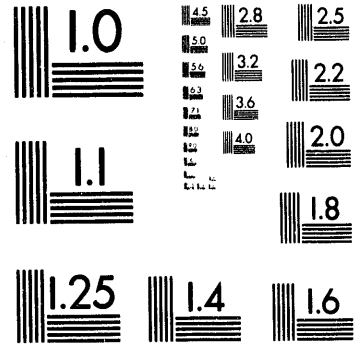
1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
7.1

1 of 1

SAND93-2289C
Conf-940301--26

Semantic Modeling for Theory Clarification: The Realist vs Liberal International Relations Perspective

Olin H. Bray
Sandia National Laboratories
Albuquerque, New Mexico
and
Political Science Department
University of New Mexico
Albuquerque, New Mexico

Prepared for delivery at the Annual Meeting of the
Western Political Science Conference
March 10-12, 1994
Albuquerque, New Mexico

This work was supported by the United States Department of Energy under Contract DE-AC04-94AL85000.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

Abstract:

This paper describes a natural language based, semantic information modeling methodology and explores its use and value in clarifying and comparing political science theories and frameworks. As an example, the paper uses this methodology to clarify and compare some of the basic concepts and relationships in the realist (e.g. Waltz) and the liberal (e.g. Rosenau) paradigms for international relations. The methodology can provide three types of benefits: (1) it can clarify and make explicit exactly what is meant by a concept; (2) it can often identify unanticipated implications and consequence of concepts and relationships; and (3) it can help in identifying and operationalizing testable hypotheses.

Introduction:

This paper describes a natural language based semantic information modeling methodology and explores its use and value in clarifying and comparing political science theories and frameworks. This methodology was developed in the information systems area as a formal methodology to clarify communications between users and system developers. As an example, the paper uses this methodology to clarify and compare the basic concepts and relationships in the realist (e.g. Waltz) and the liberal (e.g. Rosenau) paradigms for international relations. The models focus on the levels of analysis, the types of actors, their resources, and the relationships among the actors. The paper then shows the role of international regimes within each of these models. The paper shows that the methodology can provide three types of benefits: (1) it can clarify and make explicit exactly what is meant by a concept; (2) it can often identify unanticipated implications and consequence of concepts and relationships; and (3) it can help in identifying and operationalizing testable hypotheses. The paper concludes with a summary of the lessons learned and general applicability of this semantic modeling methodology for theory clarification in political science.

Theory Development and Problems:

Theory development in political science, as in any science, proceeds through a series of steps. First, concepts must be generated and clarified.

The more precisely defined they are the better for articulating the theory and its consequences. Second, relationships among these concepts must be identified and defined. Part of the clarification of these relationships involves determining whether the relationship applies to the entire concept or only to specific parts of it. Third, the concepts must be operationalized. To accurately operationalize a concept, it is useful, if not essential, to know the relationships in which that concept participates. Fourth, testable hypotheses are formulated. Fifth, the experiment and/or data collection is performed. Sixth, the hypotheses are tested using the data that was collected. Seventh, conclusions are drawn from the results and if necessary the theory is refined, which takes us back the first step one where existing concepts may be refined (or even deleted) and new concepts are generated and added to the existing set.

The above steps definitely apply under the assumption that theory is developing in a linear progression, as concepts and relationships are clarified, become more precise, and are tested and refined, as with Lakatos' progressive problemshift. However, Kuhn proposes that science progress through a series of paradigmatic revolutions. In these cases, the above steps initially crystalize and refine the anomalies that precipitate the crisis and then begins to clarify the concepts involved in the new paradigm. Although Kuhn's position is that the new paradigm is completely different from the old one and not an outgrowth or extension, at the level of individual facts there is overlap across paradigms unless all of the fundamental concepts are completely different. Determining the number of objects and facts that are reused across paradigms provides a metric for the difference between two paradigms. This would allow an actual test of Kuhn's hypothesis (actually stated as a assumption or principle) about the differences among paradigms.

Two problems that complicate the theory development process are a lack of clarity and a lack of formalism or in some cases formalizing the concepts too late in the process, such as when they are being operationalized. The lack of clarity may involve the concepts or the relationships among them. It often involves ambiguity and lack of precision. One possible solution to this problem involves a more precise formal statement of the concepts and relationships. Unfortunately, this formalism often occurs too late. — Too late for what?

The next section describes a modeling methodology to increase the precision and reduce the ambiguity with which a theory, potentially any theory, can be stated and analyzed. Furthermore, the methodology can be used from the initial statement of both the concepts and relationships throughout the refinement and testing process, i.e. throughout all of the seven steps. This approach is semantic or natural language based information modeling. The next section describes the methodology with general explanatory examples from the international relations area. The following three sections apply the methodology specifically to build basic models for parts of the realist and the liberal perspectives of international relations and for international regimes.

Semantic Modeling/Information Modeling:

This semantic modeling methodology was developed in the information systems area to address the communications problem between users, who had a problem they needed to solve, and the information system developers, who needed to completely and accurately understand the problem before they could develop an information system to solve it. The usually approach was for the user to describe the problem usually in a general narrative form. The information system developers took this narrative and translated it into a system specification, much of it in computer terminology, which the user had to agree to -- even if they cannot understand it. Once the specification was approved, the system was designed and built. Unfortunately, most of these information systems did not initially work the way the users expected. In fact, many of them never worked successfully and were scraped.

The imprecision and ambiguity in the initial problem statements concealed various mistakes, misunderstandings, gaps, and inconsistencies. (This is the same problem that often occurs when discussing political science or international relations theory.) The problem was further complicated because different people, often with different backgrounds, were looking at this ambiguous problem statement and trying to understand it and correct it on the fly. The formalism came in the system design, far too late to identify and correct the real problems. The semantic modeling methodology avoids these problems by initially

formally modeling what the user says. This is done in a structured natural language form so that every one can understand precisely what is being said. Specific examples are used to eliminate any remaining ambiguity. This approach has two benefits. First, the problem statement (or the semantic model) can be understood and critiqued by anyone working in the area. Second, the process of developing the formal model forces one to clarify exactly what is meant.

The basic construct of the methodology is an elementary sentence or fact, which specifies that an object plays a role with another object. For example, a country is a member of an alliance. The two objects are country and alliance and the role is "is a member of". An elementary sentence may include more than two objects: "a country joined an alliance in a year." A fact is elementary if it cannot be decomposed into other facts without losing meaning.

A more detailed information model includes examples and constraints that are derived from these examples. An example of a constrained fact would be "a country may be a member of one or more alliances." The complementary fact (obtained by considering the relationship in the opposite direction) is "an alliance must include one or more countries." These constraints are determined by considering a set of examples:

Country	Alliance
US	NATO
UK	NATO
Poland	WTO
US	SEATO
Country Y	-
-	Alliance X (an invalid example)

The two constrained facts were derived by realizing that only the last example is invalid. The examples clearly show that an alliance can have more than one member country, that a country can belong to multiple alliances, and that a country does not have to belong to any alliance. Only the last example is invalid -- an alliance does not exist without member countries

The benefit of this methodology in clarifying concepts and relationships becomes apparent if we consider two additional examples for this fact -- the country the US is a member of the alliance GATT and the country the US is a member of the alliance the UN. The immediate reaction is that these two examples are invalid. GATT and the UN are not alliances. One is an international trade regime and the other is an international organization. Everyone knows this.

However, according to the rules of the semantic modeling methodology, two objects are of different types only if they have different facts. Some facts may be common, but at least some facts must be different for the different objects. The methodology forces the issue. How are alliances, international regimes, and international organizations different? They all have countries as members. Countries joined them in a year. They were established on a date. They have a purpose. They all have an establishing document or treaty. If you want to deal with them as different types of objects, the methodology forces you to specify precisely how they are different.

This is not a trivial issue. If they are really all equivalent types of objects, then the research about them and the hypotheses are cumulative. On the other hand, assume for the moment that they are different, but they have fifteen facts in common and fifteen that are different. An obvious question is does hypothesis testing about the fifteen common facts provide similar results. This again raises the possibility of cumulative results. A key point is that the modeling methodology quickly identifies different concepts masquerading under the same term or conversely different terms being used for the same concept.

Consider another example that is central to discussing the realist versus the liberal perspective. For the realists, the international system consists of states, or conversely a state is a member of the international system. For the liberals a crucial distinction is that the international system consists of actors, or conversely an actor is a member of the international system. When we consider specific examples of actors (e.g. the U.S., Japan, NATO, the UN, IBM, and Greenpeace), we discover that some types of facts are common to all actors, while other facts apply only to

certain types of actors. This indicates a subtype-supertype relationship. The common facts apply to the supertype - actor. The different facts are specific to each subtype of actor, such as state, alliance, international organization, company, and transnational organization.

How do we compare the richness and power of the two perspectives? One approach is to say that any facts or hypotheses relating to any subtype of actor other than the state are excluded from or invisible to the realist perspective or model. A second approach is that what is considered a state may be different or broader than is normally assumed. For example, alliances may be treated as "quasi-states" so that facts and hypotheses related to alliances may be included. A third possibility is that the realists are simply not interested in facts and hypotheses about these other subtypes of actors, which is another perfectly legitimate way to carve up the world and the phenomena to be explained.

The next two sections briefly summarize the information models for the realist and the liberal perspective. The realist model includes more facts about the international system than about the states, while the liberal model includes more facts about states and other type of actors. The following section then discusses international regimes.

Realist Model:

For the realists, the semantic model would include a set of basic entities or concepts -- the international system itself, states, and resources. Other concepts and entity types will emerge as the model is presented. The rest of this section simply specifies some of the facts for the realist model and discusses several questions that arise from these facts.

An international system includes one or more states.
An international system has N number of member states.
An international system exhibits one or more behaviors.
(e.g. balancing or bipolarity or another type of behavior such as free trade or restricted trade)
An international system is controlled by certain rules.
(These rules define acceptable behavior. If the rules change,

then by implication the international system changes.)

A state has a certain level of resources of a resource type.
(e.g. state X has a $500 billion GDP or 15 army divisions)
A state has sovereignty over a territory.
(or a state has sovereignty over a territory for a function or a state has authority in a territory for a function)
A state performs one or more functions.
(or a state performs one or more functions with a structure.)
State 1 with respect to state 2 has a level of power with a resource type.

Although a complete model for the realist approach would include many more facts, these few facts raise a number of questions. First, the international system only includes state, no other types of actors. Second, is sovereignty complete within a territory or can it be restricted to or exclude certain functions? Perhaps this question is better phrased in terms of authority rather than sovereignty. Third, the fact that a state performs functions at least allows the possibility that different states perform different functions. At a very high level, clearly all states perform a common set of functions. However, at the detailed level required by this modeling methodology, the answer is not as clear. Are the U.S. and Japan functionally equivalent in terms of defense or nuclear deterence or control of their domestic economy? The realist answer is probably yes they are functionally equivalent and that what we are really distinguishing is level of capability not functionality. They would capture this distinction in the fact about the relative power between two states. However, it is not clear that this is always true, especially in the case of marginal or failed states, such as Bosnia or Somalia, which may not be able to perform some of the functions normally associated with states. Finally, when authority is related to a function, is an additional fact needed to indicate that a state's authority over a function is enforced using a certain type of resource. In other words, are different types of resources more effective at enforcing authority in certain functional areas -- and can this differ for different parts of a state's territory?

Liberal Model:

An international system includes one or more actors.
An international system has N number of member actors.
An international system exhibits one or more behaviors.
(e.g. balancing or bipolarity or another type of behavior such as free trade or restricted trade)
An international system is controlled by certain rules.
(These rules define acceptable behavior. If the rules change, then by implication the international system changes.)

An actor is of a type.
(state, international organization, subnational group, transnational group, alliance, regime, etc.)
(For example, Rosenau list a number of types of actors, each of which has its own set of facts, although there are some facts common to all actors.)

An actor may be a member of one or more alliances.
An actor may be a member of one or more international organizations.
An state may contain one or more subnational groups.
An state may include part of one or more transnational groups.
An actor has a certain level of resources of a resource type.
(e.g. state X has a $500 billion GDP or 15 army divisions)
A state has sovereignty over a territory.
(or a state has sovereignty over a territory for a function or a state or an actor has authority in a territory for a function)
An actor performs one or more functions.
(or a state performs one or more functions with a structure.)
Actor 1 with respect to actor 2 has a level of power with a resource type.

An international organization is founded on one or more agreements or treaties.

A multinational corporation (MNC) operates in two or more states.

These facts raise questions about actors, functions, and resources.

Do different types of actors have access to different types of resources? Do different types of actors perform different functions, with the same or different types of resources?

International Regimes:

An international regime includes one or more states.
(or more generally can members be actors of any type?)
An international regime may include a dominant state.
An international regime must have a primary issue area.
An international regime may have one or more secondary issue areas.
An international regime is guided by one or more norms.
An international regime exhibits one or more behaviors.
(Which is different from the fact that
an actor [in a regime] exhibits one or more behaviors.)
An international regime must be headquartered at a location.
An international regime may have a staff of size N.
An international regime must communicate with its members through one or more channels.
An international regime is founded on one or more agreements or treaties.
A state may be a member of one or more international regimes.
A state's norms can be compared to an international regime's norms.
A state's behavior can be compared to an international regime's behavior.

These examples raise a number of questions about international regimes. First, is an international regime only a collection of states or can it include other types of actors? If it can include other types of actors, then are there different types of regimes based on their types of members and are their norms and behaviors different? A second set of questions relate to the issue areas. Must there be any relationship among the issues areas for an international regime? Can different regimes cover the same set of issue areas? How do international regimes interact or do they even begin to merge as their issue areas expand and overlap? Third, can an international regime or must it have a headquarters and a support staff? Considering the above facts about an international regime, how is

an international regime different from an international organization or is it simply a subtype of an international organization, and if so what makes it different? Many of these questions arise if one tries to specify examples for these objects and facts. Is GATT simply an international regime or does it include an organization (formally or informally as part of each country's bureaucracy) supporting that regime? Consider the IAEA (International Atomic Energy Agency). Is it an international regime or just an international organization supporting an international regime? If it is the latter case, does this make it a special type of international organization?

Clearly, none of these three models is yet complete and rigorous. However, the purpose of this paper was not to provide a completed model. It was to show the benefits of this natural language based, semantic modeling methodology in terms of forcing concept clarification early in the process and making it easy to integrate concepts and relationships.

Conclusions:

This paper has briefly described semantic or natural language based information modeling as a technique for theoretical precision and clarification. It applied the modeling approach to the realist and liberal perspective of international relations and identified a number of differences and issues. The purpose of the paper was to show how the methodology can be used to more precisely specify and relate concepts for improved theory development and hypothesis testing.

One issue not addressed in this paper involves data collection and hypothesis testing. Recall that the methodology was developed to improve communications for defining and building information systems. Therefore, carried to completion, this methodology will help operationalize concepts and result in a theoretically correct database design for capturing and analyzing research data on any area that has been modeled. Furthermore, if existing data archives are modeled using this approach, the results will help show where and how data from these various sources can be combined.

References:

International Relations

Keohane, Robert O. (ed). 1986. Neorealism and Its Critics. New York: Columbia University Press.

Keohane, Robert O. 1992. "A Functional Theory of Regimes." in Robert J. Art and Robert Jervis. (ed). International Politics: Enduring Concepts and Contemporary Issues. 3rd Edition. New York: HarperCollins Publishers Inc.

Rosenau, James N. 1990. Turbulence in World Politics. Princeton: Princeton University Press.

Waltz, Kenneth N. 1979. Theory of International Politics. Reading, MA: Addison-Wesley.

Semantic Information Modeling Methodology

Bray, Olin H. 1988. "Chapter 6: Database Design". Computer Integrated Manufacturing: The Data Management Strategy. Bedford, MA: Digital Press.

Meersman, Robert. 1988. "Towards Models for Practical Reasoning about Conceptual Database Design." Data and Knowledge (DS-2), North-Holland.

Nijssen, G. M. and T. A. Halpin. 1989. Conceptual Schema and Relational Database Design: A Fact Oriented Approach. New York: Prentice-Hall.

Appendix: Natural language Information Analysis Methodology

This Appendix provides a more detailed description of the methodology and has been used as an introduction for training information systems and user personnel in the methodology.

NIAM provides a technology independent, fact-based model of the application area's information requirements. The information model of this real world specifies the entity types, the types of facts about each entity type, and the roles these entity types play with each other. Examples of entity types are person, department, and project. Examples of roles are: "works in", "manages", and "assigned to". Examples of facts include:

A person works in a department.

A department employs a person.

(The complement of the above fact)

A person manages a department.

A person is assigned to a project.

The information model can be represented in either of two ways -- verbally or graphically. In the verbal representation, the structured natural language sentences can be read, critiqued, and corrected by anyone who knows the subject matter with virtually no explanation of the methodology. The graphical representation shows the relationships among the entity types more clearly and concisely, but it does require a few minutes of explanation to be able to read it.

To completely capture all of this information, a deep structured sentence or fact has a specific form. It specifies the first entity type, its identifier or label type, several examples or instances of that label type, a verb phrase, and another entity's set of information (i.e. entity type, label type, and label instance). An example of a deep structured sentence is, "a person identified by social security number 123-45-6789 works in the department identified by department name engineering." If the sentence is not binary, then there are additional verbs and entity information.

Examples or fact instances are critical because they explicitly

define the data constraints, which the DBMS must enforce. Both uniqueness and total constraints can be determined from the following examples:

	Person	Department
1.	Sam	Eng
2.	Mary	Mfg
3.	Bill	Eng
4.X	Sam	Finance
5.X	Joe	-
6.	-	Accounting

When shown the above six examples, the user can quickly determine which ones are good. Examples 1 and 2 are good because there is no overlap, they are two independent fact instances. Example 3 is good because a department (Eng) can have more than one person in it. However, example 4 is incorrect because a person (Sam) can only be in one department. This defines a uniqueness constraint: a specific instance of a person can only appear once in this fact type. The possible uniqueness constraints are that the instance on the left may be unique, the instance on the right may be unique, both instances may be unique, or the combination may be unique. An example of fact with the combination is "a person is assigned to a project". A person can be assigned to many projects and a project can have many people, but you would not assign Sam to project X twice. Examples 5 and 6 specify the total constraints. The user can tell that example 5 is invalid since every person must be in a department. However, according to example 6, departments can exist without employees.

There is also an equivalent graphic representation of the information model. The solid circle represents an entity type, such as a person, a department, a part, or a release status. Dashed or dotted circles represent data objects that identify or further describe real objects, such as employee name, social security number, or release code. Boxes or rectangles represent the roles played by one object type with respect to another. The two boxes together indicate that two roles are complementary -- a person works in a department and a department employs a person. With appropriate naming, facts in a graphic model can

be read as sentences.

Figure 2 shows several basic facts (in both directions) and their constraints. The constrained facts shown include: a person must be identified by one SSN, a person must work in one department, a person may design one or more parts, a person may be assigned to one or more project, and a part must have one current release status. (Note that the model must specify "current" release status because a part will have many release statuses over time. The constraints are also shown. The V indicates a total constraint and the line over a role indicates uniqueness. Obviously there are additional constraint types and symbols, but this should provide the reader with a general understanding of the graphic model representation.

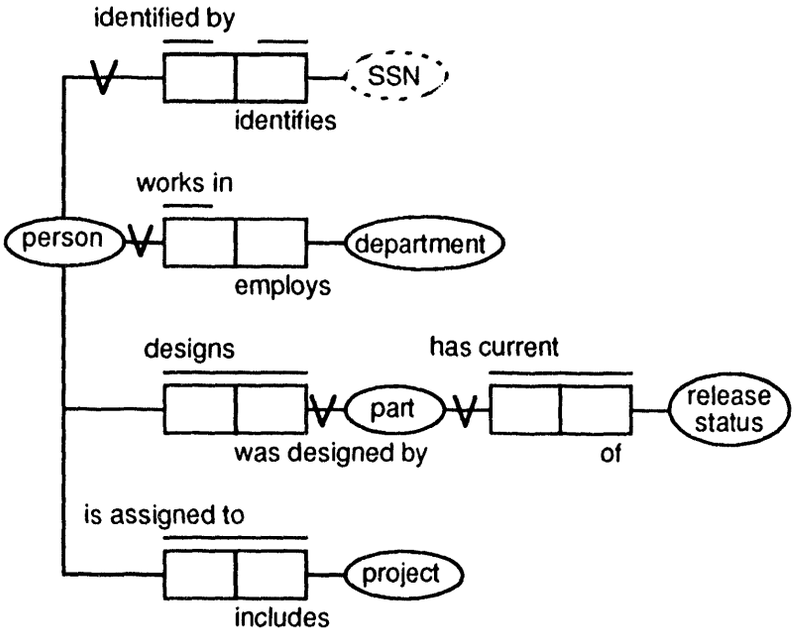

Figure 2. Example of Graphic Representation of Information Model

There is an algorithm that generates a neutral data model (NDM) from the information model. This NDM can be represented in any of the traditional data modeling notations. This NDM is then used to generate an initial database schema.

DATE
FILMED
5/31/94
END