9th International Command and Control Research and Technology Symposium
Coalition Transformation: An Evolution of People, Processes and Technology to Enhance Interoperability
14-16 September 2004
Copenhagen, Denmark

Topic: C2 Experimentation

Paper's Title: Automatic Report Processing

Author
Name: Dr. Ulrich Schade
Organization: FGAN/FKIE
Address: Neuenahrer Straße 20
53343 Wachtberg-Werthhoven
Germany
Phon: +49 228 9435 436
Fax: +49 228 9435 685
E-Mail: schade@fgan.de

Automatic Report Processing

Dr. Ulrich Schade
FGAN/FKIE
Neuenahrer Straße 20
53343 Wachtberg-Werthhoven
Germany

Abstract

The paper presents a module to be integrated into C2 Information systems. This module is meant to process military reports given in natural language. The reports are transformed into a formal structure by means of computational linguistics. Afterwards, the structure is augmented by ontological processes in order to allow the actualization of the data base (C2IEDM) as well as the visualization of the report's content on the map.

A prototype has been developed to illustrate the module's abilities by a bunch of examples. Besides, the properties of the module are demonstrated and discussed. It is shown that the module adds to flexibility and, in particular, to interoperability.

Introduction

Technological progress always influences all spheres of life. Obviously, this includes the military sphere as well. As a consequence, military research in our days aims at the exploitation of Information Age concepts and technologies. It focuses on information enabling combat power under the key words "Network Centric Warfare (NCW)", "Network Centric Capabilities", "Shared Situational Awareness", and "Enhanced Collaboration and Self-Synchronization" (cf. [1]).

The basic tenets of NCW include the following:

- Deployed forces have to be connected robustly. They have to form a network in which information exchange is done automatically, without effort, and without any breakdown.
- The quality of the shared information has to allow for shared situational awareness.
- The interoperability among the forces has to allow for shared situational awareness as well. This is of particular relevance for coalition forces.

Although the connectedness of the forces is the physical precondition of information exchange, its realization will provide the lowest levels of network-centric capabilities, only ([2]: 86ff.). In order to achieve the more mature levels, it is necessary to optimize the quality of transferred information as well as the interoperability among the forces ([3]: 109). Thus, the C2 Information systems used by the forces, especially coalition forces, have to undergo a transformation beyond pure connectability as well. In analogy to what is required for the forces (ibid: 27), the systems have to be

- adaptable to the given situation,
- interoperable to allow for shared awareness,
- competent to make high quality information available,
- trustable and reliable.

A Report Processing System as Module of a C2 Information System

This paper presents a system under development called "SOKRATES system." It will be integrated as a module into existing C2 Information systems like German army's FAUST system (cf. [4]). The SOKRATES system processes information brought in by real-world military reports given in natural language. It provides automatic semantic analyses of the reports. Thus, it can be used if the incoming reports cannot be handled anymore due to their number or if they are given in a language the addressee does not comprehend well enough. A first prototype of the SOKRATES system has been completed in order to demonstrate its capabilities.

In its present form, the prototype is able to process and to analyze reports of moving actions, e.g., "*Fünf Bradyland Haubitzen marschieren von Nederveert nach Helmond über Someren*" (*Five Bradyland howitzers moving from Nederveert to Helmond via Someren*) or "*Fünf feindliche Kampfpanzer in Zufahrt*" (*Five hostile battle tanks approaching)*. It also deals with position reports like "*Haben 31UFT785235 erreicht*" (*Arrived at 31UFT785235*). The report examples the system processes had been constructed on the base of reports recorded during German army exercises. After transcription and linguistic classification, names within the reports had been changed in order to make them anonymous for further use.

Figure 1 provides a blueprint of the SOKRATES system. It shows the stages of the process the system passes through while analyzing a report. First, the report is divided into sentence-like units. Second, these units are transformed into a formal representation by means of Information Extraction [5]. Third, the resulting formal representation is augmented by ontological processes. Finally, the augmented representation is used for post-processing. With respect to the prototype, post-processing includes the actualization of the data base and the visualization of the report's content on the map. In the following, the latter three of these stages are discussed in more detail with a special focus on the augmentation by ontological processes.

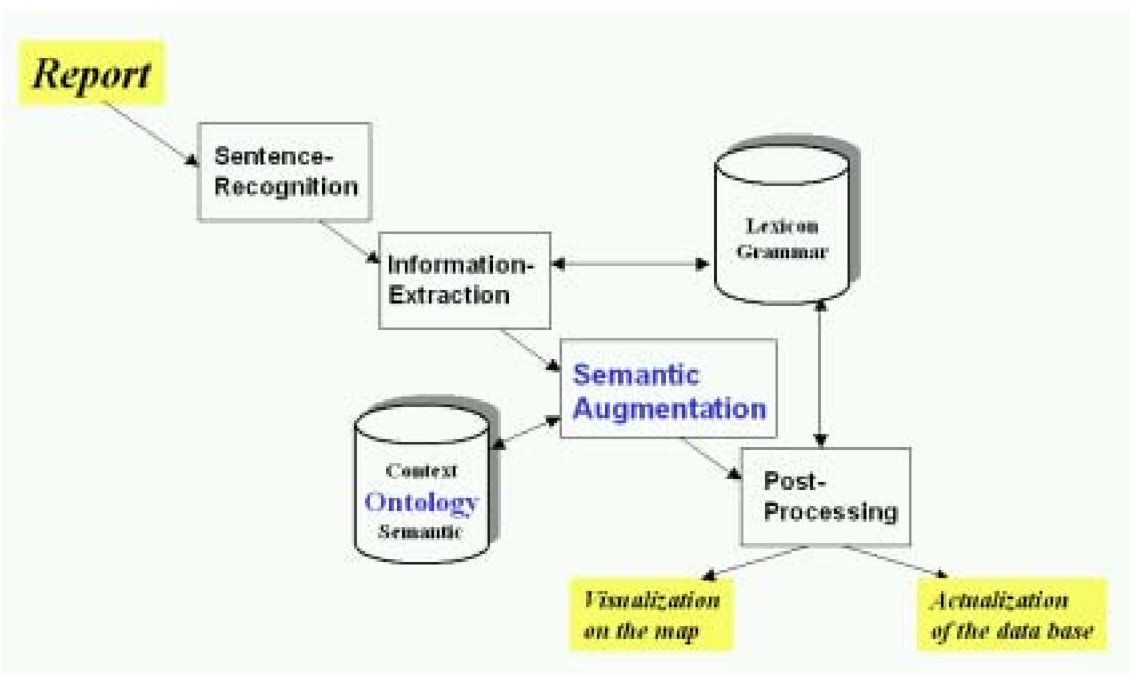

Figure 1: The blueprint of the SOKRATES system

Information Extraction

The information given by the reports in written natural language has to be represented formally. This is the task of the information extraction component. The component is based on the SMES system [6]. The result of its processing is a feature structure, the standard representation format used by unification-based processing systems in the field of computer linguistics [7, 8]. Mathematically, feature structures are finite sets of pairs of attributes and values. Attributes are atomic, and values are either atomic or feature structures by themselves. In the case of report processing, the information extraction builds up a feature structure of type “report” (cf. figure 2). This means, this feature structure includes an element of which “type” is the attribute and “report” is the value. This feature structure also includes as mandatory element one with attribute “sender.” The value of “sender” is a feature structure of type “unit” representing the sender of the report. Thus, the sender’s feature structure incorporates elements describing the relevant qualities of the unit in question, e.g., “name: 4./PzGrenBtl332, Zug C” or “size: PLT” (PLT for platoon). The content of the report is represented as value of “reporting_data” which is another mandatory attribute of the report’s feature structure.

```
type:               report

sender:             | type:      unit                    |
                    | name:      4./PzGrenBtl332-Zug C   |
                    | size:      PLT                     |

reporting_datetime:     | type:      datetime |
                        | year:      2004     |
                        | month:     9        |
                        | day:       5        |
                        | hour:      12       |
                        | minute:    18       |

reporting_data:         | type:      in_position                                     |
                        |                                                            |
                        | agent:     | type:     unit                    |           |
                        |            | name:     4./PzGrenBtl332-Zug C   |           |
                        |            | size:     PLT                     |           |
                        |                                                            |
                        | location:  | type:     town        |                       |
                        |            | name:     Nederveert  |                       |
                        |                                                            |
                        | qualifier: exactly_at                                      |
```

Figure 2: A feature structure resulting from analyzing the report “*051218ZSEP04 von 4./PzGrenBtl332-Zug C: Stehe in Nederveert*” (*051218ZSEP04 by 4./PzGrenBtl332-Zug C: Standing at Nederveert*) by means of information extraction

Details about the information extraction, its mode of operation, the resulting formal representation, and the specific attributes and values used are given in [9]. However, there are additional aspects to be noted here. First, the result of the information extraction, the feature structure, is no longer a German-specific representation. In contrast, the attributes as well as their values are derived from the C2IEDM [10], the data model used in the Multilateral Interoperability Programme [11]. This allows for an easy interaction with the data base (see below). It also ensures that the report processing results of the SOKRATES system can be interpreted and used by operators who do not speak German. Second, the order of feature structure elements (the pairs of attributes and values) is not important. Thus, this kind of representation abstracts from the surface structure of a report's sentence-like units. Therefore, an information extraction component for any language might be built up which only has to deliver a feature structure like the one at hand. It then can be connected to the rest of the SOKRATES system, and everything will operate smoothly. By the way, this aspect also projects some reports onto the same feature structure. It does not matter whether the report has been "*... fahren von Nederveert nach Helmond über Someren*" (*go from Nederveert to Helmond via Someren*) or "*... fahren von Nederveert über Someren nach Helmond*" (*go from Nederveert via Someren to Helmond*), the resulting structure is the same. Another aspect of worth is the intended incompleteness feature structure representations. On the one hand, this means that the missing of some information will not let the system fail. On the other hand, elements can be added by a well-defined method. Thus, augmentation is supported by the very nature of the structures.

Ontological Processing

There are lots of reasons to augment the result of the information extraction process. The most obvious arise from the needs of post-processing. In order to actualize the data base the inserts in question must respect the constraints given by the C2IEDM declaring some data fields as mandatory. Therefore, the respective data has to be present. If it is not in the feature structure after information extraction it has to be calculated and then filled in. Similarly, in order to visualize the report's content on a map, coordinates explicitly telling where to put the APP-6A-symbols of units and equipment mentioned have to be provided. Augmentation is done by ontological processing.

The ontology that is the basis of the ontological processes includes the usual object hierarchy to define objects and their properties. The hierarchy is defined by the ISA-relation with respect to object types. For example, it is represented that a M1A2 (Abrams) is a battle tank, a battle tank is a tank, a tank is a vehicle, and a vehicle is an object. Besides, instances of defined objects are included according to the "member of set"-relation. E.g., specific M1A2s, identifiable and distinguishable by their respective call sign, are instances of the type "M1A2 Abrams main battle tank." The ontology's domain is defined by the scenarios in which military reports occur, e.g., battle field scenarios. The hierarchy of object types had been constructed on the base of the C2IEDM, and the instances have to be added out of the actual data base. A previous version of the ontology under use had been presented in [12].

C2IEDM information is not only used for the construction of the class hierarchy, but also for the definition of the objects' properties and their values. Placing restrictions on values is a first step toward the representation of knowledge beyond the hierarchy. For example, movability (as property of land vehicles) allows the values *tracked*, *wheeled*, and *towed*. It is restricted to *tracked* in the case of M1A2s. However, the appropriate definition of the class hierarchy together with the definition of the classes' properties and a suitable restriction of their values is not enough. In order to augment the feature structures adequately, ontological processes must be added. Naturally, these processes operate on the hierarchy. In some cases, values are looked up, in other cases, they have to be calculated, and sometimes even more complex operations have to be carried out. In the following, these cases are illustrated by example.

As a first example, we take the report “*Stellung bezogen*” (*In position*) sent by 2./PzGrenBtl332-ZugB. Analyzing this report, information extraction will construct a feature structure for sender which is of type unit. By the way, this feature structure is also the value of the reported action’s agent. Thus, its symbol has to be displayed on the map during the post-processing of visualization. Information extraction provides the name of the unit (2./PzGrenBtl332-Zug B), and thus all the information needed to determine the unit’s symbol according the APP-6A can be checked up: size = PLT, category = COMBAT, arm category = INF(mechanized), mobility = TRACKED, hostility = FRIEND. Even the coordinates of the unit’s location can be checked up because the position deployed is represented in the data base and thus in the ontology as well.

The determination of values, especially, the determination of coordinates, however, is not always a simple check up. Let us assume as another example, the 2./PzGrenBtl332-ZugB reconnoiters enemy movements and reports “*Fünf Bradyland Haubitzen hinter Vinstedt*” (*Five Bradyland howitzers behind Vinstedt*). In this case, coordinates have to be calculated in order to place the howitzers’ symbol on the map. It is save to assume that the howitzers are “behind” the village of Vinstedt with respect to the position of the reporting unit. Thus, an axis is calculated from the reporting unit through the village of Vinstedt, and the howitzers’ symbol is placed on this axis next to Vinstedt, on the side where the reporting unit is not. A similar case is given, if the unit reports “*Drei feindliche T80 in Zufahrt*” (*Three T80s approaching)*. Again, coordinates have to be calculated. In this case, either the forefront of the reporting unit’s position or, if the line is unknown, the FEBA is taken as reference for the calculation as it can be assumed that the enemy approaches the forefront or comes from the direction of the FEBA, respectively. The calculations are rule-based whereas the rules can be counted as part of the knowledge the ontology holds. In the case of coordination calculation, they are based on the specific scenario (e.g., the location of the FEBA), military knowledge (e.g., positions are prepared under the expectation that the enemy will approach from a specific direction) and pragmatics, or to be more precisely, the proper conduct of reporting and the proper referring to place (cf. [13, 14]).

The last report (“*Drei feindliche T80 in Zufahrt*”) may also serve as example for more complex processing which exploits mainly military knowledge. Whenever equipment is mentioned in the report, like the battle tanks of the example, a unit determination process may check which kinds of unit hold such an equipment. With respect to the example, it can be assumed that T80s will be, most likely, operated by a battle tank unit, and because there are three of them the unit should be at least of platoon size. So, if the unit determination process runs the visualization will display the symbol for “unit combat armor, hostile” together with the size indicator for platoon, cf. figure 3b, instead of the symbol for “equipment, armoured tank, hostile” and the quantity indication (“3”), cf. figure 3a.

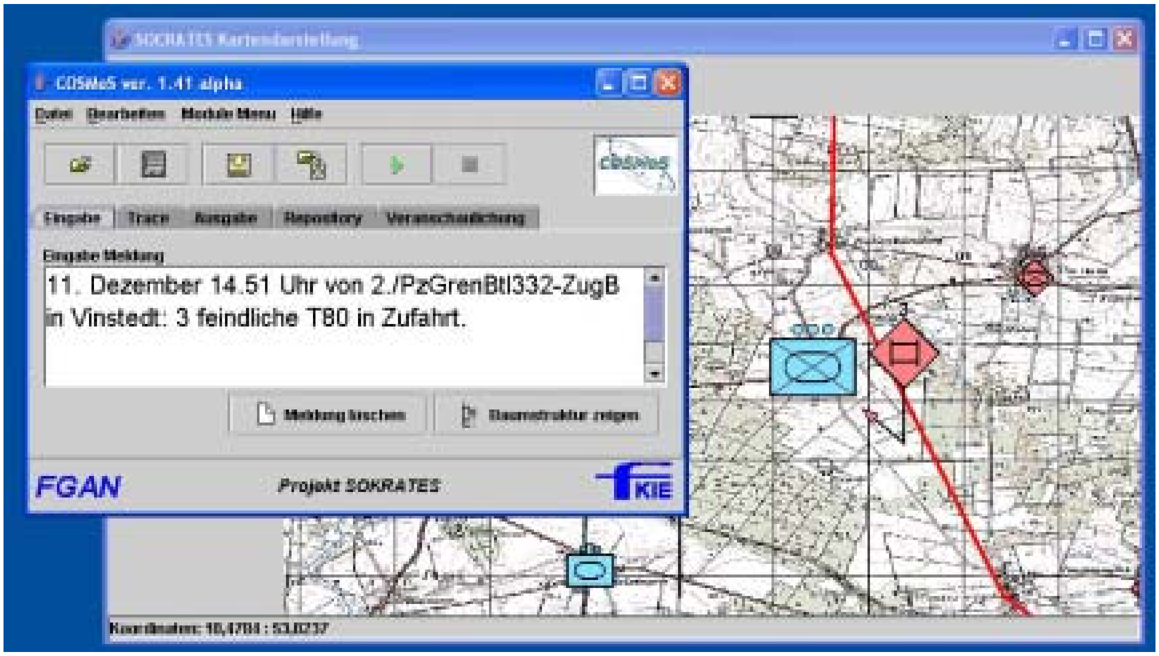

Figure 3a: Sokrates sub-modules for input and visualization. The map displays the analyzed information in the case in which the unit determination process had been inactivated and, thus, the symbol for battle tanks is displayed.

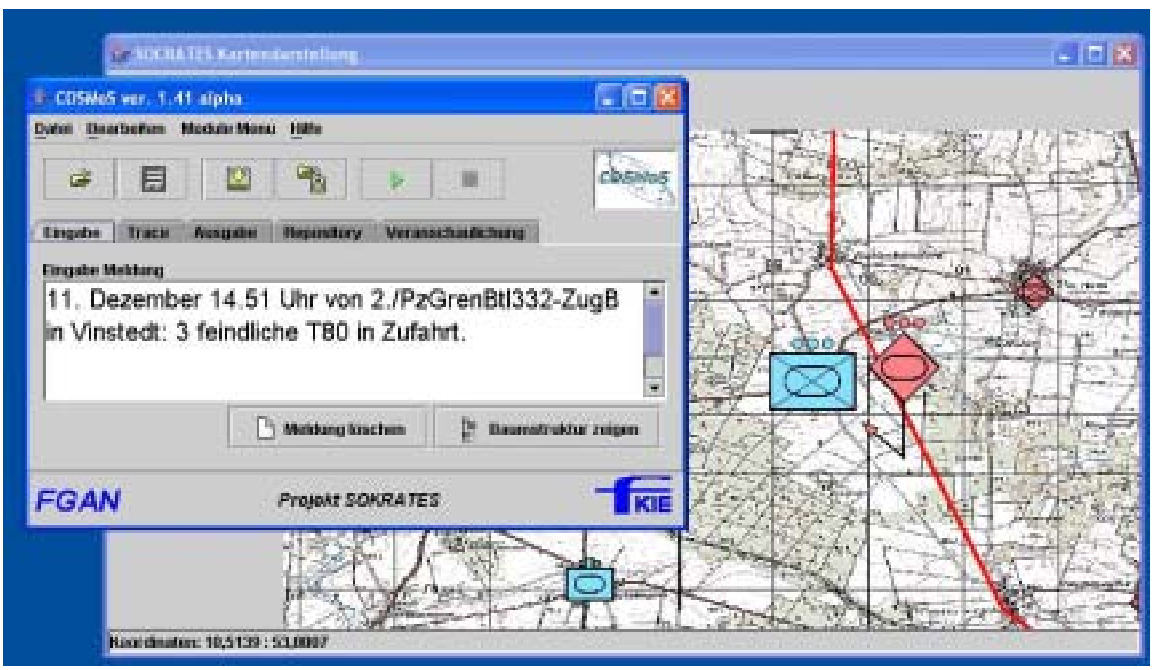

Figure 3b: In contrast to figure 3a, the unit determination process had been active.

With respect to the unit determination process, it should be mentioned that this process is facultative. The system's operator is allowed to activate or inactivate processes marked as facultative as required. In general, facultative processes automatize estimations which would be done more precisely by humans, but which sometimes have to be skipped due to a lack of time or resources. The existence of facultative processes within a system adds flexibility to the system. It can be adapted to the requirements and the resources on hand.

Post-Processing

In the prototype, post-processing means visualization of the report's content on a map as well as actualization of the underlying data base. With respect to visualization, all the units occurring in the augmented feature structure are displayed. Therefore, the ontological process had to calculated coordination for their locations. It should be mentioned that in some cases units are displayed which are not mentioned in the report, explicitly. E.g., if a barrier is mentioned ("*Drei feindliche BMP-2 vor Sperre 89 liegengeblieben*" – *Three hostile BMP-2 stopped by barrier 89*) there is a unit observing this barrier, and this unit will be added to the report's feature structure – as value of the "observer"-attribute – and, thus, displayed. In addition to displaying the report's content, the context is displayed. The context is predefined as a list of those objects (e.g., units, positions, barriers, control features) the operator would like to see. The list can be added during processing, and the objects will be displayed according to the most actual knowledge about their locations. Thus, the context represents those aspects of the operational picture the operator is interested in. In order to highlight the report's content, its symbols are displayed somewhat larger than the context's symbols.

The actualization of the underlying data base, a version of the C2IEDM, is done by a module called "fs2sql" which transforms the information kept in the report's feature structure into SQL-statements. The module is presented in detail in [15]. Like the visualization module, fs2sql is coupled to the kernel of the SOKRATES system in a way that it can be deactivated or run on a different computer, e.g., together with the data base. The modular architecture of the SOKRATES systems grants that the system can be shaped as required. This flexibility assures adaptability to the actual situation and its demands.

In future versions of SOKRATES system there will also be a post-processing generation module. Its task will be the preparation of reports or requests the operator might be willing to send as a consequence of the report processed. If, for example the report says, people are hurt then a request for rescuing will be prepared. The operator, then, may launch this automatically prepared request if and whenever it seems appropriate.

The Benefit

As has been already said in the introduction, IT systems to be used by coalition forces have to meet demands similar to those established with respect to the forces themselves. Under the perspective of "NCW" these demands are termed "agility" and "interoperability" (cf., [3]: 105). The proposed module for military report processing is a step toward an IT system which meets these demands.

With respect to the aspects of agility (robustness, resilience, responsiveness, flexibility, innovation, and adaptation; ibid: 128), adaptation was focused in the this paper because the ability of adaptation is especially important in coalition operations, ibid: 153. The

SOKRATES system contributes to the adaptiveness twofold. First, it is developed as a module to be plugged-in into a broader C2 Information system as required. In cases in which a huge amount of reports have to be processed, the SOKRATES system can be added to allow for automatic report processing. Second, the system itself is adaptable. It is composed of a kernel and sub-modules. For example, the unit determination process is such a sub-module which can be added as desired. Obviously, the modularity of the system's architecture also counts with respect to other aspects of agility. For example, the module can be distributed in order to improve its resilience.

Interoperability is the key attribute for coalition operations. As has already been stated, it is not enough that two forces are connected physically. They also have to share information, contribute to a common operational picture, and build up a shared awareness in order to collaborate. The SOKRATES system contributes to these aspects in many ways. First, it transforms natural language reports into C2IEDM entries such that the report's information become shareable according to the MIP standard. Second, it transforms the information into a map representation using APP-6A's military symbols. By this, the system makes a contribution to the common operational picture and provides a basis for shared awareness. Third, it augments the information and thus enhances its quality. On the one hand, this reduces the chance for misinterpretations which otherwise may result from ambiguity and incompleteness. On the other hand, this compensates for some of the "semantic loss" that necessarily occurs as soon as information is exchanged among systems and keeps the loss "acceptable" with respect to the operational demands, cf. [16].

References

[1] Network Centric Warfare. Department of Defense Report to Congress, July 2001.

[2] Alberts, D.S. (2002). *Information Age Transformation*. Washington, DC: CCRP.

[3] Alberts, D.S. & Hayes, R.E. (2003). *Power to the Edge*. Washington, DC: CCRP.

[4] Brendel, P., Eiselen, K. & Kuhl, H. (2003). Gefechtsfeldführungssystem FAUST. *Soldat und Technik, 7/03*, 33-37.

[5] Appelt, D. & Israel, D. (1999). *Introduction to Information Extraction Technology*. Stockholm: IJCAI-99 Tutorial, http://www.ai.sri.com/~appelt/ie-tutorial/.

[6] Neumann, G. (2003). http://www.dfki.de/~neumann/pd-smes/pd-smes.html .

[7] Shieber, S.M. (1986). *An Introduction to Unification-Based Approaches to Grammar* (= Volume 4 of CSLI Lecture Notes Series). Stanford, CA: Center for the Study of Language and Information.

[8] Bresnan, J. (2001). *Lexical-Functional Syntax*. Oxford, UK: Blackwell Publishers.

[9] Hecking, M. (2003). Information Extraction from Battlefield Reports. *Proceedings of the 8th International C2 Research and Technology Symposium (ICCRTS)*. Washington, DC: CCRP.

[10] Multilateral Interoperability Programme (2003). *The C2 Information Exchange Data Model (C2IEDM)*, Working Paper 5-5, Version 6.1.

[11] MIP website, http://www.mip-site.org

[12] Schade, U. (2003). Towards an Ontology for Army Battle C2 Systems. *Proceedings of the 8th International Command and Control Research and Technology Symposium*, June 17-19, 2003. National Defense University, Washington, DC.

[13] Nunberg, G. (1979). The non-uniqueness of semantic solutions. *Linguistics and Philosophy*, *3*, 143-184.

[14] Fillmore, C. (1982). Towards a descriptive framework for spatial deixis. In: Jarvella, R.J. & Klein, W. (Eds.), *Speech, Place, and Action*. Chichester: John Wiley.

[15] Casals Elvira, X. (2004). Translation from semantically enriched linguistic representations to SQL statements for the Project SOKRATES. *FKIE-Bericht 76*. Wachtberg: Forschungsgesellschaft für Angewandte Naturwissenschaft e.V.

[16] Dorion, E. & Boury-Brisset, A.-C. (2004). Information Engineering in Support of Multilateral Joint Operational Interoperability. Proceedings of the "Command and Control Research and Technology Symposium: The Power of Information Age Concepts and Technologies." June 15-17, 2004. San Diego.