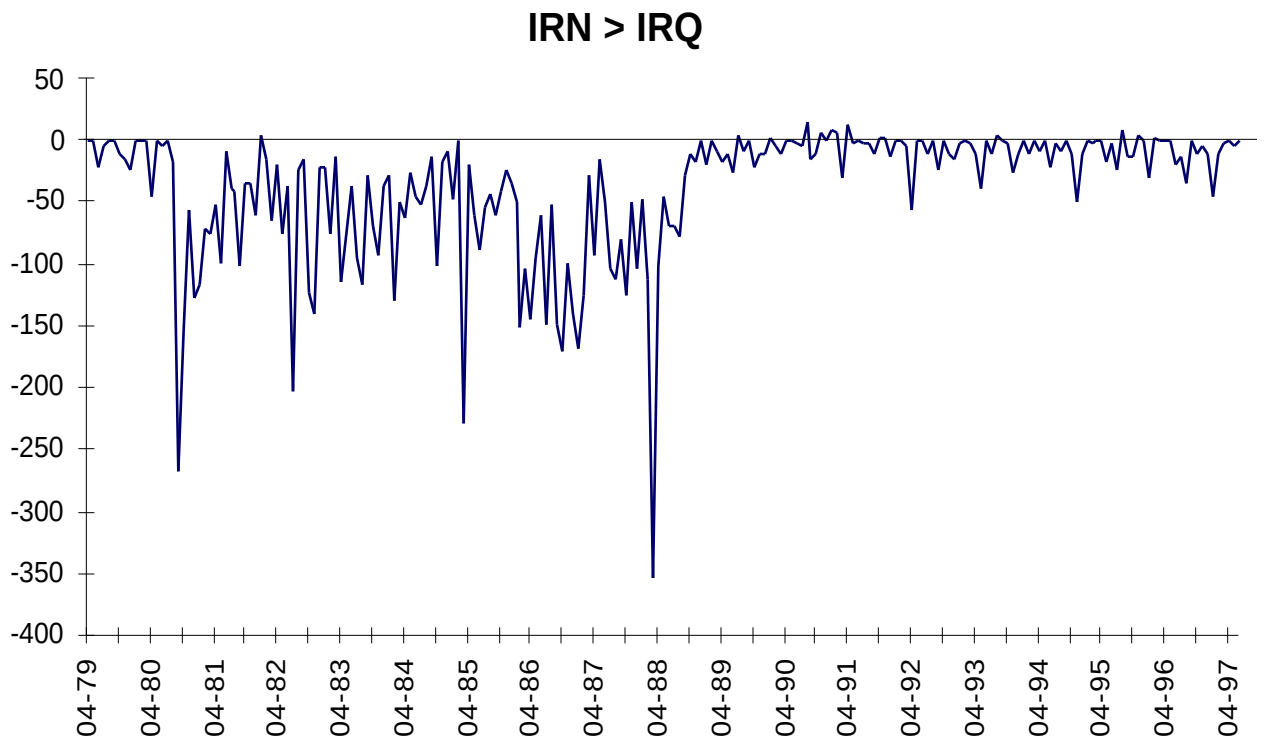


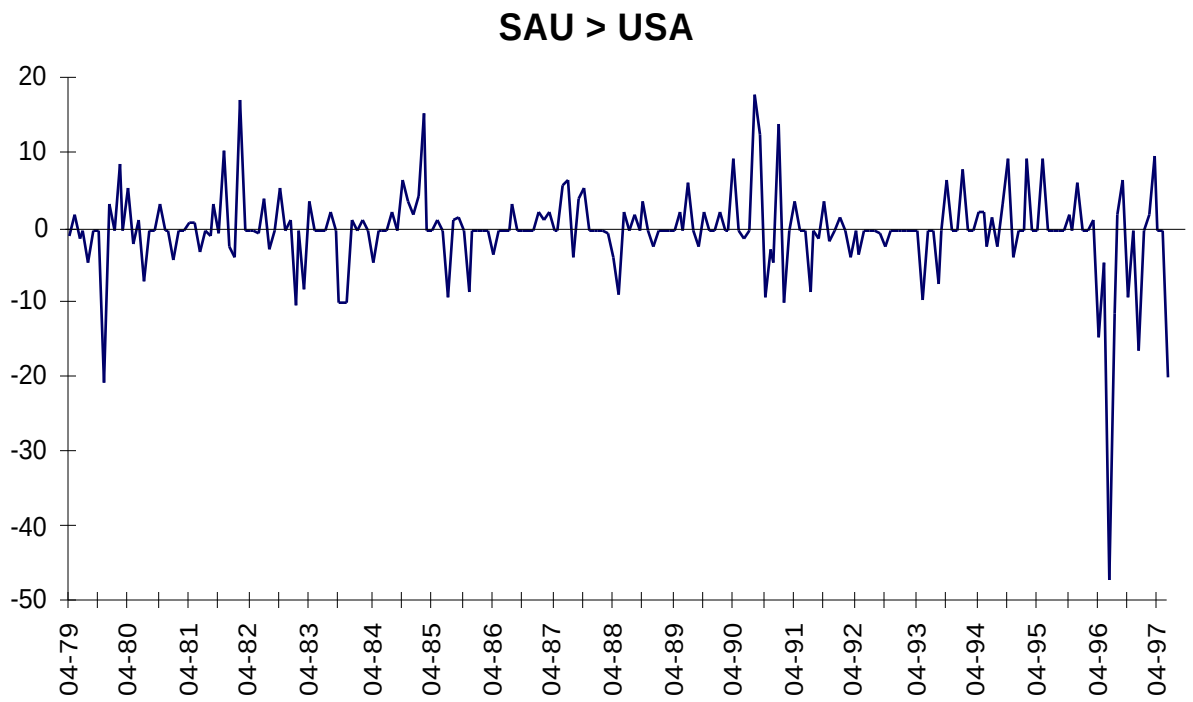
Figure 3.3 shows the events in this data set that were initiated by Iran and directed to Iraq, aggregated using Goldstein's scale.[10]. The major offensives of the Iran-Iraq war are clearly visible, as are the negotiations and overtures that Iraq made to Iran following Iraq's invasion of

[10] The Iraq to Iran sequence is similar: the two correlate with r = 0.84

Kuwait. The 1990s, as expected, are characterized by sporadic disagreements but only limited activity compared to the 1980s.

In contrast to the largely conflictual behavior in Figure 3.3, the graph of the Saudi Arabia to U.S.A. relationship in Figure 3.4 shows a sequence of ebbs and flows in the relations between the two states (also note that the vertical scale is about one-tenth that of Figure 3.3). An examination of the individual events shows that the positive values primarily correspond to meetings, agreements and statements of policy support. Negative values correspond to diplomatic disagreements between the two governments, for example on the controversy over the sale of AWACS aircraft in the early 1980s, and over U.S. policy towards Iran and Israel later in that decade.

Figure 3.4. Goldstein series for Saudi Arabia → U.S.A., 1979-97



The very large negative spike in 1996 is due to the Khobar Towers bombing in 1996. This is far the most part due to the fact that the data set did not differentiate sub-state actors, and therefore the bombing was coded as a Saudi attack on the U.S. However, additional legitimate negative events were generated by the diplomatic disagreements between the two governments

over the investigation of that incident. The only large anomalous negative value occurs in January and February 1991, where there are a number of incorrectly-coded uses of force. These come from reports about the Second Gulf War where the machine coding system failed to accurately determine the appropriate actors.[11]

Finally, Figure 3.5 shows the UAE → Kuwait series for both the full-story coding and the lead-sentence coding; the latter series has been displaced by -20 and rests on the bottom of the graph. In this dyad, where both of the actors receive relatively little attention from the international media, full-story coding produces a far more detailed set of events than lead-sentence coding produces (in fact lead-sentence coding shows almost no interactions at all in the dyad). However, we have found that in dyads that are closely the monitored by Reuters (for example Iran-Iraq), the full-story and lead-sentence series are highly correlated, and therefore lead-sentence coding gives adequate coverage.

As expected, most of the scaled events are positive, but there are a few instances of negative values despite the fact that Kuwait and the UAE are allies. These are due to three sources:

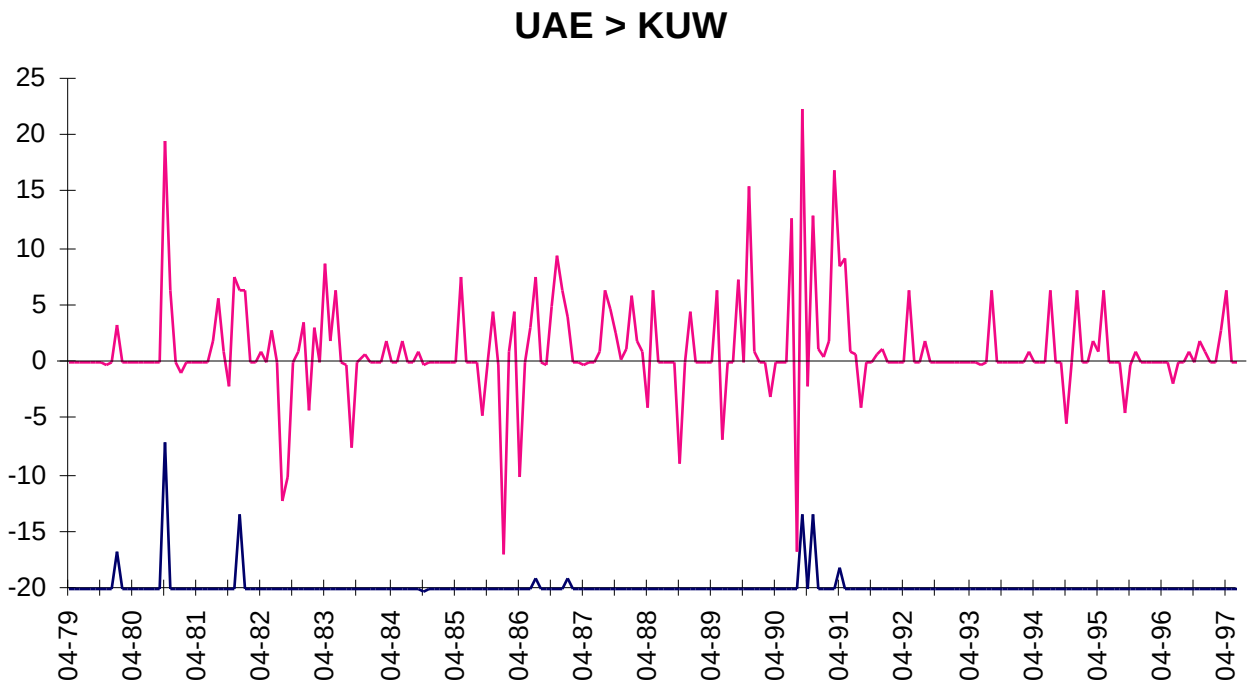
1. Disagreements over policy
2. Violent events due to criminal activity
3. Violent events that are due to coding errors.

The first category is legitimate—many of the cooperative events in the data set are the result of meetings, and sometimes meetings result in agreements, sometimes in disagreements. The criminal activity—usually a citizen of one state murdered or kidnapped in the other—is more problematic: this is not overtly political but it can have political implications. In the sequences we examined in detail, the violent events that are due to coding errors are relatively rare—the standard dictionaries deal successfully with most ambiguous phrases—but they still exist.

---

[11] The KEDS coding dictionaries, which were developed to code general international behavior, were prone to errors when coding stories dealing with military activities in the Second Gulf War. Descriptions of these events involved vocabulary and sentence constructions not encountered in other contexts, and sentences were sometimes misinterpreted.

Figure 3.5: Goldstein series for UAE → Kuwait, full-story (top) and lead-sentence events



Erroneously coded violent events (WEIS cue categories 20 through 22) can be problematic in aggregated data because they have high Goldstein scores. One solution to this problem would be to simply filter out such events in the computation of the Goldstein score for any dyad that was known, *a priori*, not to be engaged in military conflict. This would automatically remove the reports of deaths and seizures of persons and property that were due to criminal activity or incorrect coding; any remaining negative scores would be known to be caused by policy disagreements.

3.3. Former Yugoslavia

The dictionaries for coding events in the former Yugoslavia were originally generated by Jon Pevehouse for the data used in Goldstein & Pevehouse (1997) and Pevehouse & Goldstein (1999). We subsequently extended the data series in time and made some additional changes to

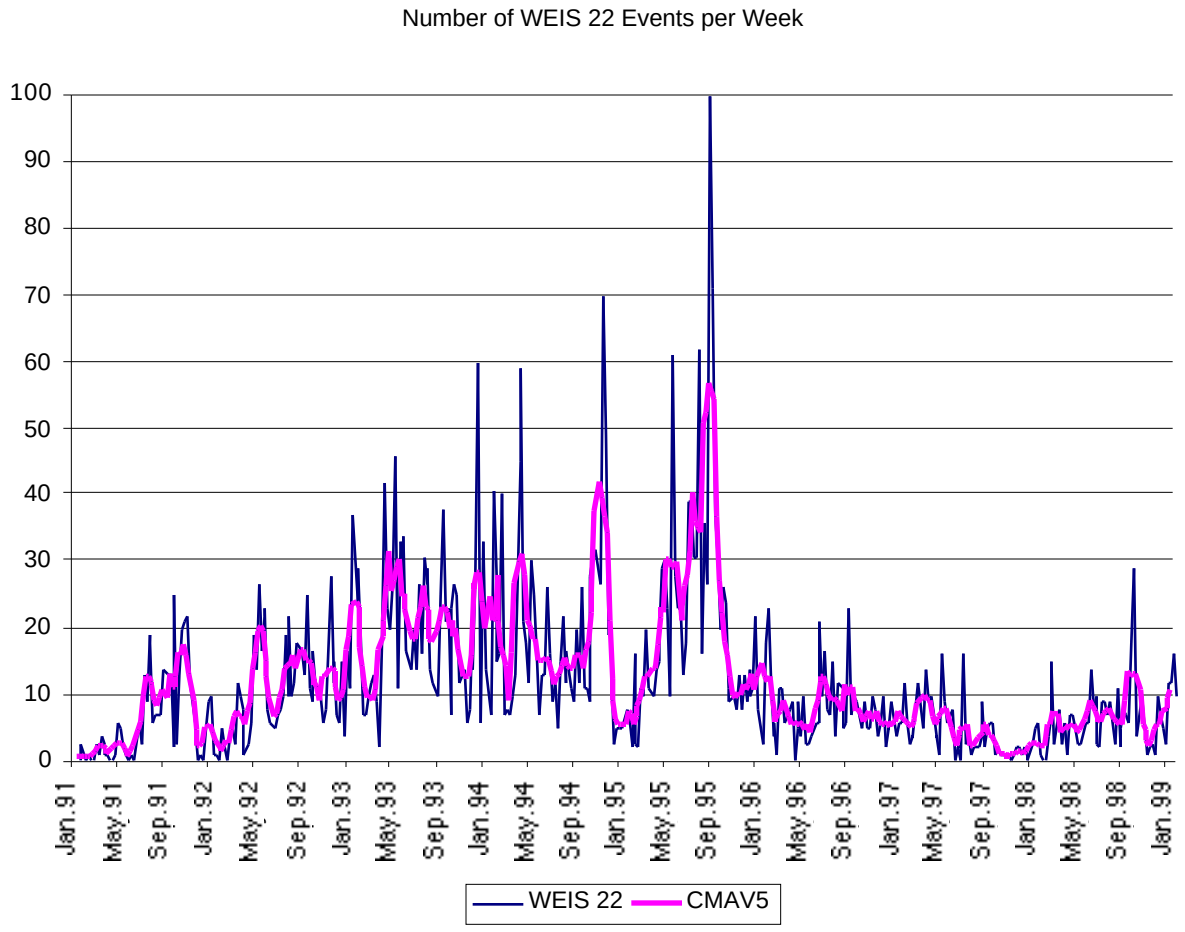
the dictionaries, particularly as Kosovo emerged as a focus in the late 1990s. Schrodt (2000a) uses these data to estimate hidden Markov models for predicting conflict in the region.

The case of the former Yugoslavia illustrates several of the strengths of the machine coding approach. First, by extending the Pevehouse dictionaries, we were able to make use of all of earlier efforts of Pevehouse with essentially no loss of information. Human coding projects, in contrast, can exchange codebooks and training protocols, but because most of the coding expertise lies in the "human capital" of the coders, there are very large start-up costs to any new project. Second, the Balkans involved a large number of sub-state and international actors, some of which require specification in excruciating detail (for example the dictionary includes phrases such as `SARAJEVO'S_WARRING_ MOSLEMS_AND_SERB [BFRMOS/BFRSER]`, `SERB_MINORITY_ LEADERS_IN_BOSNIA [BFRSER]` and `NATO'S_BOSNIA_PEACE_FORCE [NAT]`). Finally, because we were developing the data set for an on-going conflict, we were continually adding new actors as these became conspicuous in the Reuters reports. Because recoding the entire data set can be done quickly, it was a simple matter to go back and code any previous occurrences of these actors.

Figure 3.6 shows the total number of WEIS "Force" events (cue category 22) in the region by week from January 1991 to January 1999; the heavy line (magenta in the color version) shows a five-week centered moving average of these totals. As expected, the weekly series is very erratic—the conflict itself went through rapid increases and decreases in intensity—but the moving average reflects the various phases, with a peak of activity in the summer of 1995 that corresponded to the combined activities of the Bosnian-Croatian counter-offensive against Serbian forces, and the brief NATO military intervention against Serbia.

Figure 3.7 shows Goldstein-scaled net cooperation from Serbia to Bosnia; in this graph Bosnia Serbs are included in the "Serbian" activity, although in the original event data they are assigned a distinct code.[12] The major periods of conflict are evident, as is the period of the Dayton agreement. The dyadic summary correctly does *not* show conflict during the periods when the major source of regional military activity was between Serbia and Croatia (1991) or Serbia and Kosovo (1998).

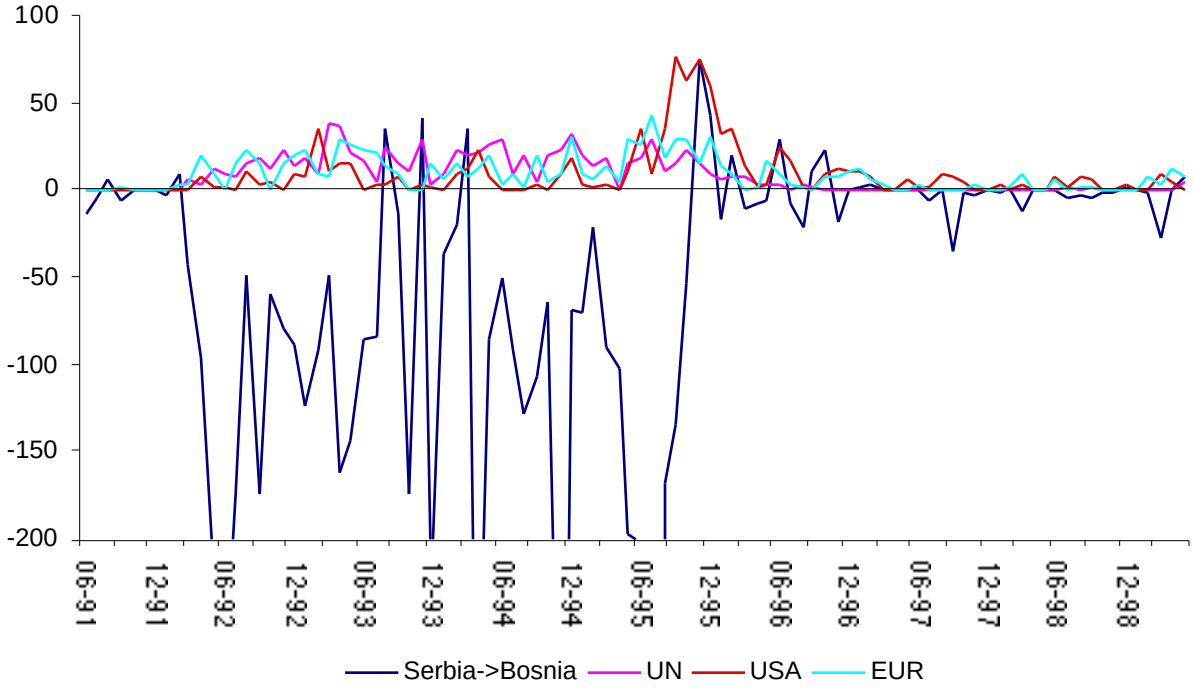
Figure 3.6. Number of WEIS "Force" events per week in the former Yugoslavia; CMAV 5 shows a five-week centered moving average



The lines labeled "UN", "USA" and "EUR" count the number of mediation events involved the United Nations, United States and major European state (plus the EU) respectively. As in Figures 3.1 and 3.2, "mediation" is the pattern of the mediator meeting with both parties in the dyad within a period of seven days. As we demonstrate in Gerner and Schrodt (2001), even this relatively simple measure of mediation is sufficient to show distinctions in cross-correlation tests between the effects of the three mediating groups: UN efforts were associated with subsequent increased levels of conflict; United States efforts were associated with decreased levels; and there was no discernible change following European efforts.

[12] As in Figure 3.1, the values of the largest negative months have been truncated.

Figure 3.7. Serbia-Bosnia Goldstein values and mediation



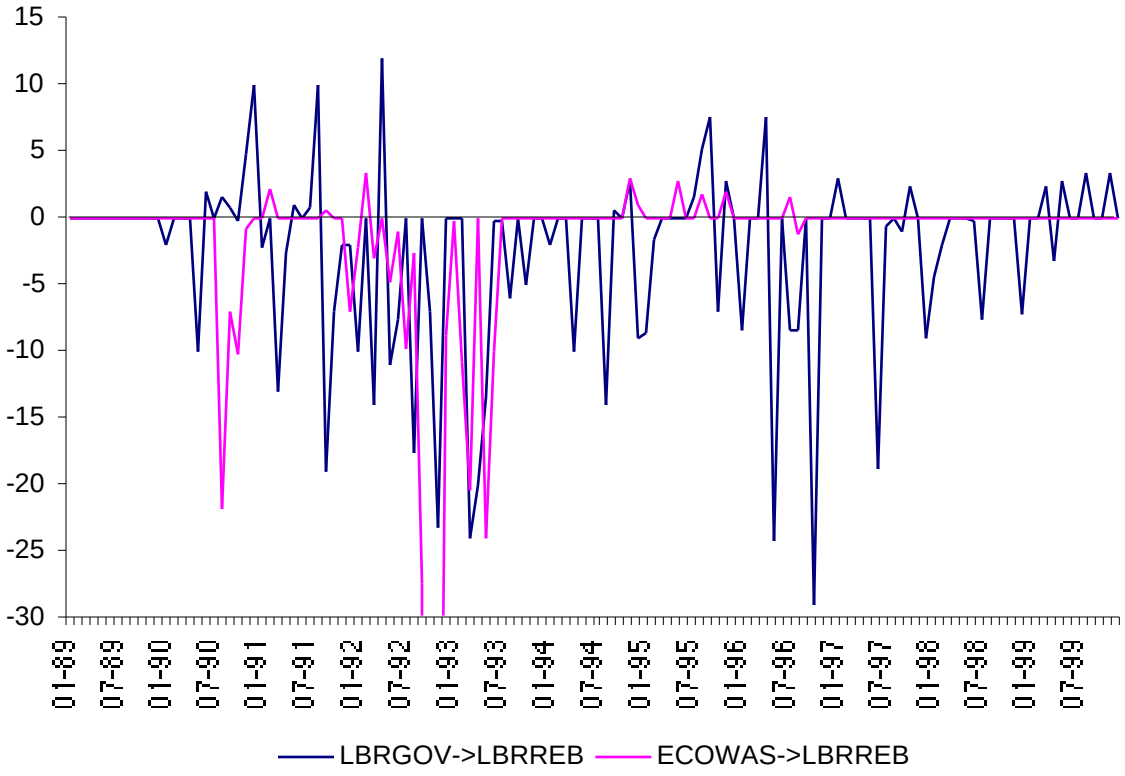
3.4. West Africa

Huxtable's 1997 dissertation research demonstrated that news reports on West Africa, while difficult to code and subject unreliable and inconsistent coverage, would nonetheless produce event data that could yield meaningful analytical results. We have continued to do some work with this region[13]—we are currently monitoring all of the littoral states from Mauritania to Cameroon, plus Burkina Faso, Mali and Niger—and we may use some of it in our mediation research. In contrast to Huxtable, who used multiple news sources (Reuters, *Agence France Presse*, Associated Press and BBC), we are currently coding only from Reuters. Like Huxtable, we used full-story, rather than lead-sentence coding, because Huxtable found this to be essential for the region.

Figure 3.8 shows the Goldstein-scaled series for the civil wars in Liberia. The Liberia series shows a fair amount of detail, both with respect to the internal conflict and the ECOWAS

intervention in the conflict. (In this series, "ECOWAS" includes activities involving ECOWAS explicitly, but also those of Nigeria and Ghana, the primary contributors to that force, and activities Reuters attributed to "West African peacekeeping forces"). Nonetheless, the Goldstein values are roughly one-tenth those found in the former Yugoslavia in Figure 3.7, despite the fact that during much of the period, the level of violence was probably comparable or worse than the levels in the Balkans. Since the Goldstein scaled value is roughly proportional to the total number of reports, this is indicative of the lower level of coverage in West Africa compared to the Balkans.

Figure 3.8. Goldstein series for Liberian government and ECOWAS actions toward Liberian rebels, 1989-99

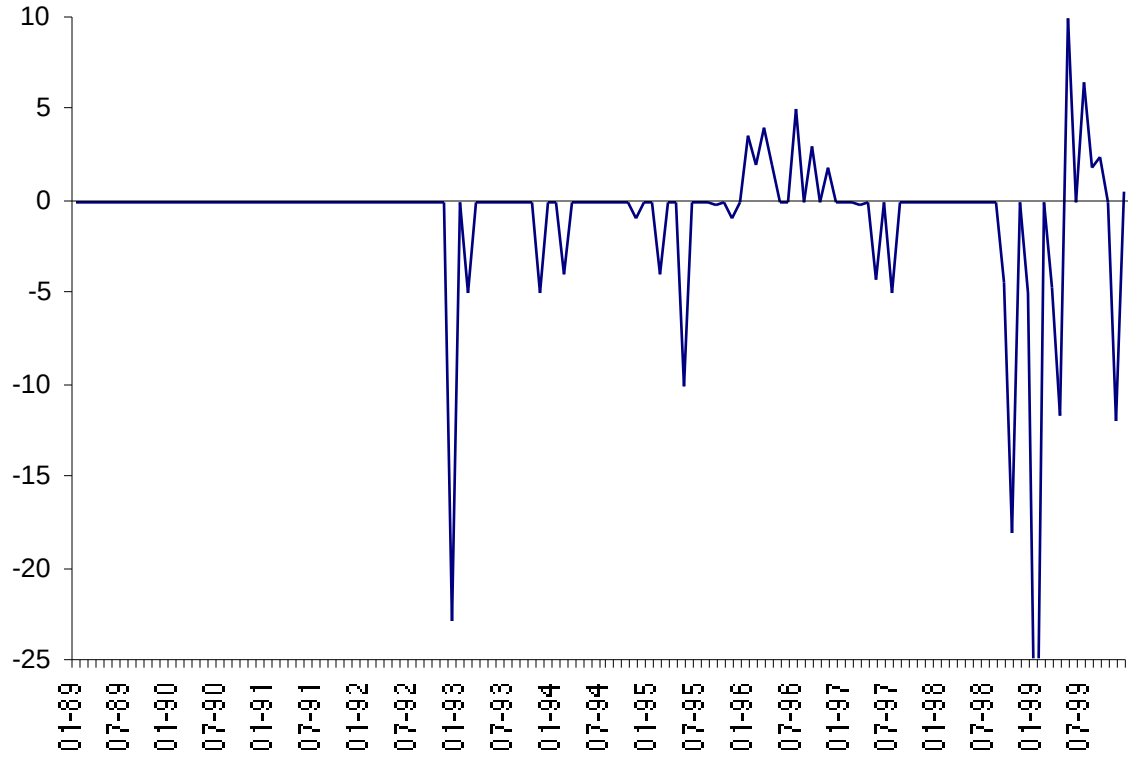


The coverage of Sierra Leone is even worse, as illustrated in Figure 3.9. The data show some indications of the major outbreaks of conflict in the early 1990s, and the regionally-mediated

[13] This data set is currently not on the web site because it contains a one-month "hole"—November 1995—due to an error in the source from which we obtained the data. We have subsequently obtained the text required to fill

peace talks and elections in 1995 and 1996, but very little else until the late 1990s. The increase in reported events in the late 1990s are, in all likelihood, due primarily to the increased media attention to Sierra Leone that resulted from well-publicized reports of atrocities against civilians and the emerging issue of "blood diamonds" fueling the war. In the late 1990s Liberia and the "Great Lakes" region became relatively quiet and the limited journalistic attention devoted to Africa may have shifted to Sierra Leone. But overall the event data are very sparse, and the situation looks particularly hopeless with respect to finding sufficient information for early warning.

Figure 3.9. Goldstein series for Sierra Leone government actions toward Sierra Leone rebels, 1989-9992



Several factors combine to make West Africa difficult. First, as noted above, we are coding only Reuters, and Huxtable makes a convincing case that in West Africa the use of multiple sources is essential. Second, West Africa is generally a very marginal area for the international

this in, but have yet to get around to recoding it. The data are, nonetheless, available from us via email.

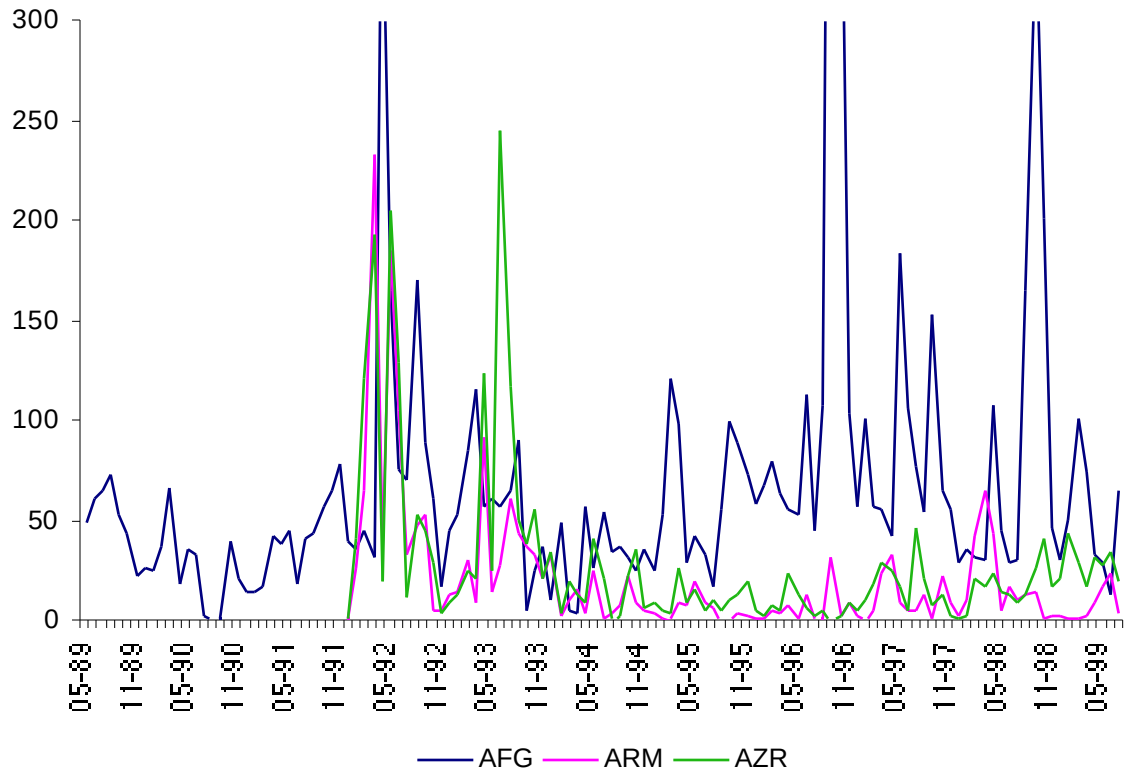
media, both because it is relatively poor and because reporting is very difficult in unstable areas. (This stands in contrast to the Levant and Balkans where most of the major conflict areas, while at times dangerous, could usually be reached from open international airports in a couple of hours.) Third, it is more difficult to code "sides" in these conflicts, since much of the violence comes from very loosely organized guerrilla groups and warlords rather than the well-defined political organizations and states in the Levant, and the ethically-defined antagonists in the former Yugoslavia. Some of these problems may be due to insufficient dictionary development—we have spent substantially less time on this region than on the Levant and Balkans cases—but other problems are inherent in the reports, which are often vague as to the identity of the individuals involved in activities.

3.5. Central Asia

Our final case is Central Asia, where we have coded events from texts downloaded from Reuters Business Briefing on Afghanistan, Armenia, Azerbaijan, and the former Soviet republics of Central Asia. In the case of Afghanistan, we have coded the most conspicuous of the militia groups and political factions (for example the Taliban), and in all of the countries we have separate codes for opposition groups and the military. Figures 3.10 and 3.11 show the monthly totals for the frequency of events towards any other actor (rather than Goldstein-scaled dyadic interactions) reported for most of the states in this data set.

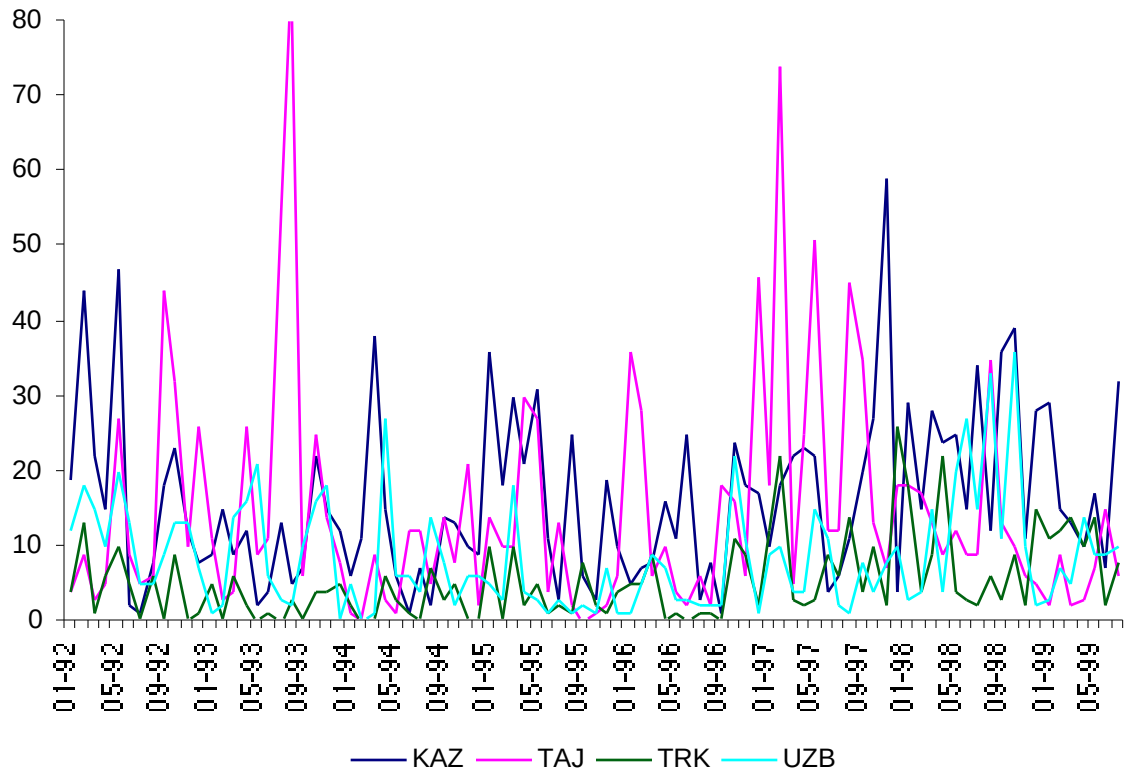
As shown in Figure 3.10, Afghanistan is the most thoroughly covered state in the region, with a continuous and relatively high number of events (particularly compared to West Africa). It is not exactly clear *why* Afghanistan gets this coverage, since the conflict is at least as isolated, and the area at least as economically depressed as anything in Liberia and Sierra Leone, but we speculate that Afghanistan got onto the international agenda during its war against the Soviet occupation and has therefore continued to maintain the attention of the media. Armenia and Azerbijan, in contrast, get considerable coverage during the period of warfare in 1991-1993, but subsequent coverage has been very scant. Coverage of Azerbijan picks up a bit in the late 1990s, but we suspect this is largely due to interest in Caspian Sea oil.

Figure 3.10. Number of events generated per month for Afghanistan, Armenia and Azerbaijan, 1989-1999



Somewhat to our surprise, there has been a fairly consistent level of coverage in the former Soviet republics, as shown in Figure 3.11. Tajikistan has received the most attention, probably due to an on-going civil war between the government and rebel forces supported by Russia. (We were also specifically looking for this conflict, and the apparent signal here may be due in part to more thorough dictionary development.) Turkmenistan has the lowest level of coverage, dropping to zero events in a number of months, and at a dyadic level Turkmenistan might look similar to the graph for Sierra Leone. As with the West African case, this coverage could probably be substantially improved if multiple news sources were coded.

Figure 3.11. Number of events generated per month for Kazakhstan, Tajikistan, Turkmenistan and Uzbekistan, 1992-1999



4. Comparing Reuters and *Agence France Presse*

The density of data available from Reuters has made it appealing to researchers, but, as will be discussed below, it is far from a perfect source. The "Reuters Question" is part of a broader debate within the event data community as to the ability of any source to accurately and reliably reflect an underlying reality. More practically, Reuters is no longer available on the data service most widely available to academic researchers, LEXIS-NEXIS. Unable to create new datasets or continue older ones, researchers are looking for alternative sources. We have begun exploring the possibilities of *Agence France Presse* (AFP) as an alternative source of event data . AFP is still available on NEXIS, and is the second-largest news gathering organization in the world after Reuters.

The second generation of event data projects that began in the 1990s—KEDS, PANDA and GEDS—initially viewed Reuters as an attractive alternative to the newspaper sources used by

WEIS, COPDAB and other first-generation projects. Reuters had far more stories each day, and while there was still an editorial process selecting these stories, the news service was less likely to be affected by the space limitation and the readership interests of the individual newspapers. Nonetheless, the news wires have their own limitations. Huxtable (1997) found that in West Africa, Reuters, AFP, and the BBC were much more supplementary than complimentary; some events were only covered by one source and not the others. While Huxtable had expected to find a systematic bias based on language and past colonial links— BBC and Reuters would provide better coverage of Anglophone countries such as Ghana and Nigeria, while AFP would be superior in Francophone countries such as Senegal and Côte d'Ivoire—this did not prove to be the case.

Similar findings were reported by Gerner et al. (1994) when comparing Reuters coverage to that of the *Journal of Palestine Studies* and the German-language periodical *Infomationen.*[14] Here again, despite seeming heightened regional sensitivity, Reuters captured some events that the regional sources did not. The regional sources did however focus on different *types* of events. For example, *Informationen* emphasized events in East Germany and relations with the Warsaw Pact more heavily than Reuters, while *Journal of Palestine Studies* reported many more events concerning internal Israeli politics and acts against Palestinians. In Gerner et al. (1994) we concluded, nonetheless, that the density of data from Reuters compensated for (or at least outweighed) these disparities.

Sommer and Scarrit (1998) report on comparisons between Reuters and regional news sources, but here the emphasis is in Africa, in particular Zimbabwe and Zambia. They compare the daily newspapers *Africa Research Bulletin* and *Zimbabwe Herald* over an eleven year period. They also conclude that Reuters and the regional sources supplement each other, covering events that others miss. They are not, however, so willing to dismiss the regional sources, particularly given

---

[14] *Infomationen* was machine-coded using an earlier version of KEDS that had been modified to code German-language sources. While we have not subsequently pursued the development of multiple-language coding, our sense is that this would be relatively easy within the sparse-parsing framework of KEDS/TABARI. This would be particularly easy for Spanish and French, which have the same subject-verb-object ordering as English, and where verb roots can generally be determined by removing tense-indicating suffixes. German and Russian—which are inflected languages where noun and verb forms differ depending on whether a word is in the subject or object—would be more difficult with the existing KEDS/TABARI program, but researcher in the field of

their differing emphasis on cooperation between Zimbabwe and Zambia. Huxtable and Pevehouse (1997) have noted the coverage disparities for Reuters between developed and developing countries; Sommer and Scarrit further confirm this conclusion. It may be that in situations characterized by a low density of events, regional sources are more important than those receiving continuous coverage.

4.1. Design

In this analysis, we used TABARI to code lead sentences from Reuters and AFP stories covering the Balkans and the Levant during 1993, 1995, and 1997. New AFP stories were downloaded from NEXIS and compared to events coded earlier from Reuters stories. A major concern when generating the AFP stories was related to the search strings used to generate the original Reuters stories. When Reuters was available on NEXIS, stories were selected using a Boolean search command. However, Reuters stories after 10 June 1997 were downloaded from the Reuters Business Briefing service, which selects stories by country name based on indexing of the stories that is done by the Reuters service; it does not have a Boolean search capability on the story itself. The potential implications of this difference will be discussed below.

We have written a Perl program—nexispider— that automates most of the process of downloading and formatting stories from the NEXIS "Academic Universe" service that is available in most North American university libraries; this program is available on our web site.[15] After entering a time interval that generates less than 1,000 articles—the limit of a single search in the Academic Universe service—one transfers the URL of the first article to the program, which then downloads the full text of each article and puts it into a KEDS/TABARI input format. An additional Perl program called nexisreverse puts the stories in chronological order.

The leads were coded using specialized regional dictionaries. However, all of these were developed and updated using Reuters sentences, so idioms and vocabulary used by AFP may not be reflected in them. The resulting events were aggregated using the Goldstein (1992) scale. For this analysis we looked at the following dyads:

---

computerized content analysis have done extensive work with these languages, and there is no reason to assume that they could not be coded.

[15] By hacker conventions, a "spider" is a computer program that automatically checks a number of pages on the World Wide Web by following the links embedded in each web page. "spider"…"web"…cute, eh?)

Balkans	Levant
SER>USA	ISR>USA
USA>SER	USA>ISR
SER>UNO	PAL>USA
UNO>SER	USA>PAL
BFRSER>BFRMOS	ISR>PAL
BFRMOS>BFRSER	PAL>ISR

We totaled the number of events per dyad for each month during 1993, 1995, and 1997, and tallied the total number of events per year per WEIS two-digit cue category.[16] These calculations were done over the entire series as well as for each individual year.

4.1. Results

The results of the comparison generally differed substantially between the Balkans and the Levant in terms of the event counts and the Goldstein-scaled series, but not in the distribution of the types of events. The total number of coded events generated by Reuters and AFP are quite comparable in the Levant, with only 1997 showing a disparity where AFP has more events. In contrast, the number of events reported by AFP in the Balkans is substantially lower than by Reuters. However, there does not seem to be a systematic or consistent bias to the coverage. While Reuters coverage is more dense in the Balkans, it also fluctuates across years and dyads; there are not simply twice or three-times as many Reuters' events as AFP. This is reflected in the low correlations for raw event counts by dyad. Similar results characterize the Levant despite the greater comparability in gross totals. Table 4.1 indicates the correlation results.

Table 4.1: Correlation for AFP and Reuters of total event count by dyad

Dyad	**SER>UNO**	**UNO>SER**	**SER>USA**	**USA>SER**	**BFRSER>BFRMOS**	**BFRMOS>BFRSER**
correlation	0.2	0.09	0.45*	0.28*	0.06	0.13
Dyad	ISR>PAL	**PAL>ISR**	**USA>ISR**	**ISR>USA**	**USA>PAL**	**PAL>USA**
correlation	0.06	0.36*	0.38*	0.35*	0.37*	0.15
* indicates significance at 0.05 level						

[16] We use the following abbreviations for actors: SER = Serbia; USA = United States; UNO = United Nations; BFRSER = Bosnian Serbs; BFRMOS = Bosnia Moslems; ISR = Israel; PAL = Palestinians

This seems to suggest that even where event density is similar—as in the Levant—there are differences in monthly-by-month coverage as well as by dyad.

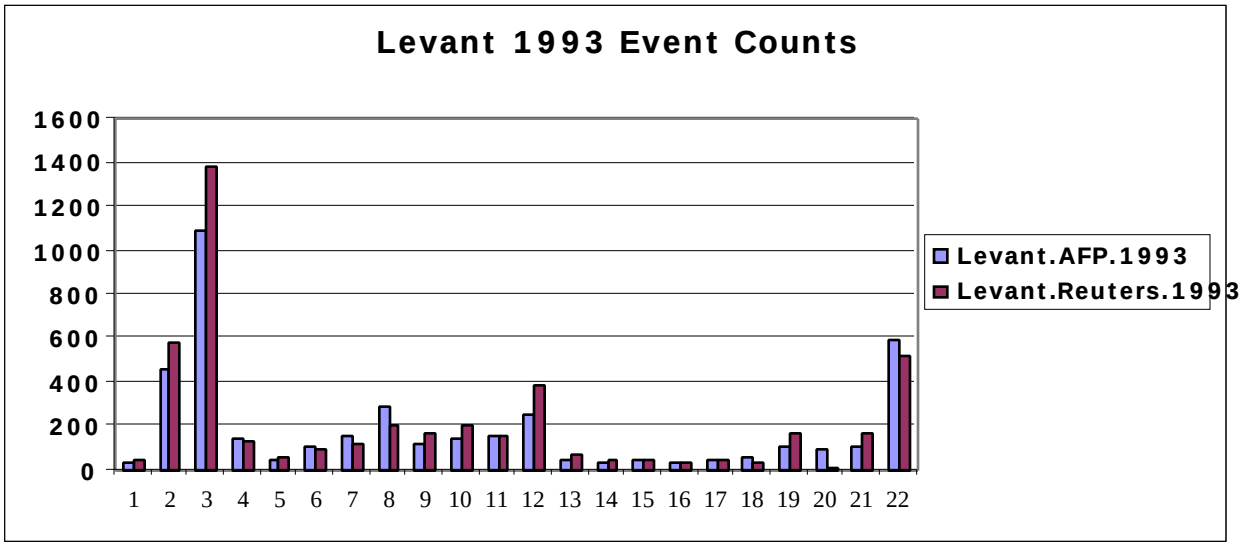
These results do not speak to whether similar *types* of events were being coded. As a first cut as to whether Reuters and AFP were covering the same types of events despite the differences in overall coverage, we tallied the total event count by WEIS code for each year. Figures 4.1 and 4.2 illustrate the results for the Balkans and the Levant, respectively. As would be expected, Reuters outpaces AFP in every WEIS category in the Balkans. This contrasts sharply with the Levant, where AFP matches Reuters in nearly every 2-digit category and outpaces it across several categories in 1997. Despite these differences in raw event totals, the two wire services seem to report generally similar information (at least in terms of WEIS categories). Table 4.2 shows the Pearson correlation between the Reuters and AFP totals for each WEIS category, by year. All values are significant at the 0.05 level. These results seem to indicate that while AFP and Reuters may cover events with different intensity, they are generally focused on similar types of events as indicated by WEIS categories.

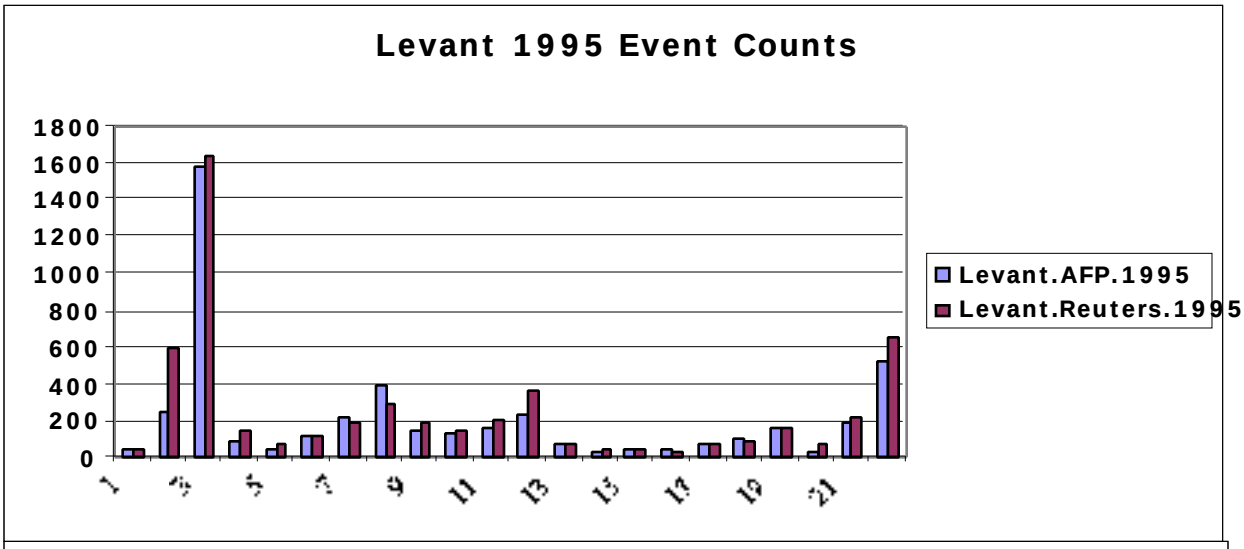
Table 4.2: Correlation for AFP and Reuters for counts in WEIS categories

	Balkans.1993	Balkans.1995	Balkans.1997	Levant.1993	Levant.1995	Levant.1997
correlation	0.95	0.93	0.91	0.98	0.97	0.96

Broadly speaking these results are confirmed by the net-cooperation scores reported in the above dyads by AFP and Reuters. When AFP and Reuters capture the same set of events (which is not always the case as indicated in Table 4.1), they tend to generate similar net-cooperation scores. However, because the Goldstein (1992) scale is cumulative, five stories about conflict will generate a lower-magnitude score than 50 stories. This effect is particularly strong in the Balkans where Reuters outpaces AFP in nearly all dyads. The primary discrepancy in the graphs is reflected when one wire service picks up an event (or series of events) while the other does not. This reflected in the low correlations between net-cooperation scores by dyad between AFP and Reuters (Table 4.3).

Figure 4.1. AFP and Reuters Event Counts by Cue Category for Levant





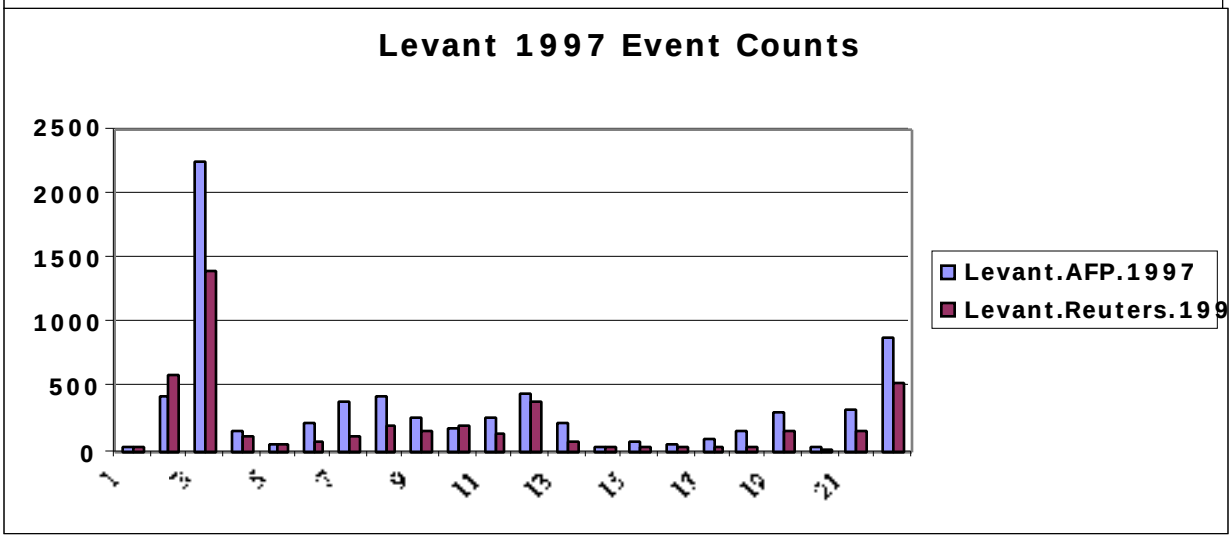
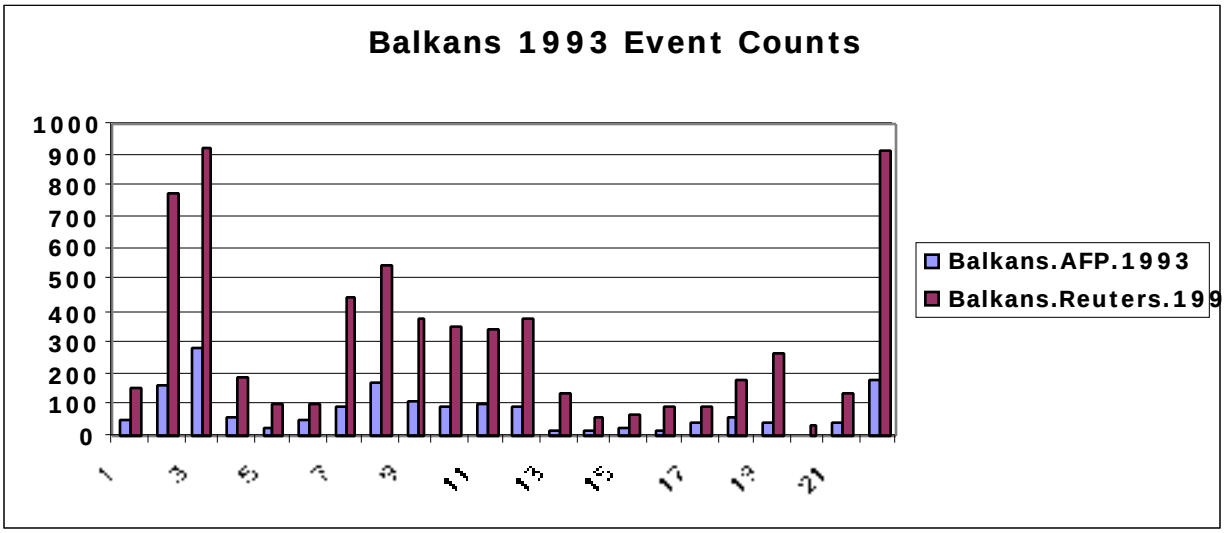
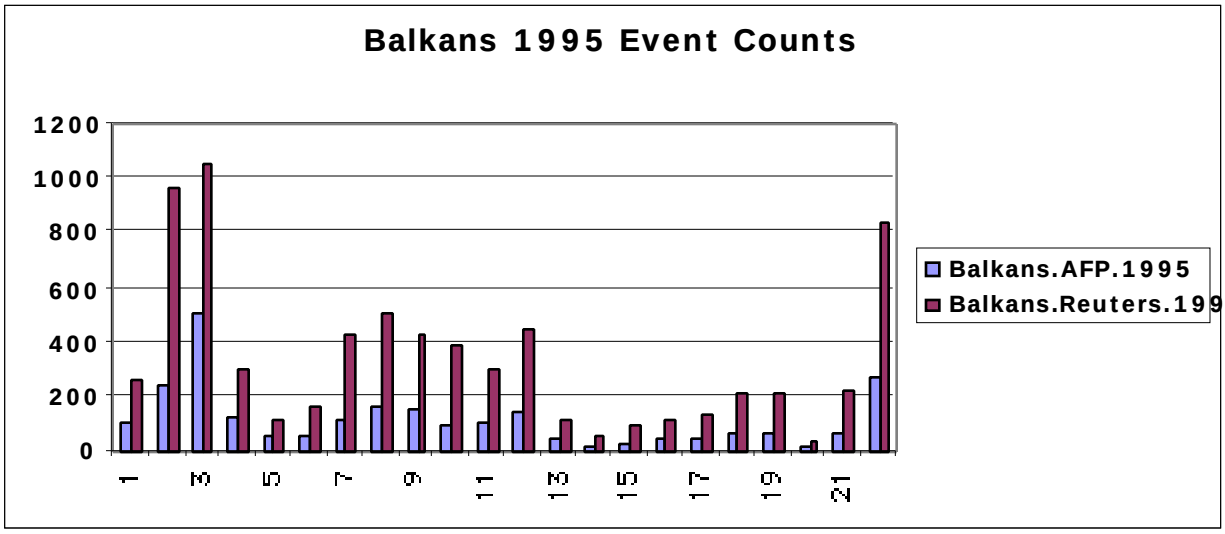


Figure 4.2. AFP and Reuters Event Counts by Cue Category for Balkans





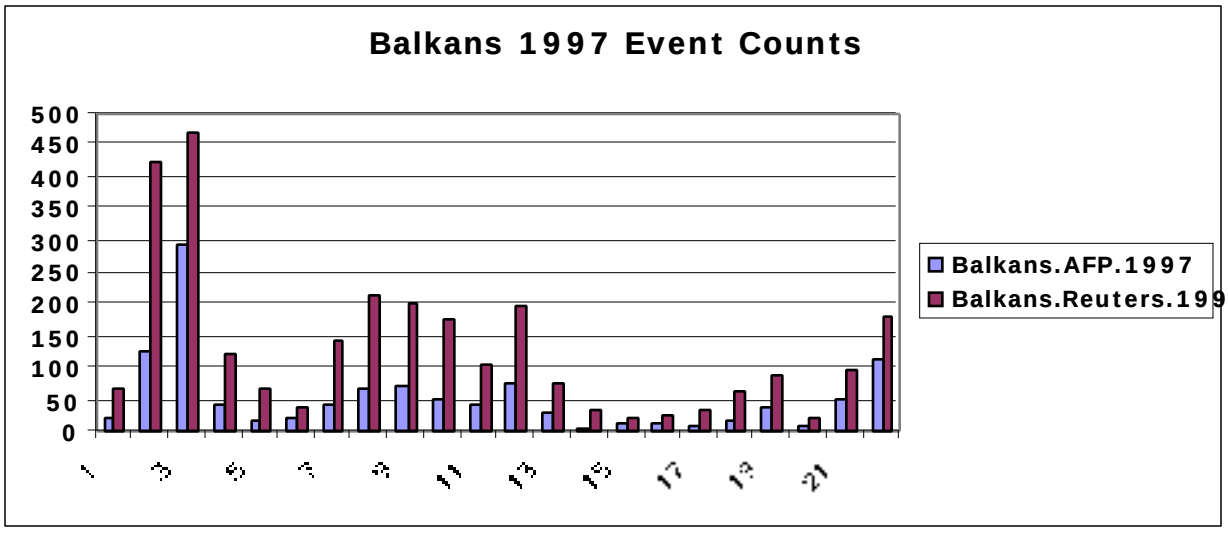


Table 4.3: Correlations for net-cooperation scores

Dyad	SER>UNO	UNO>SER	SER>USA	USA>SER	BFRSER>BFRMOS	BFRMOS>BFRSER
1993	-0.51	-0.25	n/a	0.22	n/a	n/a
1995	0.27	0.25	0.4	0.28	0.27	n/a
1997	-0.09	0.13	-0.12	0.47	-0.17	n/a
series	0.16	0.23	0.03	0.29*	-0.08	-0.09
Dyad	**ISR>PAL**	**PAL>ISR**	**ISR>USA**	**USA>ISR**	**PAL>USA**	**USA>PAL**
1993	-0.32	0.31	0.34	0.76*	-0.21	-0.33
1995	0.28	0.22	-0.2	0.18	-0.08	-0.15
1997	0.16	0.29	-0.09	0.61*	0.67*	0.56
series	-0.05	0.17	0.13	0.55*	0.22	0.39
* indicates significance at 0.05 level						

These low correlations and the problem of non-events confirm the findings of Huxtable and Pevehouse (1997) and others who note that most sources are complimentary, not supplementary. In both the Levant and the Balkans, AFP and Reuters pick up different stories. However, the correlation between the total event-count across WEIS categories suggests that they generally pick up similar types of stories in the region as a whole. Perhaps there are differences in geographic location or month-to-month emphases that create disparities between the Balkans and Levant coverage.

The low coverage of AFP in the Balkans is particularly troubling. There are two possible conclusions to be drawn from this phenomena. Either AFP did not cover the Balkans very intensely, or our Boolean search did not yield all the available stories because the search terms were not found in the headline. NEXIS Boolean searches for "NATO" during August 1995 (when NATO strikes against Serbs began) revealed several stories that were ostensibly about events in the Balkans but contained headlines such as "NATO troops deployed today." Subsequent sampled searches revealed a much higher yield when using the "Headline and Lead Paragraph" in NEXIS as opposed to the headline alone.

We experimented with an alternative download of stories from 1995—a period of high and varied activity in the Balkans—to sample the possible effects of these changes. The changes in the search strategy almost doubled the total number of both articles and events. The new AFP event totals are consistently higher than those from the original download. There are still inconsistencies in the BFRSER-BFRMOS dyads but the correlation for total event count (across all three years) by dyad improved significantly when the new data from 1995 was included. Table 4.4 reports these results.

Table 4.4: Comparison of Pearson Correlations for Total Event Count by Dyad

Dyad	SER>UNO	UNO>SER	SER>USA	USA>SER	BFRSER>BFRMOS	BFRMOS>BFRSER
Balkans.Orignal	0.2	0.09	0.45*	0.28*	0.06	0.13
Balkans.Expanded	0.29*	0.38*	0.47*	0.24	0.13	0.02
* indicates significance at 0.05 level						

In four of the six dyads, the correlations improved in the expanded 1995 data. Though the lower correlation for the USA>SER dyad is interesting, the explanation for this is unclear. The new correlation is, however, nearly statistically significant ($r > .275$ is significant at the 0.05 level).

These findings are confirmed by including the new data from 1995 in the net-cooperation time-series. Although still not generating scores as large as Reuters, the new set of AFP leads do outpace the original AFP leads in several months as well as picking up events in dyads (such as the BFRSER-BFRMOS dyad) that were previously ignored. Table 4.5 compares the Pearson correlations for all dyads in 1995.

Table 4.5: Comparison of Pearson Correlations for Net-Cooperation by Dyad

Dyad	SER>UNO	UNO>SER	SER>USA	USA>SER	BFRSER>BFRMOS	BFRMOS>BFRSER
Balkans.Original	0.27	0.25	0.4	0.28	0.27	n/a
Balkans.Expanded	0.29	0.28	0.4	0.68*	0.43	-0.015
* indicates significance at the 0.05 level						

The correlations improve in five dyads and remain the same in one (SER>USA). Although only one correlation is statistically significant (USA > SER), the results suggest that re-running all the

data with based on stories using the second search string would improve the comparability between Reuters and AFP.

4.2 Discussion

Reuters and AFP are comparable in term of the general patterns of events they report. They are not, however, identical sources of information. Confirming previous analyses, Reuters provides denser coverage in the Balkans, even after changing the AFP search strategy for 1995. The 1995 Balkans sample using the "Headline and First Paragraph" search option suggests that future research should employ similar techniques in order to capture all of the relevant stories. In the case of Reuters, relevant information is contained in both the headline and lead sentence of the story; apparently in AFP it is not. However, a similar problem did not occur in the Levant. This could be due to something as simple and idiosyncratic as a different writer and copy editor in the region.

What seems to be important here is not only that AFP differs in style from Reuters, but that there are regional differences in AFP as well. This suggests that sometimes Reuters is in the right place at the right time, and sometimes AFP is. This is likely particularly true when dealing with events "on the ground" as opposed to formal diplomatic communications. This may explain the low correlations in the UNO-SER and BFRSER-BFRMOS dyads where most interactions had to be witnessed instead of reported at a news conference.

Similar to earlier work, we did not find that AFP has one story while Reuters has five about the same event. They "see" different things. But, given the high correlation between the number of events in the individual WEIS categories, there do not appear to be systematic biases toward certain types of events for each wire service. While Reuters often had more total events per category, there was a high correlation between AFP and Reuters counts across categories.

(One reason for the lower event counts for AFP may be a result of the dictionaries being developed using Reuters and therefore not reflecting the idiosyncratic references made by AFP. This may be more problematic in the Balkans, particularly in the early years as journalists groped for coherent labels for the warring parties. This seems to have been less of a problem in the Levant where event totals were high to begin with.)

This analysis does not indict either Reuters or AFP. Previous studies had given weight to Reuters because of its density of events despite clear indication that it missed events that other wire services like AFP or the BBC picked up (Huxtable and Pevehouse 1997). In the Levant, AFP matches the density of events making it a good match for Reuters. Although the total number of stories from AFP in the Balkans is less than those from Reuters, the high correlation between the totals for the WEIS categories seems to indicate that while the two services pick up different events, they do not focus systematically on different *types* of events.

In an earlier series of tests that we did to evaluate the effect of alternative scales in a time-series cluster analysis method (Gerner and Schrodt 1998), we found that almost 50% of the variance in our results was due to the event counts themselves, rather than in the classification or weighting of events. This suggests that the important differences between sources may lie primarily in whether a report of an event was generated at all. However, Reuters and AFP were much more alike when evaluated in terms of the categories of events rather than when evaluated using aggregated values, which suggests that working with the event categories might provide more consistent results across sources.

5. Conclusion

In this final section, we will suggest four directions for future development of event data. We envision all of these as projects that could be worked on by multiple research in the event data community in North America, Europe and elsewhere, and these efforts would generally enhance our ability to use event data in both academic and policy-oriented research.

5.1. Multiple source chronologies

As our research on the differences between Reuters and AFP indicate, multiple news sources are complementary. Multiple sources of news about political activities are now readily available from data services such as NEXIS and Dow-Jones Interactive, and the texts that are downloaded for coding could be extended to regional sources such as major newspapers and news magazines in addition to the wire services. Many current news reports are now available at no cost on the World Wide Web, and it would be useful to construct a system to automatically code these in real time.

However, major events will generate duplicate stories even when only a single news source is used. A major on-going story such as an outbreak of violence or the successful conclusion of a negotiation will be repeated numerous times, and some routine stories will be repeated because of corrections to earlier stories or the addition of information. The inability to automatically detect duplicate articles has discouraged researchers from using multiple news sources except in areas such as Africa where coverage is so limited that duplication is not a major problem.

We suspect that it would possible to develop an automated system to detect duplicates and, in the process, produce an automatic "re-write" of the subject-verb-object part of the stories into a standardized form. This would have the additional advantage of getting around the intellectual property problems that prevent news text from being shared among researchers. While the contractual provisions of the licenses of the most electronic databases and the "fair use" provisions of copyright law allow texts to be coded within a single institution, there is no way to share large amounts of text between institutions. This means that one either has to repeat the process of downloading and reformatting, or else just use a previously-coded data set. As a consequence, one of the major promises of automated coding—the ability to finely tune an event coding system to fit a specific theoretical problem—has so far not been realized. A standardized chronology would not duplicate the original texts and therefore would be free of copyright restrictions, so it could be shared across institutions.

The system we have in mind would focus on coding the date, location, and subject-verb-object (SVO) structure that has been used in traditional event data. The subjects and objects are nouns. In the case of proper nouns (for example the names of nation-states, political parties, and individual leaders), we would include a standardized name for the individual. This get around the problem of multiple transliterations of names such as Libyan leader "Qaddafi," multiple languages ("Ivory Coast" versus "Côte d'Ivoire") and American versus British spelling ("Organization" versus "Organisation"). The system could also maintain a continually-updated list of the official position (or positions) that an individual holds; this could be added to the record even when it is not explicit in the story.

The coding of verbs and verb-phrases could be done on the basis of the widely-used WordNet lexical database (Fellbaum 1999; `http://www.cogsci.princeton.edu/~wn/`). WordNet, developed by the Cognitive Science Laboratory at Princeton University under the direction of

George Miller, provides a very large list of English nouns, verbs, adjectives, and adverbs organized hierarchically. The system is free of commercial license restrictions and the database, source code for using the database, and Macintosh, Unix, and Windows versions of the program are available on-line. Wordnet could also be used to code common nouns such as “police”, “demonstrators”, “settlers”, and “guerrillas”.

There are several advantages to WordNet. First, WordNet is organized around synonym sets, which means that one would be able to identify the same event reported in different sources even when the reporters are using somewhat different vocabulary; this is also likely to simplify the task of translating coding dictionaries across multiple languages. Second, the hierarchical organization of WordNet provides a method of systematically classifying verbs that have not already been assigned an event code: The system can simply back up in the synonym hierarchy until it encounters a classification level where a code has been assigned, and use that.

Finally, and perhaps most importantly, WordNet has been developed by computational linguists and cognitive scientists, and is based on extensive studies of how language is actually used, rather than assembled on an ad hoc basis by political scientists. Consequently we anticipate that WordNet classifications will be much more likely to have utility in future re-coding of the data to answer new theoretical concerns.

5.2. New coding schemes

Machine coding allows researchers to experiment with alternative coding rules that reflect a particular theoretical perspective or interest in a specific set of issues. Both COPDAB and WEIS were both developed during the Cold War and assume a "Westphalian- Clausewitzian" political world view of sovereign states reacting to each other through diplomacy and military threats. Consequently these system are ill-suited to dealing with contemporary issues such as ethnic conflict, low-intensity conflict, organized criminal activity, or multilateral intervention.[17] These systems have other problems as well: for example WEIS has only a single category of “military engagement” that must encompass everything from a shot fired at a border patrol to the strategic bombing of cities. COPDAB contains only 16 event categories (compared to the 63 of

WEIS), and these are intended to span a single conflict-cooperation continuum that many researchers consider inappropriate.

The "lock-in" of these early coding systems is readily explained by the time-consuming nature of human event coding from paper and microfilm sources. Because human coders typically produce between five and ten events per hour, and a large data set contains tens of thousands of events, experimental re-coding was not possible. Established protocols for training and maintaining consistency among coders presumably further constrained efforts to modify WEIS and COPDAB once these were institutionalized. Marginal changes might be made—as found, for example, in Tomlinson's (1993) incremental extensions of WEIS or the GEDS (Davies and McDaniel 1993) extensions of COPDAB—but these then introduce incompatibilities with the earlier data that required statistical "splicing" (Goldstein and Freeman 1990; Ruenveny and Kang 1996) and complicated analyzing the data as a single series.

Automating coding, in contrast, allows even a long series of texts spanning multiple decades to be recoded in a few minutes. Because the coding rules applied to the texts are transparently preserved in the dictionaries and algorithms used for the coding—rather than in some exponentially-decaying tribal lore of codebooks and training protocols—customized coding schemes can be developed as the field evolves theoretically. Automated coding allows a researcher to focus his or her efforts on maximizing the *validity* of a coding scheme for a particular problem, since the automated coding process itself guarantees the reliability of the system.

At the present time, the most promising new development is the IDEA—Integrated Data for Events Analysis—coding system, which is being developed by Doug Bond, Joe Bond, Craig Jenkins and Charles Taylor (see Taylor, Jenkins & Doug Bond 1999; Taylor, Bond, Bond, Jenkins, & Kuzucu 1999). This will supercede the PANDA coding scheme, and more is designed to provide a general framework for coding events.

> "The IDEA event form typology is a conceptual framework for use in coding social, economic and political events data. The IDEA framework is an extension and a refinement of, and is congruent with the World Event / Interaction Survey or WEIS. Like WEIS, IDEA is

---

[17] There have been some efforts to extend the WEIS and COPDAB—most notably Leng's (1987) Behavioral Correlates of War (BCOW) and the Bond et al (1997) Protocol for the Analysis of Nonviolent Direct Action (PANDA)—but WEIS and COPDAB still dominate the research literature

> nominally scaled, but unlike WEIS the event forms in IDEA are not bound to state actors (though some event forms are intrinsically bound to specific actors like military forces, as in military engagement). For example, the WEIS reduction in relations event form represents a diplomatic behavior and is therefore restricted to inter-state behavior, but the IDEA equivalent, reduce routine activity, refers to such reductions by individuals, groups or organizations, both state and non-state. (`http://vranet.com/idea/`; accessed 31 May 2001)

IDEA has been designed to be backwards-compatible with WEIS, COPDAB and BCOW—it contains an explicit mapping of the IDEA codes onto their equivalents in the earlier systems—but contains codes for a variety of activities that were not available in the other systems, and more detail for some of the existing categories. For example, IDEA has about twenty different codes for uses of force, whereas WEIS and COPDAB contained only three.

IDEA is also designed to be extendible, so that new codes can be added in the future to provide greater detail on certain types of events. Like WEIS, IDEA is organized hierarchically—in fact it uses a superset of the WEIS "cue categories" for its most general level of abstraction. Categories are nested to three levels, however, in contrast to WEIS's two levels: for example

```
09: Request
     093: Ask for material aid
          0933: Ask for humanitarian aid
```

When additional coding categories are added within this structure, a researcher who does not require the additional detail can simply truncate the code to a higher level. Thus code 0933 would become 093 in the original WEIS system, or 09 in the WEIS cue category system.

We are currently in the process of developing a set of IDEA coding dictionaries that will focus on international third-party mediation. This will involve adding a number of codes in categories such as 08 (Agree), 10 (Propose), 14 (Deny), and 15 (Demand), which currently have relatively few sub-categories. We are also translating our existing WEIS dictionaries to work with the new IDEA subcategories that Bond et al have already defined. We hope to have this work finished by the end of the summer, and we will produce mediation-oriented data sets for the Levant, former Yugoslavia and possibly West Africa.

5.3. Pattern-recognition methods that use discrete categories

A long-standing problem in event data research has been the disconjuncture between the nominal coding of event data and the interval-level variables used in most statistical time series

analysis. McClelland originally intended for WEIS data to be analyzed in term of patterns of discrete behaviors.[18] However, after some years of work with event data focusing on several crises, he concluded:

> It proved relatively easy to discern event patterns and sequences intuitively. We found we could follow the successions of action and response in flow diagram form. Stages of crisis and the linkage of event types to temporary *status quo* situations also were amenable to investigation. We were defeated, however, in the attempt to categorize and measure event sequences. This was an early expectation that was disappointed by the data which showed too few significant sequences to support quantitative or systematic treatment. (McClelland, 1970:33)

As a consequence of this problem, McClelland's "World News Index" project, published in the mid-1970s, used interval-level variables in its measures. With the hindsight of two decades, the failure of a discrete event approach appears due to a paucity of data and processing capability. McClelland writes of analyzing hundreds or at most thousands of events; a contemporary event data researcher has available tens of thousands of events and computer power sufficient to work with millions.

Subsequent work with WEIS has generally used event data aggregated over time—typically by week, month or year—and scaled with weights such as those found in Vincent's (1990) and Goldstein's (1992) scales. This aggregation changes the data set from an irregular, nominal-level time series to a regular, interval-level time series. All of these scales have been constructed by querying panels of experts about the relative intensity of various event categories. In most cases, the weights have been assigned on a single cooperation-conflict dimension. Beyond this, there is no consistency in the scaling—for example the Goldstein weights are roughly proportional to the logarithm of the Azar-Sloan weights for similar events.

The advantage of the interval-level approach is that a wide variety of methods are readily available in the econometric literature. The clear disadvantage is that the process of reducing

---

[18] In contrast, Azar, while using WEIS as the model for COPDAB, abandoned McClelland's nominal categories in favor of an interval-level measure and approached coding as a scaling problem. Azar and Sloan (1975) consists entirely of interval-level data and Azar emphasizes

> quantitative aggregations, called here 'analytic data', [which] are summaries of the weighted frequencies of interactions. They describe the amount of conflict or cooperation exchanged between or within nation-states over some unit of time. (Azar 1980:150)

behavior to a single dimension through scaling loses a great deal of information. In principle (although almost never in actual data), a month characterized by a large amount of conflict in the first two weeks (negative numbers on most scales), followed by a large amount of reconciliation in the last two weeks (positive numbers) could aggregate to value close to zero, which is the same value that would occur in a month where nothing happened. However, this possibility has not prevented the scaled data from being used successfully in a variety of studies, in part because event data have been primarily employed to study highly conflictual situations such as the Cold War and the Middle East, where "cooperation" is largely expressed as a reduction of conflict.

A second, more subtle, problem occurs with aggregation: it removes the analysis a step further from the cognitive and organizational processes that are generating the events. While decision-makers may do some aggregation—one of the most commonly used metaphors in political analysis is indicating whether a situation is "heating up" or "cooling down"—detailed political responses are usually triggered by specific sets or sequences of events, not by the crossing of some numerical threshold. In political activity, unlike economic activity, both the stimuli and responses are likely to be discrete, not continuous. Political events move in jumps that are predicated on the prior state of the system. The fall of a single rocket following a period of peace will trigger a major response, whereas the fall of a single rocket during a period of war usually will go unnoticed. A model that can maintain the event data in its disaggregated form is, *ceteris paribus*, more likely to be successful in predicting actual behavior.[19]

The alternative to scaling would be the development of analytical methods that work with discrete data. We have experimented in the past with a number of these, including genetic algorithms, hidden Markov models, and an assortment of sequence comparison methods (see Schrodt 1991, 1991a, 1999, 2000, 2000a), and there is some additional work on this problem coming out of an “artificial intelligence” perspective (e.g. Hudson 1990; Trappl, Fürnkatz, Petrak & Bercovitch 1997), with neural networks being the most widely employed method (e.g., Beck, King & Zeng 2000, Esty et al 1998, Schrodt 1989). To date, however, no method or set of methods has emerged that is clearly superior to the interval-level time series techniques,

[19] It is also unclear whether a single weighting system can be applied in all contexts. For example, why should the same set of weights apply to a dispute such as Israel-Lebanon, where military exchanges were very common, and USA-USSR, where military exchanges were virtually nonexistent?

particularly in terms of making inferences about which sets of events are influencing the development of a situation. For example, while both neural networks and hidden Markov models are generally very good at making *predictions* based on event data, both have very diffuse parameter structures that make it very difficult to ascertain which event categories are most important.[20] There is definitely a need for further work in this area.

5.4.Open-source tools

Our final suggestion is that additional efforts be put into the development of open-source tools for the analysis of event data in particular, and to deal with texts relevant to political behavior more generally.

Over the past thirty years the quantitative international relations community has become quite good about sharing *data*—in fact there was never really a time when that community wasn't committed to sharing data, and we are apparently considerably better at this than some of our colleagues in other subfields. But we have been less good about sharing in the development of *software.* Given the generally successful implementation of the data-sharing norm, this is probably not due to a disinclination to provide the source code for software. Instead, this may be due to the lack of any clear indication that anyone would want the source code, the fact that until about fifteen years ago computer systems were very idiosyncratic and any given program was unlikely to run on more than a few computers, and, of course, the fact that computer programmers have always resisted documenting their code to the point where it can be easily modified.

The first two situations have changed over the past ten years (the third factor will never change…). The open source software movement, pioneered by Richard Stallman's GNU project and Linux Torvald's Linux operating system, have provided both the legal framework for the sharing of source code, and a large community of programmers that are willing and able to work

[20] This can also be a problem in interval-level studies that use event data, notably with the VAR method used by Goldstein and Freeman (1990) and Goldstein and Pevehouse (1996).

with open source code.[21] The development of this community has been enhanced tremendously by the World Wide Web, which enables researchers almost anywhere in the world to collaborate.

The past twenty years have also seen the emergence of both official and de facto standards in some areas of computer programming. In particular, the ANSI C/C++ programming language has reached a level of standardization where a program can be run without any changes on multiple operating systems. For example, we regularly shift our C programs between the Macintosh and Linux systems on personal computers, as well as onto "mainframe" Unix systems running on single-processor and parallel-processor machines. The perl language, which is optimized for text processing—and itself a product of the open source movement—is similarly compatible. Perl and the GNU gcc C compiler are available at no cost, so any researcher with access to a computer and an internet connection can obtain these tools.

These developments suggest that there may be more scope for collaboration than there has been in the past, particularly on tools to solve general-purpose problems such as text formatting, word disambiguation, and parsing. Text processing—to say nothing of event data analysis—will always remain a somewhat specialized task compared to conventional statistical processing, and collaboration within this diffuse research community might be very helpful.

5.5. Final Thoughts

There are two fundamental reasons for using event data in political analysis. First, politics does not have the convenient numerical measures such as location, momentum, and temperature found in physics, or variables such as price, interest rates, and GNP found in economics. Political activity instead consists largely of discrete actions and communications directed from one actor to another over time. McClelland's original observations on the potential utility of event data as a method of addressing this problem still hold:

---

[21] This does not preclude commercial development of software, since the open-source concept is entirely compatible with for-profit programming. Successful open-source efforts should as the gcc C++ compiler, the Apache web server and the perl and Python text-processing languages have focused on solving general problems, and then commercial ventures work on specific applications. The fact that IBM—scarcely known as a charitable organization—is investing over a billion dollars in open-source development should be good evidence of this. The KEDS project has been collaborating informally with for-profit text analysis companies such as Virtual Research Associates (Doug Bond) and Social Science Automation (Michael Young). (Microsoft, in contrast, has recently condemned open-source software as "unAmerican," a characterization that we trust will serve to further increase the attractiveness of the open-source approach in Europe and elsewhere.)

> International conduct, expressed in terms of event data, is the chief dependent variable of international relations research. ... [This] starting point is provided as readily by the ordering principle of classical diplomatic history as by the basic concepts of general system analysis. Thus, we may assert that the prime intellectual task in the study of international relations is to account for actions and responses of states in international politics by relating these to the purposes of statecraft or, alternatively, we can say that the problem is to account for the relations among components of the international system by analyzing the characteristics of the various components of that system by tracing recurring processes within these components. [Both definitions] carry about the same information and involve nearly the same range of choices of inquiry and analysis. (McClelland 1970:6)

Second, human analysts have a limited ability to absorb vast quantities of largely redundant material. The text of NEXIS news wire leads covering only Israeli-Palestinian interactions for 1989 alone runs to some 300 pages. The full articles would fill perhaps 2000 pages; we suspect that few researchers would read all of these. The task becomes even more formidable if one is dealing with a long time series such as the Cold War: just what were the USA. and USSR. doing on 16 August 1955? While most human analysts can memorize the day-to-day details of a short time period such as the Cuban Missile Crisis, or the major events of a long period such as the Cold War, we are skeptical about the human ability to memorize, much less analyze, day-to-day details for a long time period.

Event data fill that gap. The text of the journalistic sources provide memory and a variety of statistical and other computational methods can provide analysis. Between the text and analysis, one needs something similar in content to event data.

Science magazine once surveyed how new techniques in the physical and biological sciences sometimes revolutionized not just the methodologies, but also the theories, within their fields:

> Not everybody appreciates the importance of technique. Many scientists, in fact, are "theory snobs" who dismiss technique as a kind of blue-collar suburb of science. . . . [But there is,] clearly, enormous transforming power in techniques. In the absence of an essential technique, a researcher or a field flounders, developing elegant theories that cannot be decisively accepted or rejected—no matter how many intriguing circumstantial observations are available. But with a key technique in hand, the individual and field move ahead at almost terrifying speed, finding the right conditions to test one hypothesis after another. Conversely, new techniques often uncover new phenomena that demand new theories to explain them. (Hall 1992: 345)

The research environment in international relations, and much of comparative politics, is arguably theory rich and data poor, with too many theories are chasing too few facts. At the

same time, the interactions in international system are becoming more complex with the end of the Cold War and the need is greater than ever to be able to systematically study alternative theoretical explanations for that behavior.

One of the favorite parables employed by evangelical preachers is that of a sailing ship becalmed for weeks in the Atlantic, its crew slowly dying of thirst. Sighting a passing vessel,[22] the beleaguered crew appeals frantically for water. The crew of the other ship replies, "Throw down your buckets; you are surrounded by fresh water!": they are resting in the outflow of the mighty Amazon River.

The quantitative international relations community has often felt becalmed with respect to data. We have no American National Election Study, no U.S. Census or National Institutes of Justice data, and only so many ways one can analyze the World Handbook, Correlates of War, WEIS and COPDAB. But in fact, we are sitting amid a river of political data—both event-oriented and contextual—flowing past us every day from journalistic sources. Those sources are increasingly machine-readable, and if we can find a means of tapping them using the natural language capabilities of contemporary computers, we will find ourselves awash in data.

[22] As with many parables beloved of evangelical preachers, the movement of this second vessel under windless conditions is not explained...

Bibliography

Azar, Edward E. 1980. "The Conflict and Peace Data Bank (COPDAB) Project." *Journal of Conflict Resolution* 24:143-152.

Azar, Edward E. 1982. *The Codebook of the Conflict and Peace Data Bank (COPDAB).* College Park, MD: Center for International Development, University of Maryland.

Azar, Edward E. and Thomas Sloan. 1975. *Dimensions of Interaction*. Pittsburgh: University Center for International Studies, University of Pittsburgh.

Azar, Edward E., and Joseph Ben-Dak. 1975. *Theory and Practice of Events Research*. New York: Gordon and Breach.

Azar, Edward E., Richard A. Brody, and Charles A. McClelland, eds. 1972. *International Events Interaction Analysis: Some Research Considerations.* Beverly Hills: Sage Publications.

Beck, Nathaniel, Gary King, and Langche Zeng. 2000. "Improving Quantitative Studies of International Conflict: A Conjecture." *American Political Science Review* 94, 1: 21-36.

Bond, Doug, Brad Bennett and William Vogele. 1994. "Data development and interaction events analysis using KEDS/PANDA: an interim report." Paper presented at the annual meeting of the International Studies Association, Washington.

Bond, Doug, J. Craig Jenkins, Charles L. Taylor and Kurt Schock. 1997. "Mapping Mass Political Conflict and Civil Society: The Automated Development of Event Data." *Journal of Conflict Resolution* 41, 4: 553-579.

Burgess, Philip M., and Raymond W. Lawton. 1972. *Indicators of International Behavior: An Assessment of Events Data Research.* Beverly Hills: Sage Publications.

Davies, John, and Barbara Harff. 1998. "Dynamic Data for Conflict Early Warning." In *Risk Assessment and Crisis Early Warning Systems*, ed. John Davies and Ted R. Gurr. Pp. 79-94. New York: Rowman and Littlefield.

Davies, John L., and Chad K. McDaniel. 1993. "The Global Event-Data System." In *International Event-Data Developments: DDIR Phase II*, ed. Richard L. Merritt, Robert G. Muncaster, and Dina A. Zinnes. Ann Arbor: University of Michigan Press.

Dixon, William J. 1986. "Reciprocity in United States-Soviet Relations: Multiple Symmetry or Issue Linkage." *American Journal of Political Science* 30:421-45.

Edwards, George C., III, and B. Dan Wood. 1999. "Who Influences Whom? The President and the Public Agenda." *American Political Science Review* 93, 2: 327-344.

Fellbaum, Christiane, ed. 1999. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Gerner, Deborah J., Philip A. Schrodt, Ronald A. Francisco, and Judith L. Weddle. 1994. "The Machine Coding of Events from Regional and International Sources." *International Studies Quarterly* 38, 1: 91-119.

Gerner, Deborah J.and Philip A. Schrodt. 1998. "The Effects of Media Coverage on Crisis Assessment and Early Warning in the Middle East." In *Early Warning and Early Response*, ed. Susanne Schmeidl and Howard Adelman. New York: Columbia University Press-Columbia International Affairs Online.

Gerner, Deborah J. and Philip A. Schrodt. 2001. "Analyzing the Dynamics of International Mediation Processes in the Middle East and the former Yugoslavia." Paper presented at the annual meeting of the International Studies Association, Chicago.

Gerner, Deborah J., and Ian S. Wilbur. 2000. "Semantics or Substance? Showdown between the United States and the Palestine Liberation Organization." Pew Case Studies in International Affairs, Institute for the Study of Diplomacy, Georgetown University.

Goldstein, Joshua S., and John R. Freeman. 1990. *Three-Way Street: Strategic Reciprocity in World Politics*. Chicago: University of Chicago Press.

Goldstein, Joshua S. 1992. "A Conflict-Cooperation Scale for WEIS Events Data." *Journal of Conflict Resolution* 36, 3: 369-385.

Goldstein, Joshua S., and Jon C. Pevehouse. 1997. "Reciprocity, Bullying and International Cooperation: A Time-Series Analysis of the Bosnia Conflict." *American Political Science Review* 91, 3: 515-530.

Goldstein, Joshua S. , Jon C. Pevehouse, Deborah J. Gerner, and Shebli Telhami. forthcoming. "Dynamics of Middle East Conflict and US Influence." *Journal of Conflict Resolution*

Hall, S. S. 1992 "How Technique is Changing Science." *Science* 257:344-349.

Hudson, Valerie, ed. 1991. *Artificial Intelligence and International Politics*. Boulder: Westview

Huxtable, Phillip A. 1997. *Uncertainty and Foreign Policy-Making: Conflict and Cooperation in West Africa*. Ph.D. dissertation, University of Kansas.

Huxtable, Phillip A. and Jon C. Pevehouse. 1996. "Potential Validity Problems in Events Data Collection." *International Studies Notes* 21, 2: 8-19.

Laurance, Edward J. 1990. "Events Data and Policy Analysis." *Policy Sciences* 23:111-132.

Leng, Russell J. 1987. *Behavioral Correlates of War, 1816-1975.* (ICPSR 8606). Ann Arbor: Inter-University Consortium for Political and Social Research.

Kovar, K., J. Fürnkranz, J. Petrak, B. Pfahringer, R. Trappl, and G. Widmer. 2000. "Searching for Patterns in Political Event Sequences." *Cybernetics and Systems* 31, 6.

McClelland, Charles A. 1970. "Some Effects on Theory from the International Event Analysis Movement." mimeo, University of Southern California, February, 1970

McClelland, Charles A. 1976. *World Event/Interaction Survey Codebook.* (ICPSR 5211). Ann Arbor: Inter-University Consortium for Political and Social Research.

Merritt, Richard L., Robert G. Muncaster, and Dina A. Zinnes, eds. 1993. *Theory and Management of International Event Data: DDIR Phase II*. Ann Arbor: University of Michigan Press.

Moore, Will H. 1995. "Action, Reaction or Rational Expectations? Reciprocity and the Domestic: International Conflict Nexus during the Rhodesia Problem." *Journal of Conflict Resolution* 39, 1: 129-167.

Pevehouse, Jon C., and Joshua S. Goldstein. 1999. "Serbian Compliance or Defiance in Kosovo? Statistical Analysis and Real-Time Predictions." *Journal of Conflict Resolution* 43, 4: 538-546.

Reuveny, Rafael, and Heejoon Kang. 1996b. "International Conflict and Cooperation: Splicing the COPDAB and WEIS Series." *International Studies Quarterly* 40,2:281-305.

Schrodt, Philip A. 1989. "Short Term Prediction of International Events using a Holland Classifier." *Mathematical and Computer Modeling* 12: 589-600.

Schrodt, Philip A. 1990. "Parallel Event Sequences in International Crises." *Political Behavior* 12:97-123.

Schrodt, Philip A. 1991. "Classification of Interstate Conflict Outcomes using a Bootstrapped ID3 Algorithm." *Political Analysis* 2: 31-56.

Schrodt, Philip A. 1991a. "Prediction of Interstate Conflict Outcomes Using a Neural Network." *Social Science Computer Review* 9, 4: 359-380.

Schrodt, Philip A. 1999. "Early Warning of Conflict in Southern Lebanon using Hidden Markov Models." In *The Understanding and Management of Global Violence*, ed. Harvey Starr. Pp. 131-162. New York: St. Martin's Press.

Schrodt, Philip A. 2000. "Pattern Recognition of International Crises using Hidden Markov Models." In *Political Complexity: Nonlinear Models of Politics*, ed. Diana Richards. Pp. 296-328. Ann Arbor: University of Michigan Press.

Schrodt, Philip A. 2000a. "Forecasting Conflict in the Balkans using Hidden Markov Models." Paper presented at the American Political Science Association meetings, Washington, DC

Schrodt, Philip A., and Deborah J. Gerner. 1994. "Validity assessment of a machine-coded event data set for the Middle East, 1982-1992." *American Journal of Political Science*, 38: 825-854.

Schrodt, Philip A., and Deborah J. Gerner. 1997. "Empirical Indicators of Crisis Phase in the Middle East, 1982-1995." *Journal of Conflict Resolution* 41, 4: 529-552.

Schrodt, Philip A., and Deborah J. Gerner. 1998. "Cluster Analysis as an Early Warning Technique for the Middle East, 1979-1996." In *Risk Assessment and Crisis Early Warning Systems.* ed. John Davies and Ted R. Gurr. Pp. 95-107. New York: Rowman and Littlefield.

Schrodt, Philip A., and Deborah J. Gerner. 1998a. "An Event Data Set for the Arabian/Persian Gulf Region 1979-1997." Paper presented at the International Studies Association, Minneapolis, March 1998.

Schrodt, Philip A., and Deborah J. Gerner. 2000. "Cluster-Based Early Warning Indicators for Political Change in the Contemporary Levant." *American Political Science Review* 94,4.

Schrodt, Philip A., Shannon G. Davis, and Judith L. Weddle. 1994. "Political Science: KEDS—A Program for the Machine Coding of Event Data." *Social Science Computer Review* 12, 3: 561-588.

Sherman, Frank L. 1994. "SHERFACS: A Cross-Paradigm, Hierarchical and Contextually Sensitive Conflict Management Data Set." *International Interactions* 20, 1-2: 79-100.

Sherman, Frank L., and Laura Neack. 1993. "Imagining the Possibilities: The Possibilities of Isolating the Genome of International Conflict From the SHERFACS Dataset." In *Theory and Management of International Event Data: DDIR Phase II*, ed. Richard L. Merritt, Robert G. Muncaster, and Dina A. Zinnes. Pp. 87-112. Ann Arbor: University of Michigan Press.

Sigler, John H., John O. Field, and Murray L. Adelman. 1972. *Applications of Events Data Analysis: Cases, Issues and Programs in International Interaction.* Beverly Hills: Sage.

Sommer, H. and J. Scarritt. 1998. "The Utility of Reuters for Events Analysis in Area Studies: The Case of Zambia-Zimbabwe Interactions, 1982-1993." *International Interactions* 25: 29-59.

Taylor, Charles L., and Michael C. Hudson. 1972. *World Handbook of Political and Social Indicators,* 2nd ed. New Haven, CT: Yale University Press.

Taylor, Charles L., and David A. Jodice. 1983. *World Handbook of Political and Social Indicators,* 3rd ed. New Haven, CT: Yale University Press.

Taylor, Charles Lewis, J. Craig Jenkins, and Doug Bond. 1999. "Trends in Protest and Mass Contention: Integrated Data for Events Analysis." Paper presented at the American Political Science Association meetings, Atlanta.

Taylor, Charles Lewis, Joe Bond, Doug Bond, J. Craig Jenkins, and Zeynep Kuzucu. 1999. "Conflict-Cooperation for Interstate and Intrastate Interactions: An Expansion of the Goldstein Scale." Paper presented at the International Studies Association meetings, Washington.

Thomas, G. Dale. 1999. *The "Strange Attractiveness" of Protracted Social Conflict in Northern Ireland.* Ph.D. dissertation, University of South Carolina.

Trappl, Robert, J. Fürnkatz, J. Petrak, and J. Bercovitch. 1997. "Machine Learning and Case-Based Reasoning." In *Learning, Networks and Statistics*, ed. Riccia, G.D., R. Kruse and H-J. Lenz. Vienna: Springer-Verlag.

Vincent, Jack E. 1983. "WEIS vs. COPDAB: Correspondence Problems." *International Studies Quarterly* 27: 160-169.

Vincent, Jack E. 1990. "Interdependence Reexamined." *International Interactions* 16:91-107.

Ward, Michael and Sheen Rajmaira. 1992. "Reciprocity and Norms in U.S.-Soviet Foreign Policy." *Journal of Conflict Resolution* 36, 2: 342-368.

Wood, B. Dan, and Jeffrey S. Peake. 1998. "The Dynamics of Foreign Policy Agenda Setting." *American Political Science Review* 92, 1: 173-184.