12th International Command and Control Research and Technology Symposium
"Adapting C2 to the 21st Century"

Semantic Interoperability: Revisiting the Theory of Signs and Ontology Alignment Principles

Coalition Interoperability, Ontology alignment, Semantic distance

Eric Dorion
Situation Analysis and Information Fusion Group
Decision Support Systems Section
Defence R&D Canada – Valcartier
2459, Pie-XI North,
Val-Bélair, Québec Canada, G3J 1X5
Eric.Dorion@drdc-rddc.gc.ca

Stéphane Fortin
Information Technologies Consultant
DMR Consulting (Fujitsu Consulting)
Place Iberville Trois
2960, Boul. Laurier, Suite 400
Québec, Québec, Canada, G1V 4S1
Stephane.Fortin@dmr.ca

Abstract. The necessity to conduct military operations in coalition has been established many times throughout history. It is a sociological, political and technological requirement. From a technology point of view, having the militaries work together in coalition imposes technical interoperability requirements on their respective supporting C2 information systems (C2ISs). Since these C2ISs are not developed concurrently, it ensues that while the semantic concepts they handle are similar, they are not expressed with the same data structures. To technologically enable the military coalition at the semantic level, there are 2 possible solutions: Either through promoting usage of a single semantic representation or ontology, or through a translation or mapping of the concepts from one data representation to the other. The latter approach is termed ontology alignment. This paper addresses this approach from an ontology engineering perspective. We explain the challenges of aligning ontologies, the possible consequences and means to assess the semantic distance between them. To this effect, we will revisit semiotic – the theory of signs – as a philosophical foundation to support the ontological engineer in aligning ontologies.

Introduction

The necessity to conduct military operations in coalition has been established many times throughout history. It is a sociological, political and technological requirement. From a technology point of view, having the militaries work together in coalition imposes technical interoperability requirements on their respective supporting C2 information systems (C2ISs). Since these C2ISs are not developed concurrently, it ensues that while the semantic concepts they handle are similar, they are not expressed with the same data structures. To technologically enable the military coalition at the semantic level, there are 2 possible generic approaches: The first is through promoting the usage of a single semantic representation or ontology. For example, The Multilateral Interoperability Programme (MIP) [1], a NATO effort (STANAG 5525), designs its *MIP solution* which is two-fold: The Joint Consultation Command and Control Information Exchange Data Model (JC3IEDM) and the Data Exchange Mechanism (DEM). Both are enriched with thousand pages of documentation in order to explicitly define the concepts, their relationships and constraints that are relevant for their operational community (the military coalition). The second approach is through a translation or mapping of the concepts from one data representation to the other, hereafter termed ontology alignment. An effort to translate some of the OTH-T GOLD messages into the JC3IEDM data format and vice-versa conducted here in DRDC – Valcartier exemplifies this approach [2].

In either approach, there are fundamental principles that must guide the ontologist[1] in his task to specify and formalize the concepts and reify them into symbols. This paper aims at describing some of these principles and to illustrate them with the use of the semiotic triangle which in turn exemplifies some of the principles of the Theory of Signs or

[1] In this text, an *ontologist* designates the person that assesses the semantic meaning of terms or symbols used to designate the concepts under consideration within a specific community. It is an encompassing term for *data modelers*, *knowledge engineers*, *ontological engineers*, *information analysts, translators* and the like.

Semiotic[2]. Also, since semantic interoperability is involved, we will describe the concept of semantic distance (or conversely semantic similarity) that is introduced in ontological alignment. Finally, we will describe some of the work that is on-going at DRDC – Valcartier that relates to this.

Revisiting the Semiotic Triangle

The semiotic triangle (Figure 1) was popularized by Ogden and Richards in [3] and draws heavily upon Charles Peirce Theory of Signs or Semiotic.

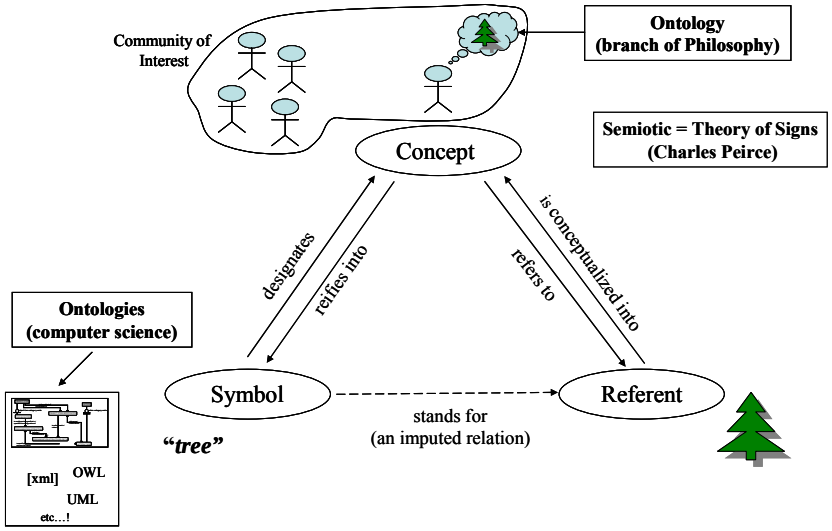


Figure 1: The Semiotic Triangle

It all starts with an object, a *referent*, something that *exists in itself* and that is to be the focus of a particular interest. On the choice of "*referent*" to represent the object, Ogden and Richards argue that:

> "The word 'thing' is unsuitable for the analysis here undertaken, because in popular usage it is restricted to material substances – a fact which had led philosophers to favour the terms 'entity', 'ens' or 'object' as the general name for whatever is. It has seemed desirable, therefore, to introduce a technical term to stand for whatever we may be thinking of or referring to. 'Object', though this is its original use, has been adopted, though its etymological form is open to question when considered in relation to other participial derivatives, such as agent or reagent. But even in Latin the present participle occasionally (e.g. *vehens* in equo) admitted of variation in use; and in English an analogy with substantives, such as 'reagent', 'extent', and 'incident' may be urged. Thus the fact that 'referent' in what follows stands for a thing and not an active person, should cause no confusion."

[2] We simply use the term *semiotic* at the first degree in this paper. We did not reflect on possible differences between this term and *semeiotic*.

Referent was the classical term since then. As already noted, this referent is to be the focus of interest of someone. Through our limited senses and intellect, we form a certain *conceptualization* of this referent biased by a specific interest or point of view. For Kant [4] this conceptualization cannot encompass all of the essence of the referent but only some of its exposed phenomena. Furthermore, the categorical capacity of humans to 'seize' or 'abstract' the phenomenon exposed by the referent is determined *a priori*! Whether this view represents the actual agreement in Philosophy or not is out of the scope of this paper, but is useful in noting that the branch of Philosophy that aims at identifying these categories is the Ontology. This term was seemingly used for the first time around the 12th century but was the focus of work of many philosophers since Aristotle. As a branch of Philosophy, Ontology's object has yet to be fully circumscribed.

As a surrogate (however imperfect) for the real world, the set of conceptualizations one uses enables him to act according to certain goals. However, this mental representation is only useful for the bearer of the conceptualization unless it is *shared* with others. In order to do anything that transcends the boundaries of the intellect, a set of *symbols* that *denote* the concepts must be explicitly defined. Indeed, symbols are the only means by which concepts can be shared among people. That is, in order to share concepts within a *community of interest*, there is a need to *reify* concepts into symbols.

Figure 1 shows that a symbol *stands for* a referent (the dotted line between symbol and referent vertices). The only valid relationship between a symbol and a referent is articulated through its specific conceptualization. The choice for a symbol per se is completely arbitrary (Ferdinand de Saussure's *arbitraire du signe)* [5]. However, the choice for a symbol must be *consensual* in the sense that the community of interest must agree about the concept it will denote. In practice, consensus is usually reached when the proposed symbol shows a quality of *naturalness* in its representation of the concept. Although qualitative in nature, naturalness increases with time and exposure. For example, the symbol "1" that denotes the concept of unity is very familiar to any reader of this paper since he was exposed extensively to this symbol from his early youth. There is also a fair chance that we will agree on what "1" means, hence the consensus.

Sets of symbols may be organized in such constrained structures that they will enable the denotation of complex thoughts. Natural languages are used precisely for that purpose. In computer science, a similar exercise leads to the construction of symbol structures that are called *ontologies*[3]. Studer *et al* [6] define an ontology as "*a formal, explicit specification of a shared conceptualization*" which naturally fits the concepts we try to explain in this paper. Under this definition, any data structure[4] that explicitly defines and formalizes a set of concepts that are shared by a specific community of interest is an

[3] It is somewhat paradoxical that the symbol "ontology" in Philosophy and Computer science does not always denote the same concept. In Philosophy, it seems to be far more encompassing while in Computer Science it is more restrictive. There are similarities of course but this, more often than not, leads to confusion.

[4] We use « data structure » herein as an encompassing term for any method that aims at formalizing a conceptualization. It includes data models, OWL ontologies, UML diagrams and the like.

ontology. That includes OWL representations, taxonomies, data models, UML diagrams and the like. The choice of a specific formalism depends on the degree of expressiveness needed to capture the semantics of the domain under consideration. Whatever the data structure, the ontologist is the one that must define symbols denoting a certain conceptualization in such a way that:

- They formalize consensual knowledge of the community of interest,
- They are specific in the sense that they denote the concepts unambiguously,
- They respect a certain quality of naturalness,
- The relationships and constraints between concepts are also formalized.

The Case for Systems Interoperability

The need to conduct military operations in coalition implies that information systems supporting them have to maintain some degree of interaction. These systems, although they may address the same functional area (e.g. command and control) are dissimilar on many aspects. First, they are usually built and maintained by the participating countries that have their own systems engineering practices and distinct capability acquisition processes. This results in systems that are dissimilar on many levels and "plugging" them together poses significant technical challenges, even more under stringent operational requirements. Nevertheless, system-to-system interactions happen all the time in the real world so this question must be given some thought.

At the abstract level, systems intercommunication can be thought of as an *addresser* sending a *message* **μ** to an *addressee,* where two information systems can take both these roles in turn. For some reason, **μ** is subject to some transformation and **μ'** is the message perceived instead. That is, **δ = μ – μ'** represents a distorsion that happened in the original message. It is a function of several variables, subject of research in many fields. In linguistics one could try to characterize **δ** against Jakobson's conceptual model issued from his theory of communicative functions (Figure 2). Taking this approach, we could say that:

$\Delta = f(r, p, e, c, ph, m)$, where r, p, e, c, ph and m correspond to Jakobson's communicative functions.

Although initially intended to represent verbal communications, this model has been used also for other conceptual domains where information exchange between agents take place. A similar assumption could be made for systems interoperability. At the very least, this could help ontologists to avoiding over-simplifying the task of symbol translation between ontologies. Specifically to this paper, the metalingual communicative function represents the *code* or the ontology that is to be agreed upon by the addresser and addressee.

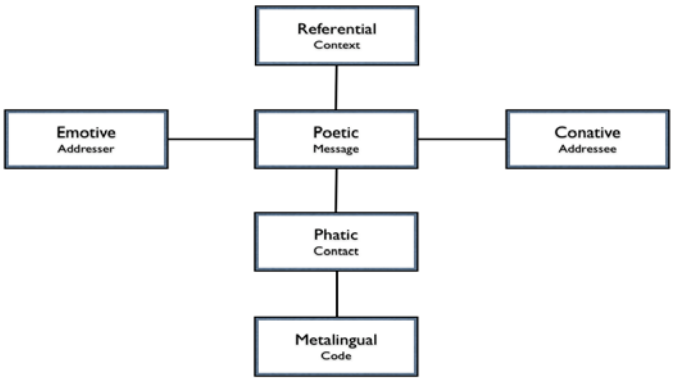


Figure 2: Jakobson's Theory of Communicative Functions [7]

Emotive function
This function relates to the "addresser", as indicated in the diagram. This means that it refers to the initial intention, the idea and/or message to be transmitted by the addresser.

Conative function
This is the function that refers to the "addressee". It is the function that acts upon the receptor of the information, idea and/or message, transmitted by the addresser. One common example of this function can be seen through publicity where the message is directly aimed at the listener (ie. the "addressee"). Another broader use of this function can be identified through direct orders, or the use of the vocative.

Phatic function
The phatic function refers to the channel of communication. It acts as a mean to establish, maintain or end the communication. One of the most common exemple would be the various interjections we use such as "Hello" at the beginning of a conversation, "Roger" or "Over" in military radio communications.

Metalinguistic function
This function refers to the code itself. It would be a direct link to the symbols in ontology. Some would describe it as the function of language about language.

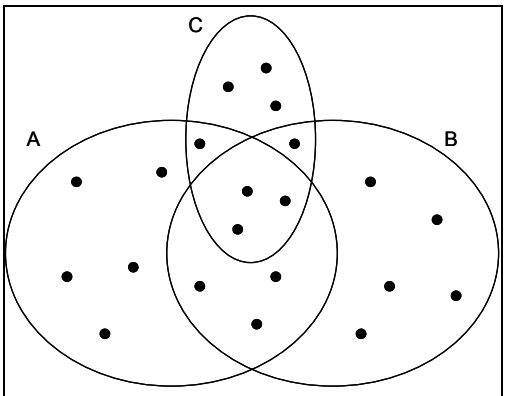
Referential function
As its name states it, this function refers to the context or the referent. An exemple would be the word "platform" which would have one specific meaning in construction, yet a totally different one in a military context.

Poetic function
The poetic function is not limited to poetry, unlike some people would like to believe. It refers to the form of the message where it carries an expressive value. This function will materialize the signs and codes of the communication into a message that will generate indirect messages. This is often seen in sentences having a double meaning. This double meaning is usually the work of the poetic function.

The Case for Semantic Interoperability

While we recognize that there must be some common semantic ground (the *code* in Jakobson's model) between an addresser and an addressee (or between agents) we must also recognize that the addresser and the addressee often bathe in different ontological contexts (e.g. Two Navy and Army officers). This doesn't deter the need for them to communicate and the same goes for their systems. [2] and [8] state that systems interoperability is achievable only for the concepts that are shared by both systems (Figure 3). [2] also states that the main driving force to justify interoperability is the operational context. That is to say that the justification in trying to interconnect two systems lies in the operational need to do so. This also exacerbates the information exchange requirements that will be considered in the engineering of the solution.

For those concepts in Figure 3 that are comprised as elements of the set defined by $A \cap B \cap C$, a semantic interoperability solution is found by establishing a relationship of equality between the symbols of A and B that denote those concepts. When conceptualizations are modeled with the use crisp sets, the semantic interoperability problem seems trivial. However, we claim that a better model to represent conceptualizations would be *fuzzy sets* where concept elements can be members of several conceptualizations according to certain membership functions [9]. To us, this uncertainty on the membership of conceptual elements to conceptual sets better reflects the reality that concepts are a little different from a conceptualization to another. The establishment of relationships between elements of conceptual sets is far more subtle than in the latter model which also reflects reality. To convince ourselves of this fact and for the sake of illustrating this, let us consider two domain ontologies where a list of country codes is a concept comprised in both ontologies.



Let A and B be the sets of concepts Systems A and B encompass in their respective ontologies.

Then, A ∩ B represents the set of concepts that are understood by both systems. Interoperability can only take place with these concepts.

Let C be the set of concepts that are manipulated within the realm of a military operation.

Then, an interoperability solution that involves systems A and B can only address a subset set of the operational needs by manipulating the concepts comprised in the set defined by A ∩ B ∩ C.

Figure 3: Semantic Interoperability

We would find values that correspond to Canada, United States and so on. The arbitrariness of the sign principle shows us that in fact "CAN" or "CA" could be used to represent Canada (or "United States" although the naturalness principle would prevent it) and it may be so in our two ontologies. The ontologist would be rightly tempted here to establish a relationship of synonymy between "CAN" and "CA", reflecting that these symbols both correspond to the concept of Canada as a country. But what our ontologist really knows about these two ontological subsets is conditioned by his own *interpretation* of the domains. The concepts denoted by the symbols "CAN" and "CA" might present subtle differences in their ontological usage. [2] and [10] reflect on some of the tough choices we had to make in mapping data elements between the Over-The-Horizon Targeting GOLD (OTH-T GOLD) message format and the C2IEDM (later to be named JC3IEDM).

The Concept of Ontology Alignment

The example above illustrates the activity of *ontology alignment*[5]. We define it as

> *The act of establishing a relation of correspondence between two or more symbols from distinct ontologies, for those symbols that denote concepts that are semantically identical, or similar.*

[5] In this paper, *ontology alignment* will be used as an encompassing term for *ontology matching*, *ontology mapping*, *data mapping* and the like.

We say that concepts are similar when they point to referents that are, for all practical purposes, close in meaning. Indeed, individually we all have distinct abstractions of what is a "tree" because we have different perceptions on this object due to our senses, mental capacities, etc. Therefore, in a strict sense, concepts from distinct ontologies referring to a same referent cannot be aligned! However, for all practical matters we agree on what a "tree" is as a concept and are not afraid to establish a correspondence of similarity between the symbols "tree" and "arbre" for example.

Ontology alignment has been practiced for a long time by ontologists, almost as an art form. Jakobson gave a central role to the *poetic* function in his model. There are however several research efforts that aim at formalizing and even automate the process (see [11] for a state of the art on these efforts). Every approach has its own merits with varying degrees of success, but each and every one of them present an *ontological commitment* [12] to the concept of *semantic similarity* or conversely *semantic distance*[6].

The Concept of Semantic Distance

When two symbols denote concepts that are close but are not synonyms or completely congruent, we say that there is some sort of similarity (or distance) between these terms. In trying to align distinct ontologies together, we establish correspondence relationships between symbols where semantic distance is an inherent part of the resulting matching solution. In practice, we often discard such considerations, but this may result in semantic loss. We would like some sort of mechanism to alleviate the uncertainty on the exactness of a translation. To do so, we need some sort of a method to measure that distance, but even before that we are confronted with the very nature of that distance, what it represents and how it is defined. Of course, there are several views on this from many research efforts. Rodriguez in [13] gives an account of 4 classes of models for semantic similarity assessment and how it is defined within those contexts:

- *Feature-based models* define the semantic similarity by comparison of concept characteristics such as properties, roles, rules, etc.

- *Models based on semantic relations* make use of semantic networks and how far 2 nodes (concepts) are from one another. This model is not foreign to a certain practice motivated by structuralism.

- *Models based on information content* establish semantic similarity between concepts by comparison of how much information there is to support the concept and how specific it is. Its underlying assumption is that abstract concepts present less information and influence some function occurrence within a *corpus* of text. This model is less dependent on structure but very sensitive to aspects related to its assumption. It might not be the right model for semantic interoperability since two distinct ontologies might not present

[6] We will assume that *semantic distance* and *semantic similarity* are purely converse. We accept that one may find us wrong on this matter.

the same characteristics of abstractedness.

- *Context-based models* consider syntactic, semantic, pragmatic and stylistic factors to establish how words can be substituted within a sentence. The higher this degree of substitution the higher the semantic similarity. This model fits Jakobson's approach minus the *poetic* function.

Rodriguez also proposes the *matching-distance* model and tries to integrate the advantages of all previous approaches while avoiding the pitfalls. One of its underlying assumptions however is that it only considers the concepts of a single ontology, which seems to limit its applicability to semantic interoperability.

At the time of writing this paper, it is yet unclear to us what method is the most applicable to a coalition interoperability situation and how such measures would be defined and used at the application level. We are absolutely certain though that disregarding semantic distance in ontology alignment may lead to erroneous handling of information at the semantic level. Nonetheless, this approach is rather the norm than the exception when we devise interoperability solutions. We think that this should be studied more carefully in the future for command and control information systems put in contexts of military coalition operations.

Conclusion

In this paper, we tried to give the reader a philosophical perspective on the act of matching distinct ontologies. We revisited the semiotic triangle as a means to understand how symbols denote concepts that in turn refer to real objects. We justified its consideration in the context of coalition interoperability, where a number of systems must be interconnected. On the consequences of such requirements, we explained the concept of semantic distance that is an inherent characteristic of imperfect alignment of concepts. Future research at DRDC – Valcartier will lead us to the study of such measures and how they would be applied, for example, in multi-sensor data fusion systems.

References

1. Multilateral Interoperability Programme Website: http://www.mip-site.org/.
2. E. Dorion and A.-C. Boury-Brisset. *Information Engineering in Support of Multilateral Joint Operational Interoperability*. CCRTS, San Diego, 2004.
3. C. K. Ogden and I. A. Richards. *The meaning of meaning*. Harcourt Brace Jovanovich, San Diego, 1989, 363 pages.
4. I. Kant. *The critique of pure reason*. 1781.
5. F. de Saussure. *Cours de linguistique générale*. 1916
6. R. Studer, V. R. Benjamins, D. Fensel. *Knowledge Engineering: Principles and Methods*. Data & knowledge engineering, ISSN 0169-023X, 1998, vol. 25, n° 1-2, pp. 161-197.
7. R. Middleton. *Studying Popular Music*. Philadelphia: Open University Press. ISBN 0335152759.
8. N. Guarino, *Formal ontology and information systems*. Proceeding of the first international conference on formal ontologies in information systems, FOIS'98, Trento, Italy, IOS Press, June 1998, pages 3-15.
9. G. J. Klir and T. A. Folger. *Fuzzy sets, uncertainty, and information*. Prentice-Hall, NJ, USA, 1987.
10. E. Dorion, C. J. Matheus, M. M. Kokar. *Towards a formal ontology for military coalitions operations*. 10th International Command and Control Research and Technology Symposium. Washington, June 2005, 11 pages.
11. P. Shvaiko and J. Euzenat. *A survey of schema-based matching approaches*. Journal on Data Semantics IV, volume 3730/2005, December 2005, pp. 146-171.
12. A. Rayo. *Ontological commitment*. Forthcoming in Philosophical Compass. http://web.mit.edu/arayo/www.
13. M. A. Rodriguez. *Assessing semantic similarity among entity classes*. PhD. Thesis, University of Maine, 2000, 179 pages.