8th International Command and Control Research and Technology Symposium,
June 17-19, 2003
National Defense University
Washington, DC

<u>Topic</u>: C2 Experimentation

<u>Paper's Title</u>: Towards an Ontology for Army Battle C2 Systems

<u>Author</u>

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

Towards an Ontology for Army Battle C2 Systems

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

Abstract

The analysis of battle field reports by a system can result in an update of a corresponding data base and the actualization of the map of the battlefield. These kinds of automations speed up the C2 process and reduce the amount of data the command posts' staffs have to process themselves. Such a system should include an ontology component for semantic processing, especially if the reports to be analyzed are given in natural language. The ontology should be based on expert knowledge of the domain in question, the battle field. This knowledge is partly captured in military data models which therefore can be exploited for the ontology's development. However, the ontology has to have additional resources in order to meet its requirements, i.e. to tackle the problems of vagueness, ellipsis, and ambiguity. Additional knowledge must be represented, and algorithms for reasoning and association processes have to be included as well.

Introduction

According to Gruber [1], an ontology is an explicit specification of a shared conceptualisation. It represents knowledge, especially the knowledge human beings take for granted. But what goes without saying in communication among humans often has a significant impact on the interpretation of statements. Let us consider the example of a squad leader receiving the order "*Cunctator to all, position 3.1 advance! Over*." The squad leader knows that the squad is ordered to advance to the position coded in plan 3.1 to be the squad's one. A C2 system without ontology component however lacks the knowledge that in contrast to units positions do not move. Thus, it will infer on the base of syntactic analysis that it is position 3.1 that has to advance.

In this paper, a C2 system will be presented which extracts meaning out of reports communicated in natural language. Emphasis is placed on the system's ontology component. Therefore, after an outline of the C2 system's general architecture, the problems of vagueness, ellipsis, and ambiguity will be discussed which have to be tackled by the ontology if natural language reports are under analysis. With these problems in mind, those principles will be presented which apply to the construction of the ontology component. Finally, it will be discussed how the ontology addresses these problems.

The SOKRATES system

As the "position advancing" example indicates, any system which has to assign meaning to a natural language report must include an ontology component. The ontology presented in this talk is indeed part of such a system (the SOKRATES system) which is under development (for the developmental process of components of this system, cf. [2, 3, 4, 5, 6]). This system is intended to analyse battlefield communication, in particular reports which renege the restrictions of the ADat-P3 format [7] and use natural language instead.

Natural language in reports

Considering the existence of strictly specified report formats like ADat-P3, one might ask if and why natural language is still used in reports. There are at least two answers to this question. First, not everything which happens on a battlefield can be expressed adequately in strictly specified formats. Since formal formats must be developed and trained in advance they cover standard situations. However, it is an old truth that unexpected things happen as soon as battle starts. These things are not covered but nevertheless must be reported. So natural language has to stand in. Even more, the rate of natural language use increases with respect to non-conventional contexts, e.g. reports in the contexts of peace supporting and peace enforcing operations, as in these contexts the range of what might happen is even broader.

Second, there is a psychological advantage in using natural language. The gap between sender and addressee is felt to be narrower and the communication is more personal if the sender reports by the means of natural language instead by the means of clinically filled out prim forms [8].

Processing a report within the SOKRATES System

The SOKRATES system takes reports as input. In a first step it divides a report incrementally into sentence-like units ("propositions" in the terms of [9]). These "sentences" are parsed syntactically. As a result, a feature-value structure (the standard format used in unification based grammars, cf. [10]) is generated which specifies actions and their corresponding role entries. E.g., a "move" is specified by assigning the moving unit to the agent role and the respective locations to the roles "source location" and "target location". This feature-value structure is input to a semantic processor which revises the structure. As will be discussed in the following section, this revision relies on the ontology component. It is meant to reduce vagueness, to add missing information, and to resolve ambiguities. The revised structure is input to some output components. By these components, the map of the battlefield is actualised, the information in the data base is updated, and some text is generated which can be copied into own reports and requests (to be sent to superior head quarters). This process flow is illustrated in Figure 1.

The SOKRATES system might be integrated into a C2 system like FAUST ("Führungsausstattung, taktisch"). This is the German version of the Force XXI Battle Command Brigade and Below (FBCB2) system (cf. [11] for the German classification of C2 systems, and [12] for the US view). The meaning extracted from the reports by SOKRATES might then complement the information provided by GPS or other sensors. It will add to the shared common picture of the battle space and to increase the situational awareness for commanders, staffs, and soldiers even further.

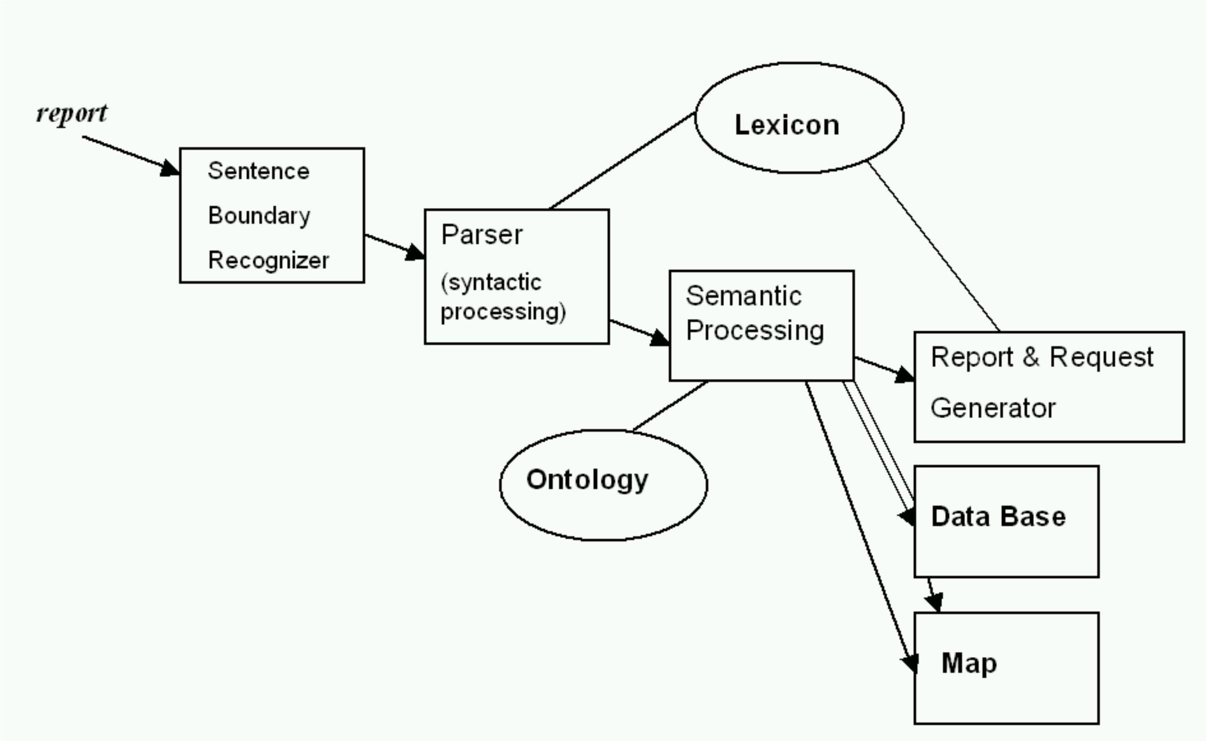

Figure 1: The figure shows the steps a report passes through while under processing within the SOKRATES system.

Problems to be tackled

The main tasks the ontology component has to perform in the SOKRATES system are to solve the problems of *vagueness*, of *ellipsis*, and of *ambiguity*. With respect to vagueness, the reported data may include information in a manner which is not explicit enough to process it adequately. If, for instance, the report says "*bei Meisenhain KPz in Stellung, links davon zwei weitere in Zufahrt*" (*at Meisenhain, a battle tank in position, two more approaching*), the affiliation of the battle tanks is not mentioned. However, the information that they are hostile ones, most probably can be inferred from the context. This inference is a process of completion, a process which handles the problem of ellipsis. Vagueness, however, is in the term "battle tank" itself. It might be crucial to know the specific types of the battle tanks, e.g., in order to estimate their cannons' ranges. If so, "battle tank" should be substituted by the specific tank type during the ontological process.

As the example above suggests, the problem of ellipsis is tackled by a completion process. Battlefield reports communicated in natural language are extremely elliptical. Thus, lots of completion has to take place. Whenever a component regarded as essential (Who, What, When etc.) is not mentioned explicitly in the report its referent has to be filled in. In some cases this seems to be an easy task. For example, if the report says "*Cunctator, hier B, Stellung bezogen*" (*Cunctator, this is B, deployment completed*), neither the agent (Who?) nor the target location (Where?) of the deployment is mentioned explicitly. The missing information can be looked up in the header of the report since the agent of the reported action is the reporting unit ("*B*"), and the action's location is the one this unit is in at the moment the report is given. However, this is not self-evident. If the report says "… *beziehe Wechselstellung*" (… *switching position*), again, the actor is the reporting unit but the location the unit is moving towards is not in the header. Instead, it has to be looked up in the data base where it is listed as the unit's evasion position.

Beside the problems caused by the elliptical nature of natural language reports, ambiguity might also be a problem. In artificial intelligence, ambiguity emanating from homonyms is a main problem [13]. With respect to battle field reports, however, this kind of ambiguity is negligible because the domain is restricted: "*Leopard*" will most probably refer to a [German] battle tank since big cats are rarely seen on modern battlefields. Nevertheless, ambiguity creeps in. Especially, prepositional phrases provide a source of ambiguity. A simple example may illustrate this. A report might say "…, *zwei Kampfpanzer vor Sperre 7 liegengeblieben*" (…, *two battle tanks stopped in front of barrier 7*). In order to catch the ambiguity, the prepositional phrase "*in front of barrier 7*" has to be examined carefully. Human beings often fail to notice an ambiguous term because they unconsciously resolve it while listening or reading by resorting to their outstanding expertise in language processing. However, there are at least three ways to interpret the phrase in question. For one, it may be construed deictically in a way that the speaker is the primary reference point and the barrier is the secondary one. The report then is interpreted in the sense of "From my point of view, the tank is in front of the barrier." If so, the tank would be somewhere in between the speaker and the barrier. For second, the primary reference point may be the addressee: "From your point of view, the tank is in front of the barrier." This also is a deictic interpretation meaning that the tank is somewhere between the addressee and the barrier. For third, the tank may be at the forefront of the barrier. This is the intrinsic interpretation. It does not depend on the position of the speaker or the position of the hearer. However, to allow for an intrinsic interpretation, the location taken as reference (the barrier) must have an intrinsic orientation, e.g. a forefront, in this case the side the enemy approaches to.

The examples given above illustrate that the information which is conveyed by natural language often is "hidden". Although, human beings have the means to catch the information usually, it has to be become explicit before systems can represent and process it. It is the task of the ontology to make the information an explicit one.

The construction of the ontology

An Ontology has to be built up according to the guidance of construction principles. Some of these guidelines are self-evident. In instance, an ontology should be internally coherent, it should be extensible, and it should be restricted to the domain in question (cf. [14] for a detailed discussion of these and other construction principles). Naturally, the ontology should also be constructed in an object oriented manner such that features and assigned values can be inherited among related representation classes.

In the following, two more principles are presented which result from the specific function, the ontology has to play in the SOKRATES system. First, the relation between the ontology and the ATCCIS data model [15] will be discussed. This data model is intended to store the core data needed to describe information to be exchanged between C2 systems with respect to the battlefield domain. Second, lexical processing is mentioned. This principle is to be considered whenever natural language is processed. The discussion of both principles will regard the task of the ontology within the system, in particular, the tackling of the aforementioned problems.

The role of the data model

The ontology executes a principle task in the SOKRATES system. It provides the semantic and pragmatic base to analyse natural language communication. In order to build up an ontology one has to identify the objects to be represented as well as their relevant features and the relations among them. With respect to battlefield communication one can rely on the ATCCIS data model in many cases. If, for example, the object "minefield" is included into the ontology, the data model recommends among others "type", "depth", "pattern", "persistence", "purpose", and "stopping power" as relevant features. It also suggests possible values for each of these features. As long as the data model's advocacies are adopted, the interoperability between the ontology on the one hand and a data base on the other hand is granted. Entities could be easily transformed from data base entries to ontology instances and vice versa. Therefore, it is a principle of the ontology's construction that objects, features and values should be taken from the data model whenever this is possible.

However, the world is not a perfect one, and the data model at least is not "complete". The data model is meant to contain the "minimum set of data to be exchanged" in a C2 environment [15, section 3.2.4]. Thus, it cannot be expected that it covers all objects and features an ontology component needs. With respect to the minefield example, within the ontology it should be expressed if a specific unit or a specific soldier monitors a specific minefield. In theory, this can be done by using data model's organization-facility association. However, it is not specified whether the "monitoring" relation has to be expressed by the organization-facility-association category "controls", "is responsible for", "provides sustainment for", or "uses". An even better example is the "forefront" feature that is not part of the data model but should be attached to each location which has an intrinsic orientation within the ontology. Otherwise, the disambiguation of prepositional phrases cannot be done properly because it would be unknown whether a prepositional phrase allows for intrinsic interpretation.

Lexical processing

Lexical processing means that lots of information needed for the processing of language is stored in lexical entries but not in "rules" which abstract from these entries [16]. With respect to the battlefield ontology, this principle is applied to the representation of "actions". According to the data model, action concepts are used to represent activity, e.g., the movement of a unit.

In natural language, verbs are used to verbalize actions, and each verb has its own frame. Verb frames describe the kind of objects the verb might take including restrictions specific to the verb. For instance, the verb "to succeed" needs a prepositional object whose preposition has to be "in" for completion ("the unit *succeeds in* holding its position"). In some sense, the same is true for concepts representing actions. They frame roles which must be filled. Some (battlefield) actions set up role slots for agent and location (e.g. "*to rest*"), some need an agent and a direction (e.g. "*to move*"), some ask for objects (e.g. "*to intercept*"). Even more the roles are restricted. Objects might be of affiliation opponent ("*to attack*") or of affiliation friend ("*to rescue*").

In the ontology, each represented action has features that specify the roles the action requires. Even more, the values these features can take are restricted action-specifically. The principle to include roles and their value restrictions within the representations of actions is the ontology's equivalent to lexical processing. It is exploited to tackle the problem of ellipsis in two ways. First, if an instance of an action is created to represent an action mentioned in the report, a value has to be assigned to all role features of this action instance. A role feature which does not receive a value reveals an elliptical gap in the report. By this way, the gaps are identified. Second, since the role features bear restrictions, these restrictions can be used to narrow down the set of possible values to be assigned to the unsealed feature.

The representation of knowledge and the implementation of reasoning

In order to bear down the problems of natural language reports caused by language features like ellipsis or ambiguous prepositional phrases, the ontology had been supplied with knowledge and reasoning tools. Some of this knowledge is domain specific and known only by experts. Thus, experts of the field, in our case military experts, have to be involved in the development of the ontology.

Some knowledge can be put into the ontology directly. For example, let the unit X be attached to unit Y. If this attachment is listed at the instance which represents X in the ontology (by making "Y" the value of X's feature "attached_to"), it will automatically be listed at the instance which represents Y as well (by adding "X" to the set which is the value of Y's feature "attached_units"). In addition, the attachment will be listed at the instance which represent the unit X belongs to (by adding "X" to the set which is the value of its feature "detached_units"). This portion of knowledge about attachment is "obvious". No military expert has to be questioned in order to derive it. It nevertheless has to be made explicit and therefore to be represented within the ontology. The military expert is needed to specify what kind of units will be attached to what other units under certain conditions. Obviously, this portion of knowledge about attachment cannot be represented that easily. However, some specifications can be formulated in order to restrict the assignment of unit instances to the "attached_units" feature of other unit representations. This restrictions then have to be complemented by the association of constraints in order to represent the expert knowledge adequately. In addition, there has to be a routine which checks for the constraint violations and generates warnings if necessary. E.g., if a brigade is attached to a squad, the size constraints of attachment are violated, and at least a warning should be given.

In addition to knowledge, directly implemented or represented by constraints, there have to be tools for reasoning to enable the ontology to tackle problems like the aforementioned ones. For example, spatial reasoning has to be utilized in order to resolve statements of location. In principle, the natural language descriptions have to be transformed into geographic coordinates in order to allow for precise mapping. Obviously, precision can be vitally important, e.g., for knowing the borders of a minefield or for coordinating fire support. In order to illustrate the ontology's spatial reasoning, let us assume that some military trucks are reported to be "*south of* (the village) *Lintzel moving westwards*". To calculated the trucks' position, a fan is constructed. The fan's inner angle is set to 45 degrees, and its center line originates in Lintzel and is aligned toward south. The fan's area is not allowed to cover another settlement of the same size or larger than Lintzel. Thus its border is set north of the village of Brambostel. Because trucks normally move on roads, the intersection of the fan area and the represented roads is calculated. If only one road intersects with the fan, this road probably is the road the trucks use. Otherwise an ambiguity results (cf. figure 2) which has to be resolved by competition (cf. below).

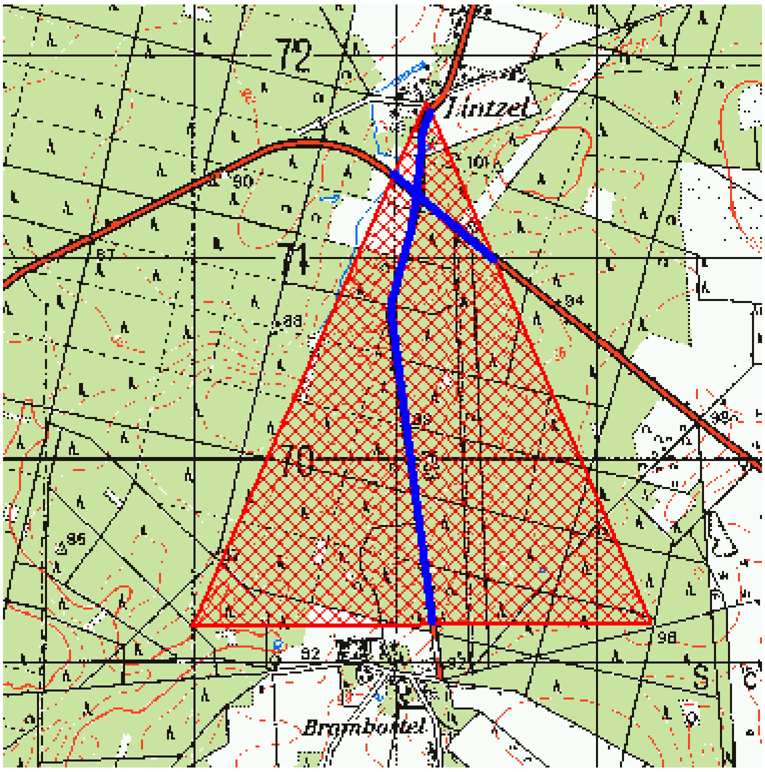

Figure 2: In order to resolve a report saying "*South of Lintzel, three trucks, moving westwards*", the intersection of a fan originating in Lintzel and roads is calculated. The blue line show the result.

The reasoning tools are complemented by a mechanism based on psycholinguistic knowledge about language comprehension using ideas proposed by Walter Kintsch [17, 18]. It interprets the objects of the ontology as well as their features as nodes of a spreading activation network [19]. Starting from those objects and those features which are explicitly mentioned in the report and which get high activation due to this mentioning, activation spreads through the ontology. According to the activation algorithm, a node sends activation to another node if the sender is activated and if there is a connection between both. Thus, objects activate their features, instances activate the concrete classes they belong to, and classes activate their super classes. Besides, all these connections are bi-directional. They convey excitatory activations. If nodes represent competing items, e.g. a class' sub-classes or the values of a feature, these nodes exchange inhibitory activation. The amount of activation sent correlates to the sending node's own activation value. Thus, highly activated nodes have a larger impact on the activation spreading than lowly activated ones.

After some cycles of activation spreading, a normalization steps in. Due to this normalization process, only those objects and features stay activated that had amassed higher than average activation. The resulting activation pattern then serves as basis for problem handling. With respect to vagueness, if a term is represented by a class which has a subclass which is activated after activation spreading and normalization, the term is substituted by the term the subclass represents. With respect to completion, the extant objects, features, and also the relations among them, are supposed to be likely candidates to fill the gaps. With respect to disambiguation, the competitor with higher activation is chosen. In the "trucks moving south of Lintzel"-example, the competition between the roads "B 71" and "K 33" is won by "B 71" since it runs (more or less) from west to east, whereas "K 33" runs (more or less) from north to south. "B 71" receives additional activation from its "direction" feature which had been activated by "*westwards*" directly from the report.

Conclusion

In the paper a system under development has been presented. It includes an ontology component for semantic processing. The examples given illustrate how reports can be analyzed by a system even if these reports include natural language. This analysis will result in updating of the data base, in actualizing the map and in providing parts of requests and own reports. Thus, the system promises a speed up of the C2 process as well as a reduction of the information which has to be processed by the command posts themselves.

References

[1] Gruber, T.R.: A translation approach to portable ontology specifications. *Knowledge Acquisition, 5 (1993)*, 199-220.

[2] Hecking, M.: *Natural Language Access for C2 Systems*. Paper presented at the RTO IST Symposium on „Information Management Challenges in Achieving Coalition Interoperability", held in Quebec, Canada, 28-30 May 2001, and published in RTO MP-064.

[3] Hecking, M.: *Analysis of Spoken Input to C2 Systems*. In: Proceedings of the 7th International C2 Research and Technology Symposium (ICCRTS), Quebec, Canada, 2002.

[4] Casals-Elvira, X.: Mapping of logical formulae to SQL-statements for a natural language access component. *FKIE-Bericht 52*. Wachtberg: Forschungsgesellschaft für Angewandte Naturwissenschaften e.V. (FGAN), 2002.

[5] Frey, M.: The role of data representation in sentence boundary disambiguation with neural networks. *FKIE-Bericht 46*. Wachtberg: Forschungsgesellschaft für Angewandte Naturwissenschaften e.V. (FGAN), 2002.

[6] Schade, U.: Ontologieentwicklung für Heeresanwendungen. *FKIE-Bericht 57*. Wachtberg: Forschungsgesellschaft für Angewandte Naturwissenschaften e.V. (FGAN), 2003.

[7] ADatP-3 Baseline 11.0.1 (CD-ROM). Brüssel, Belgien: HQ-NATO C3S-IOB, 1999.

[8] Bloom, J.N. & Farber, A.M.: *Art and Requirements of Command (ARC). Volume 1: Summary Report*. Philadelphia, PA: The Franklin Institute Research Laboratories, 1967.

[9] Clark, H.H. & Clark, E.: *Psychology and Language: An Introduction to Psycholinguistics*. New York: Harcourt Brace, 1977.

[10] Shieber, S.M.: *An Introduction to Unification Based Approaches to Grammar*, CSLI Lecture Notes Series, Number 4. Stanford, CA: CSLI, 1986.

[11] Hodouschek, T.: Eine leistungsfähige Datenverarbeitung: Führungsinformationssysteme des Heeres – Sachstand und Planungen bis 2010. *Truppenpraxis / Wehrausbildung 44 (2000)*, 193-197.

[12] Stanley, E.A.: *Evolutionary technology in the current revolution in military affairs: The army tactical command and control system*. Carlisle, PA: Strategic Studies Institute, U.S. Army War College, 1998.

[13] Stevenson, M.: *Word Sense Disambiguation*: *The Case for Combinations of Knowledge Sources*. Stanford, CA: CSLI, 2003.

[14] Uschold, M. & Gruninger, M.: Ontologies: Principles, methods, and applications. *Knowledge Engineering Review 11* (1996), 93-155.

[15] *ATCCIS WP 5-7, Edition 5.0: Overview of the Land C2 Information Exchange Data Model (LC2IEDM)*. NATO: ATCCIS PWG, 2002.

[16] Bresnan J.: *Lexical Functional Grammar*. Oxford, UK: Blackwell Publishers, 2001.

[17] Kintsch, W.: The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review 95 (1998)*, 163-182.

[18] Kintsch, W.: *Comprehension*. Cambridge, UK: Cambridge University Press, 1998.

[19] Rumelhart, D.E. & McClelland, J.L.: An interactive activation model of context effects in letter perception: Part 1: An account of basic findings. *Psychological Review 88* (1981), 375-407.