Understanding the Role of Context in the Interpretation of Complex Battlespace Intelligence

Gerald M. Powell
U.S. Army RDECOM CERDEC I2WD
Fort Monmouth, NJ, U.S.A
gerald.m.powell@us.army.mil

Christopher J. Matheus
Versatile Information Systems, Inc.
Framingham, MA, U.S.A
cmatheus@vistology.com

Mieczyslaw M. Kokar
Northeastern University
Boston, MA, U.S.A
mkokar@ece.neu.edu

David Lorenz
EWA Government Systems, Inc.
Herndon, VA, U.S.A.
david.lorenz@us.army.mil

Abstract – A key challenge presented by the increasing volume and complexity of information collectable from battlespace situations is the development of intelligent applications able to automatically analyze the information and identify critical enemy activities in a timely manner. What makes this information fusion problem particularly difficult is the strong contextual dependency of the interpretation of complex battlespace information. This paper examines the dimensions of intelligence information, establishes a framework for exploring the effects of contextual information and provides an illustration of its use. The paper concludes with a discussion of the ramifications of context on the development of intelligent applications intended to aid in the interpretation of complex battlespace intelligence.

Keywords: Automated analysis, battlespace intelligence, context dependency, METT-TC, multi-source fusion.

1 Introduction

Today the stream of battlespace intelligence information is frequently more than can be handled by available human analysts. The future battlespace will be characterized by an overwhelming volume of information collected from a vast networked array of increasingly more sophisticated sensors and technologically equipped troops. Maintaining information supremacy will require an increasing dependence on automated processes that can intelligently analyze and make at least partial sense of what the enemy is doing and predict how it might threaten friendly forces/resources now and in the future.

Efforts to develop intelligent applications for the automated analysis of battlespace intelligence have been underway for years but they have mostly focused on level 1. It was clear from the 2005 Fusion Conference Invited Panel on “Issues and Challenges of Knowledge Representation and Reasoning Methods in Situation Assessment (Level 2 Fusion)" that there remains a significant need for higher levels of information fusion such as those required for generic situation awareness, prediction of enemy courses of action ECOA and threat assessment. In general, effective processes for automated analysis at these higher levels tend to be highly-specific point solutions based upon simplifying assumptions that render them brittle and inapplicable to novel situations.

An earlier paper [1] introduced the need to recognize, and address, the problem of context-dependence in any attempt to perform battlespace analysis and interpretation; the particular focus was answering priority intelligence requirements. The present paper goes significantly beyond [1]. The thesis of this paper is that a major factor contributing to the difficulty of interpreting battlespace intelligence is the significant influence of *contextual factors* and the fact that you need to account for all of them in unison. By contextual factors we mean 1) information in a report, 2) information in a set of reports, 2) general background knowledge e.g., doctrine, techniques, practices) plus 4) known situation-specific information (e.g., mission, terrain, weather) that human analysts incorporate into their interpretation of both the meaning and relative importance of battlespace intelligence obtained during the course of an engagement.

We view this problem of interpreting battlespace intelligence as analogous to the problem of understanding and participating in natural language discourse (see [2]). To be an active participant in a two-way conversation requires not only a wealth of background knowledge pertaining to the topic of discussion and the particulars of where you are and how you came to this conversation, but also the ability to develop plausible hypotheses about the intentions of the other person. Furthermore, you must be able to adapt your responses as the conversation evolves and your hypotheses change, and even the topic might meld into something new. We will return to this analogy in Section 11 where we consider the ramifications of context on the development of intelligent systems.

This paper investigates the problem of interpreting battlespace intelligence and the way it is influenced by contextual factors. We begin by describing the challenge of intelligence analysis and propose a formula for it that elucidates its constituent parts which we decompose into smaller, more easily identifiable elements. We then introduce a *context influence matrix* that provides a framework for the further exploration of how contextual factors can influence intelligence report interpretation. We follow this with a concrete example to show how the interpretation of an intelligence report is highly dependent upon contextual factors. In our concluding sections we discuss some important ramifications for the development of automated applications intended for the higher-level analysis of complex battlespace intelligence and consider some related work.

2 Intelligence Analysis

The challenge at the heart of this paper is analyzing large volumes of intelligence reports in real-time as they come in from the battlespace with the intent of figuring out what the enemy is doing and what type of threat such activities might represent. This is currently and primarily a human task conducted by intelligence analysts; but as the volume of battlespace data continues to increase it has become imperative that automated solutions to at least portions of the problem must be developed. Our approach has been to study doctrine and pedagogy associated with the performance of intelligence analysis, but we have also placed a heavy emphasis on trying to understand how analysis is actually carried out. We believe an understanding of how it is practiced provides a useful basis for identifying tasks that represent significant cognitive challenges, and, consequently, potential opportunities for automated support.

The report analysis process may be approached from two slightly different angles. In the first case it is highly focused on answering specific questions of critical importance to the commanding officer; a priority intelligence requirement (PIR) of this sort, for example, might be to answer the question, "Will the enemy attack the bridge with artillery fire in the next 24 hours?" In this case specific ISR assets will have been strategically placed with orders to collect information that will one way or another answer the question. In the second case, the analysis approach is more open ended with the intent of answering the more general question "what is the enemy doing?" in a process referred to as situation development. In this latter case the analyst is as interested in the wide breadth of intelligence derived from all information sources, not just those allocated to answer specific questions.

In either case, PIR answering or situation development, the primary challenge remains the same: how to quickly analyze and interpret massive volumes of intelligence reports and select all of those that are relevant, credible and require further attention.

3 Intelligence Analysis Formula

As intelligence reports R_n come in from the battlespace, the information it contains needs to be analyzed in the context of 1) the report itself R_n, 2) the collection of reports already processed (e.g., $\{R_1, R_2, \dots R_{n-1}\}$), 3) a body of static background knowledge B and 4) situation-specific (possibly time-dependent) information S. The intended outcome of this analysis is an interpretation I of what the enemy might be doing and what sort of a threat such actions represent. The analysis process is summarized by the following formula in which the arrow represents "implies": $(R, B, S) \Rightarrow I$

Reports R are derived from various forms of physical sensors as well as by direct human observations. Background information is the body of knowledge developed through training and experience about friendly and enemy characteristic modes of operation. Situation specific information concerns current information about specifics of the situation being confronted. Each of these sources of input to the analysis formula is decomposed into its constituents in subsequent sections respectively.

An interpretation I can be viewed as a mapping that assigns to each plausible ECOA hypothesis a specific threat T: $I(ECOA) = T$

An ECOA may be based on a single report or on multiple fused reports. ECOAs may overlap with one another (e.g., unit E_1 is performing reconnaissance, units E_1-E_2-E_3 are performing an approach march), or they may compete with one another (e.g., unit E is retreating, unit E is setting up an ambush) and they may appear at various levels of abstraction. The associated threat T may be a relative quantitative measure.

4 Intelligence Report Dimensions

We now turn attention to the dimensions or attribute of intelligence report data. We start with the dimensions defined by the SALUTE format for spot reports [3]: **Size**: the number of observed enemy soldiers and vehicles, which can be equated with echelon level (e.g., squad, platoon, company); **Activity**: what the units were doing (e.g., emplacing mines in the road); **Location**: the location of the observed units in terms of latitude/longitude or UMT; **Unit**: either the specific unit designation (e.g., 1st Plt, B Co, 3rd Bn, 1st Inf.) or its unit function type (e.g., reconnaissance, supply or combat); **Time**: the time of the observation; **Equipment**: a list of all the observed equipment the enemy is wearing or using (e.g., tracked vehicles, protective masks)

SALUTE format is intended for spot reports by troops in the field but the basic dimensions can also be applied to sensor reports, although all dimensions may not be present in all reports. For example, an unattended ground sensor (UGS) provides information about the dimensions of size, activity (movement speed and direction), location, time and simple equipment type (e.g., tracked vehicle, wheeled vehicle or human) but is unable to provide specifics of the unit or unit type.

In our experience, the activity of **movement** is important enough to separate it into a dimension of its own in terms of a vector defining direction and speed. One reason for this distinction is that direction and speed are often critical components in determining ECOAs and their possible threat. Another reason is that a unit can be moving while also performing some other activity, such as firing or putting on protective masks. We will refer to this augmented set of fields as SALUTE+M.

In addition to the SALUTE-based dimensions two additional higher level dimensions are important: report accuracy and report credibility. Report **accuracy** and **credibility** represent meta-data about a report that indicates how much certainty or confidence one should put on the reported information. For certain types of sensors, known sensitivity information (e.g., detection range, measurement uncertainty) can be used to help judge report accuracy (e.g., error ellipse} and reliability measures (e.g., MTBF) can be used to help in assessing report credibility. Note however that these sensitivity and credibility measures are context dependent, particularly with regards to terrain and weather conditions. For reports involving human judgment, accuracy and credibility can be more subjective and will depend upon the source's experience, training, disposition, what the source has been told to look for, the knowledge the source has of how the information may be used, as well as other contextual factors (e.g. current weather conditions). For example, a report from a novice soldier on his first day in the field during a torrential downpour stating that he has spotted the entire Red army bearing down on him is not likely to be accurate or credible.

5 Local and Global Report Context

Assume for a moment that the only information we have about a battlefield situation is a collection of intelligence reports R structured according to the SALUTE format. Without even considering background knowledge B or any situation specific information S there are two types of contextual influences that can affect the interpretation of aspects of a report. The first is called Local Report Context and refers to the influence that knowledge of the value of one field within a single report can have on the interpretation of the value of another field within the same report. The second type is called Global Report Context and refers to the influence that the processing of previous reports might have on the interpretation of the fields of the current report. This distinction is analogous to one made in linguistics in comparing context issues that arise within a sentence and those that arise from the context of surrounding sentences [2].

To illustrate how these inter- and intra-report influences can occur we will explore how each of the fields in a SALUTE+M report can singularly affect (or not) the interpretation of other fields in the same report. We call these effects single-order because they take into account how a single field by itself might affect other fields. At the end of this illustration we will consider some examples of higher-order influences resulting from local knowledge of more than one field. The claims made in this illustration are based on conversations with subject matter experts and are illustrative rather than definitive.

We begin with the ***activity*** field (as it is the most influential) and ask the question, "How can our knowledge of the observed activity affect our interpretation of each of the other five report fields?" Let us consider each in turn and identify with a specific example those cases where knowing the activity (with certainty) can have an influence:

Size/count: Perhaps. Certain activities (e.g., building a bridge, moving in a convoy) require more than one entity and so one could argue that knowing that the observed enemy is involved in such an activity that its count must be greater than one. Some activities imply that a certain rough order of magnitude of units are involved.

Location: No. Knowing the activity alone cannot influence our interpretation of the raw coordinates defining the unit's location. If we knew something about the terrain (see $S^{t}_{T/W}$ below) in the specified location then we might be able to infer something from knowing the activity (e.g., "constructing a bridge" would imply the unit is at a river) but right now we are only considering "Local Report Context" influences.

Unit designation/function: Yes. If the observed activity is "firing artillery" it is safe to assume the unit's function is fire support.

Time: Yes. Certain activities are less likely in daylight (e.g., burying mines in open areas).

Equipment: Yes. Certain activities require specific types of equipment (e.g., "refueling", "hauling supplies", "preparing artillery). This may allow you to infer the presence of equipment not directly observed/reported.

Movement: Yes. If the observed activity is "emplacing mines in road" then the units certainly are not moving at moderate or high speed in any specific direction.

Space does not permit us to fully illustrate the possible single-order effects of the other SALUTE+M fields but we have summarized them in Table 1. Note that this table is symmetric indicating that the influence among fields is (or at least appears) to be symmetric.

Table 1. Single-order Intra-Report Field Influences

	Influenced Value						
Known Value	Size	Act.	Loc.	Unit	Time	Equip.	Move
Size		?	?	Yes	No	No	No
Activity	?		No	Yes	Yes	Yes	Yes
Location	?	No		No	No	No	No
Unit	Yes	Yes	No		No	Yes	Yes
Time	No	Yes	No	No		No	No
Equipment	No	Yes	No	Yes	No		Yes
Movement	No	Yes	No	Yes	No	Yes	

One reason for walking through this illustration was to show the complex, involved nature of the context problem when one merely considers single-order information local to a single report. And this is only the very beginning; higher-order effects involving the combination of two or more fields are not only possible, they are prevalent. Consider for example that we know the time is the middle of the night and we know that enemy units have been observed moving at high speed (e.g., over 30 kph). It would then be possible to infer that the observed units are equipped with night-vision capabilities. Note that this inference requires knowledge of both time and movement; either one separately tells us nothing about whether they are so equipped.

The context provided by knowledge of other reports must also be considered when interpreting a specific report. Fortunately such information can often help to fill in missing aspects of a report or increase ones confidence in some field value if it is at least consistent with, and possibly supportive of previous reports. Of course a new report may contradict previous reports, which can lead to the need to support multiple plausible interpretations (see Ramifications section below).

6 Background Knowledge

Background knowledge is the body of relatively static knowledge about the general doctrine, tactics, techniques & procedures and equipment type capabilities of all entities that might be involved in the situation. We separate blue doctrine, techniques & practices from those of the red force and denote them as D_{Blue} and D_{Red}, respectively. These doctrinal knowledge sets represent the known or expected ways that the two forces are organized and have been trained to operate under abstract, idealized conditions. D_{Red} would contain, for example, definitions of the way that the Red force would conduct a Battalion Approach March, with specifics about where subunits would be positioned relative to one another and how they would move in unison. The specifics of what is in D_{Blue} and D_{Red} would depend upon which specific armies were defined as the Blue and Red forces for the current situation; the contents of D_{Blue} and D_{Red} would stay the same for the duration of the situation but over longer periods of time they would evolve as our understanding of the corresponding forces evolved.

The nature of the equipment that is involved in a situation can also be viewed as relatively-static background information which we will denote as E_{Blue} and E_{Red} in a similar fashion to the use of D_{Blue} and D_{Redd} above. For equipment however it seems to also make sense to have a designation for civilian equipment E_{Civil} which is increasingly being employed in asymmetric battlespace situations. The content of this knowledge about equipment would include items such as its type (e.g., wheeled vehicle, acoustic sensor, RPG), its potential usage (e.g., explosive device, ISR sensor, river bridging), important operating characteristics (e.g., fire rate, MTBF, cargo capacity), etc. Note that this information is not about the specifics of what equipment is present in the situation (that would be situation-specific information) but is instead concerned with describing the capabilities and characteristics of types of equipment that may appear in a wide variety of situations.

We represent all contributions of background knowledge as the mapping of D and E into B:

$$\nabla(D_{Blue}, D_{Red}, E_{Blue}, E_{Red}, E_{Civil}) = B$$

7 Situation-Specific Information

Situation-specific information is, as the name implies, information that is specific to the current situation. A useful construct for understanding situation-specific information is the METT-TC set of factors [4]. In the brief descriptions of these factors (that follows) we note that each of them has the potential to change during the course of a situation:

Mission: the mission of the Blue force $S^t_{Mission}$. Missions almost invariably change to some extent during the course of their execution.

Enemy: analysis of the Red force, including current information about its strength, location, activity and capabilities S^t_{Enemy}. As the situation progresses enemy units may become disabled or new, unexpected units may appear and our awareness of their location and activity will likely change.

Terrain/Weather: the specific terrain features for the area encompassed by the situation as well as the weather that is observed and predicted for that area $S^t_{T/W}$. Both terrain and weather can change during the course of operations; terrain, for example can change as a result of weather (e.g., dirt surfaces becoming muddy, a paved surface becoming slippery) or as a side effect of the battle itself (e.g., routine damage due to vehicle use or the effects of munitions).

Troops: the composition and disposition of the Blue troops including their training, experience and current conditions S^t_{Troops}. This also defines the equipment they have available to them as well as intelligence, surveillance and reconnaissance (ISR) assets, both of which can change over the course of a situation.

Time: time available for the Blue mission S^t_{Time}. As time goes by the amount available decreases (unless the mission changes). It is also possible for the time available to be reduced due to enemy activities S^t_{Enemy}.

Civilian Considerations: relevant considerations pertaining to the civilian population in the area in and around the situation (S^t_{Civil}). Such considerations might change if the enemy or civilians act other than expected.

Taken collectively, these six METT-TC factors contribute to the situation-specific information that needs to be considered during the analysis of intelligence report data, as follows:

$$\nabla(S^t_{Mission}, S^t_{Enemy}, S^t_{T/W}, S^t_{Troops}, S^t_{Time}, S^t_{Civil}) = S$$

8 The Context Influence Matrix

Our objective is to figure out how to interpret intelligence report data in the context of background knowledge and situation-specific information. To assist in this process we offer the Context Influence Matrix shown in Table 2. Along the row headings are the contextual factors defined by the background knowledge and situation-specific information. Along the column headings are the symbols used to represent the intelligence report dimensions. The content of each cell of this matrix is meant to answer the question: "How can knowledge of this specific contextual factor affect the interpretation of the intelligence report dimension?"

Table 2. Context Influence Matrix

		Intelligence Report Dimension								
Context Factor	Symbol	Time	Size	Unit Type	Location	Movement	Activity	Equipment	Accuracy	Credibility
Mission	$S_{Mission}$									
Enemy	S_{Enemy}									
Terrain/Weather	$S_{T/W}$									
Troops	S_{Troops}									
Time	S_{Time}									
Civilian	S_{Civil}									
Blue Doctrine+	D_{Blue}									
Red Doctrine+	D_{Red}									
Blue Equipment	E_{Blue}									
Red Equipment	E_{Red}									
Civil Equipment	E_{Civil}									

For example, the cell indexed by the contextual factor S^t_{Enemy} and the report dimension "Unit Type" corresponds to the answer to the question: "How does knowledge of the enemy's strength (entities and readiness), locations, activities, capabilities, most likely ECOAs, and potential and/or actual threats influence the interpretation of the unit type information provided by an intelligence report?" Consider the following example that illustrates how one might answer this question in a specific situation. If an intelligence report identifies the type of the observed units as "tracked vehicles" and we have situation-specific information about the enemy's expected make-up S^t_{Enemy} that states that they have only one tank battalion TB and no other tracked vehicle units, then (lacking other contradictory information) we will be inclined to conclude that the unit described in this report is very likely to be tank battalion TB.

At present, this matrix is not intended to provide many answers as to how contextual factors influence the interpretation of intelligence reports. Rather it is meant to provide a framework from which to explore the possible influences and determine how they will need to be incorporated into automated applications that will assist in the process of report analysis. The matrix does contain a bit of insight in terms of how the Time dimension of reports might be influenced (or not) by contextual factors. Recall that Time is the specific time of observation that is provided by the intelligence report. This dimension is not (in any reasonable way) affected by most of the contextual factors, and therefore the cells for most of these are hatched out. There are two cells that remain open to consideration. The first is the influence that knowledge of blue troops might have. More specifically we have in mind the knowledge one might have about the troops responsible for the report and the possibility that we have learned that these particular troops consistently provide inaccurate or imprecise time information. Similarly, it might be the case that we have learned that a particular sensor has a bad clock and that the time stamps it puts on its reports are always off by 30 minutes.

We believe the matrix holds potential for aiding in the further analysis of the complexities involved in incorporating contextual factors into the intelligence report interpretation process. There is however a limitation that needs acknowledgment. In the initial formulas for the intelligence analysis process provided in Section 3 we used the union symbol $\cup$ to indicate that each element can be defined separately and that it is the collective union of all the knowledge, information and data that contributes to the analysis process. When, however, we consider more fully the process of combining this collection of elements, things become considerably more complicated and the union symbol is more appropriately replaced by a symbol representing the fusion of the elements. In [5] the symbol ∇ was used for this purpose; it denotes the combination or fusion of theories (i.e., collections of knowledge, information and data) in such a way as to preserve semantic consistency. This becomes important, for example, when aspects of one piece of contextual information (e.g., the location of the center of a cold front in $S^t_{T/W}$) needs to be equated with aspects from another factor (e.g., the location of a Blue objective in $S^t_{Mission}$). Since the same information might be represented in different ways across contextual factors there must be a way to semantically combine them. In category theory [6] this sort of combination of theories is accomplished using the colimit operation.

9 A Concrete Example

We will now consider a concrete example of how various contextual factors come into play in interpreting a specific intelligence report. Assume that you are functioning as intelligence officer for a mechanized infantry brigade with the task of analyzing real-time battlespace intelligence as it comes in from the field with the intent of determining the enemy's course of action and possible threat. At time T you receive a report from an unattended ground sensor (UGS) capable of providing unit count, location, speed/direction and simple type information (i.e., tracked vehicle, wheeled vehicle, or human). The report identifies three wheeled vehicles at location L moving due east at 25 kph. How do you assess the possible threat posed by the vehicles identified in this report? Without additional information beyond what's in the report there is very little that you can say about the posed threat or even if there is one. Without knowledge of the weather conditions $S^t_{T/W}$ and their potential impact on the operation of the UGS E_{Blue} you cannot even be sure of the report's accuracy or credibility; for example, how might a torrential rain or several feet of snow $S^t_{T/W}$ alter your confidence in the supposedly observed units being wheeled as opposed to tracked vehicles, or even vehicles at all?

Since the UGS has no way of detecting a moving object's affiliation you cannot reliably assess the probability that the observed vehicles are enemy units without additional information. If you had knowledge about the location of blue forces S^t_{Troops} and possible civilian presence S^t_{Civil} you could then reason about the likelihood that the observed vehicles are in fact enemy units. Assume, for example, that you have knowledge that no friendly troops are anywhere near the sensor and that all civilians had been relocated out of the area. If you have no other reason to doubt the reliability of the sensor, your confidence in there being enemy units at location L would be fairly high (although you might still entertain the possibility that all civilians were in fact not evacuated or that some blue troops are other than where they are supposed to be). If on the other hand, you knew that civilians might be in the area S^t_{Civil} you would have to entertain the prospect that the moving vehicles are possibly enemy units or possibly civilians.

Now assume that you have some terrain information $S^t_{T/W}$ and know that the UGS that provided the report was placed along a narrow wooded trail that will not permit the passage of large vehicles, wheeled or tracked. If you trust your sensor and terrain information then you may deduce that the detected wheeled vehicles are small, perhaps motorcycles or powered scooters (E_{Red}+ E_{Civil}). Are these more likely to be enemy or civilian? That depends in part on the capabilities of the enemy forces you're facing and whether local civilians use motorcycles/scooters. Does the enemy order of battle include motorcycle equipped units S^t_{Enemy}? If not, does the enemy doctrine or their known patterns and techniques D_{Red} suggest that they might be capable of exploiting civilian equipment in innovative ways (e.g., asymmetric behavior)?

Assume at this point that we cannot rule out the possibility that the observed objects are in fact motorcycle mounted enemy units. You then must answer the question of what threat might they pose if they are enemy units moving at 25 kph to the east of point L? The answer is highly dependent upon the blue mission $S^t_{Mission}$ and perhaps also on the makeup and disposition of the blue troops S^t_{Troops} as well as the time available S^t_{Time}. If the blue mission $S^t_{Mission}$ is to hold and defend bridge B for the next two hours S^t_{Time} and B is located 50 km to the west of point L, then the likely threat from the motorcycled units is probably negligible even if they are heavily armed enemy units. If on the other hand the units are headed towards the bridge and could come within striking distance of it or our defending troops then the threat is potentially significant and deserves further attention and possible action.

Assume we do not yet know the identity or intent of the moving objects but if they are enemy units and continue on their current course then they could – based on our knowledge of the terrain/weather $S^t_{T/W}$ and their demonstrated mobility – pose a real threat to the current mission, particularly if they are possibly equipped with heavy mortars or multiple RPG ($S^t_{Enemy} + R_{Enemy}$). Do you immediately inform your commanding officer or do you wait until you can at least verify their identity as either civilians or enemy units. The answer depends upon additional contextual factors, in particular whether or not you have surveillance assets or troops S^t_{Troops} that could attempt to observe the moving objects within a reasonable amount of time relative to the mission $S^t_{Mission}$ and remaining time available S^t_{Time}. For example, if you knew they where headed for a clearing with no trees $S^t_{T/W}$, that they would be there in a few minutes and that you had a UAV that could immediately identify them, you would likely wait for that confirmation rather than risk the chance of needlessly notifying your superior about the movement of a few civilian motorcyclists.

This example was rather simple; it focused on a single type of sensor, one intelligence report and a rather trivial scenario. Reasoning involved in the fusing of intelligence from multiple sources over successive periods of time will clearly require a more complex interpretation process in order to appropriately incorporate contextual factors. Our example also did not take into account the fact that the situational specific contextual factors (e.g., the METT-TC factors) are themselves time dependent, meaning that they can change during the operation. Another important aspect ignored by this simple example is the relative ranking of the threats posed by multiple competing reports; if our UGS report is competing with reports pertaining to the preparation of in-range artillery batteries or the imminent approach of a tank battalion it might fail to exceed the threshold that would warrant paying it further attention.

An even more challenging problem comes about in the interpretation of contextual information in cases where combinations of two or more factors interact in complex (i.e., non-additive) ways. For example it might be the case that a specific Red force will conduct a specific type of operation in one way if the terrain is mountainous, another way if it is snowing and yet another way if it is mountainous and snowing. This implies that all contextual information needs to be considered in whole rather than as an isolated problem that can be handled separately. This has important ramifications when considering the development of automated systems.

10 Context Integration

The examples shown in Section 9 seem quite natural and relatively easy for humans to process, provided the amount of contextual information is relatively low. The process involves various operations that humans perform subconsciously, without explicating various processing details. For instance, in almost all of the examples of Section 9, there was an implicit assumption that all of the sentences were referring to the same "location", in both the meaning and the value. In many cases, however, it is not so obvious that we are actually referring to exactly the same meaning of "location". In natural language, we often use this term in various different meanings such as: XY coordinates in some coordinate system, a relative position with respect to a specific point, a geographic region (like a specific mountain, lake, river, county or city), a specific building, room, or floor, and many others. To make things more complicated, multiple meanings of the term may be present within the same document. And even worse, one document can have a term that represents a given notion of location explicitly, while another can have the same notion defined in terms of other notions.

When multiple documents are integrated into one, all of such relationships must be stated explicitly so that a computer program does not confuse various notions, such as various notions of "location". There are two ways of expressing such relationships. First, we could use the same name (identifier). This could be achieved by translation - all of the notions that are to have the same meaning would be mapped to a common name. Another way is to use a special relation, like "sameAs" and then letting the inference process understand such a relation. No matter which solution we choose, in order to achieve the required result, a statement must be made about "the sameness" of two items from two different information sources. Such a statement is not part of any of the documents; it is necessarily external to the documents.

Another issue that needs to be addressed is whether two items that have the same name in two different information sources should be treated as having the same meaning. As we indicated in the previous discussion of various meanings assigned to the term "location", this is not necessarily the case and thus should not be accepted as a default assumption. Again, in order to say that, we need an external statement to this effect.

The need for adding external statements about multiple information sources has relatively far reaching implications. In short, this means that we need a meta-level of representation in which such statements can reside. In the formal approach to information integration, collections of information are treated as logical theories. Logical theories are part of logic. Traditional logic does not provide any means of reasoning about theories. Normally, logical reasoning is carried out *within* a logical theory and not *about* logical theories. However, to achieve integration of context with current information we need to treat various logical theories as objects that can be manipulated by a formal mechanism that is outside of any of the theories being manipulated. In other words, we need an algebra in which logical theories are objects. The algebra needs to have operations that allow for combining objects whose results would be other objects in the algebra. One possibility is to treat logical theories as simply collections (sets) of sentences in a formal language and use set theoretic operations of union, intersection or Cartesian product. This approach, however, would not account for different meanings of the same term and same meaning of different terms, as discussed above. A theory that satisfies such a requirement is category theory [6]. In category theory, logical theories can be treated as *category objects*, and relations among theories can be treated as category *arrows* (or *morphisms*). It also provides the operation of *colimit* – a counterpart of the union operator, and the operation of *limit* – a counterpart of the Cartesian product operator. The important aspect of category theory is the fact that morphisms can be used to identify the terms that have the same meaning; the morphism are then utilized by the colimit operation in such a way that the identified terms are treated as the same in the resulting theory.

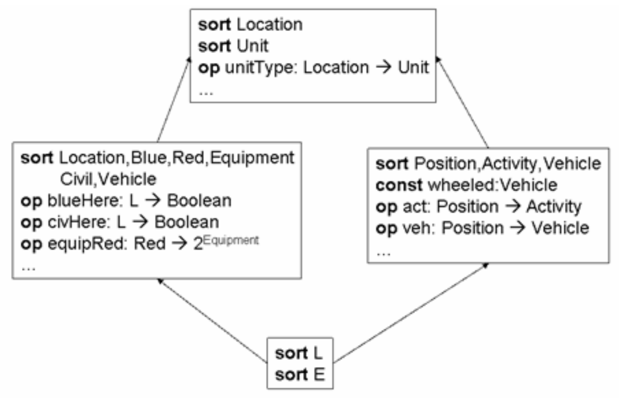

Figure 1. A diagram : colimit of Situation and Report

Now we present a simple explanation of the above concepts on the example discussed in Section 9. For this purpose we use the diagram of Figure 1, which shows four rectangles, each representing a logical theory. The left middle rectangle represents situational knowledge about the troops and civilians. As in the example of Section 9, assume that for a given location *l*, both blueHere and civHere return "false", i.e., there are no civilian and Blue forces in this location. Moreover, assume that the background knowledge includes the fact that the Red force has "wheeled" types of its equipment. The right rectangle represents the report. As the example states, presence of "wheeled" Vehicles and the Activity of "moving" are reported at the given Position. This is formally captured by the functions of *veh* and *act*, respectively.

The relationships among the theories are represented by the appropriate morphisms of the diagram. In this case, L and E are "dummy" identifiers used to unify Location in the situational theory with Position in the report and Equipment with Vehicle, respectively. At the same time, even though Vehicle appears in two specs, the meanings are different. This is because, when mapping the two theories to the top theory, everything that is not identified by the morphisms is treated as a representation of a different concept, independently of the names.

None of the theories in Figure 1 is shown in its complete form. In particular, due to space limitations, we have not shown any of the axioms of the theories. All we have shown are sorts and signatures of some of the operations. Given the diagram that includes three bottom theories, the colimit can be computed automatically [7]. This will constitute the major part of the top theory. It may need to be extended by additional axioms, but this issue is beyond the scope of this paper. Standard logical reasoning can be carried out with the top theory. The colimit operation allowed us to combine both report and situational context information in a consistent way.

11 System Development Ramifications

The problem of interpreting battlefield intelligence can be viewed as analogous to that of interpreting the discourse of a conversation. As such, we can leverage work done in the domain of discourse interpretation which includes Jerry Hobbs' framework for a theory of discourse [2] in which the following capabilities are identified: Knowledge Representation, Syntax and Semantic Translation, Knowledge Encoding, Deductive Mechanism, Specification of Possible Interpretations, Specification of the Best Interpretation. To these we add the need for "Explanation and Transparency" and then consider some implications of each capability in turn.

Knowledge Representation. Hobbs acknowledges the need of a formal means for representing knowledge. Such a requirement is critical if we wish to develop automated systems that can incorporate contextual information—yet there are very few standardized formal languages designed for such purposes (some efforts towards this end include BML [8] and UOB DAT [9]).

Syntax and Semantic Translation. For Hobbs this issue is one of translating natural language into the chosen knowledge representation. In the context of battlefield intelligence it comes down to the question of how intelligence information is collected and encoded. The process of converting the raw information into a formal representation is one that is poorly addressed by today's systems. We need a way to formally encode the syntax and semantics of the battlespace domain such as is permitted by the definition of a formal "ontology"; an effort towards this end is exemplified by the formal ontology for situation awareness described in [10].

Knowledge Encoding. Once we have agreed upon a formal method for representing both report and extra-report information there is the issue of how to capture it in that form. Today reports frequently arrive as relatively unstructured free text which must be read and interpreted by an intelligence analyst. Background knowledge resides in the analysts' heads or in the natural language used in field manuals and related documents; this includes everyday common sense and qualitative reasoning. Situation specific information is often verbally communicated, hand written or sketched on map overlays. Far too little of this information is currently captured in a standardized form that is easily rendered machine readable. This issue remains a significant hurdle preventing more rapid automation of the analysis process.

Deductive Mechanism. Assuming we can capture the relevant knowledge about a battlespace in a machine readable form there is the question of how to make use of it. This requires a formal mechanism for processing the knowledge and drawing deductions about what can be inferred from it regarding the enemy's actions and posed threats. If we view the knowledge as "axioms", as does Hobbs, then what is required is an automated theorem prover or deductive inference engine that can draw appropriate conclusions. Generic reasoners of this sort are available today and with the current state of processing power and memory capabilities it is possible for arrays of computers to apply these deductive mechanisms to the processing of large quantities of knowledge in real time.

Specification of Possible Interpretations. While the mechanism for deriving conclusions from axioms is deductive the over-reaching process that controls the reasoning need not be purely deductive. In fact in most cases there will not be just one inferable interpretation that consistently explains all of the report information in the context of the other available knowledge; due to uncertainty and lack of complete knowledge of the situation there will nearly always be several plausible interpretations. The system must therefore be capable of abductively generating multiple plausible interpretations from partial, incomplete knowledge [11].

Specification of Best Interpretations. One problem with generating multiple interpretations over partial and uncertain knowledge of a situation is that there will likely be an overwhelming number of them.. A critical requirement is therefore the ability to identify and present only the best interpretations for presentation to the analyst. What defines "best"? At least three factors should be considered. First, as suggested by Hobbs (c.f. [11]), there is the notion of economy; if we have two interpretations that describe the same thing (and would require the same response) the one that is more concise and thus faster to process (by the human) should be preferred. Second, the degree of certainty that can be assigned to an interpretation based on the accumulated certainties associated with all of the information used in its construction should play a significant role in evaluating an interpretation. Third, the potential threat posed by the enemy must be a key factor in interpretation evaluation.

Explanation and Transparency. An analyst is not going to accept an interpretation from a black box if it is not clear how the interpretation was obtained and what it is based on. For the sake of economy, interpretations must be presented at a level that abstracts out much of the detail. Inevitably there will be cases where the presented interpretation leaves out aspects that are critical to its comprehension; for example, there may be key assumptions made about the unknown location of enemy troops that do not directly appear in the interpretation. In such cases the analyst must have a means for asking for further explanation about why/how the interpretation was made. If the system is unable to provide some transparency into its reasoning process and assure the user its interpretations are plausible, the system is unlikely to be trusted and thus unlikely to be used.

12 Conclusion

A significant degree of complexity in battlespace analysis and interpretation is caused by context-dependency. In this paper, we have taken initial steps to illuminate the complexity and to provide a framework within which it can be investigated further. It appears that automated solutions to these fusion tasks will require the explicit identification of all factors underpinning context-dependency; how contexts of different types can be identified, bounded and used to reduce complexity; how these different contexts influence one another; and what knowledge types as well as what knowledge representation formalisms and inferencing mechanisms are required to allow context to be used effectively. Some promise for addressing these issues is suggested by the body of knowledge in computational discourse understanding and category theory.

References

[1] G. Powell, Army Intelligence Analysis and Interpretation: Exploring the Utility and Limitations of Computational Diagnostic Reasoning. In Proc. of DoD CCRTS, San Diego, CA, June 2004

[2] J. Hobbs, ``On the Coherence and Structure of Discourse'', Report No. CSLI-85-37, Center for the Study of Language and Information, Stanford University, 1985.

[3] FM 21-75, *Combat Skills of the Soldier*, August 1984.

[4] FM 3-0, *Operations*, June 2001.

[5] M. Kokar, J. A. Tomasik, and J. Weyman, "A formal approach to information fusion", Proceedings of 2nd Intern. Conf. on Information Fusion, vol. I, pp. 133--140, 1999.

[6] B. C. Pierce. Basic Category Theory for Computer Scientists. MIT Press, 1991.

[7] J. Tomasik and J. Weyman, Category semantics for fusion and refinement of multi-sorted specifications. In Proc. of 9th Intern. Conf. on Information Fusion, July 2006.

[8] A. Tolk, M. Hieb, K. Galvin and L. Khimeche, Modelling and Simulation to Address NATO's New and Existing Military Requirements, In Proc. of RTO NMSG Symp., Koblenz, Germany, October 2004.

[9] https://www.dmso.mil/public/thrust/ki-data/uob

[10] C. Matheus, M. Kokar and K. Baclawski, A Core Ontology for Situation Awareness. In Proc. of FUSION'03, Cairns, Queensland, Australia, pp. 545-552, July 2003.

[11] Burstall, R. M., and Goguen, J. A. Putting theories together to make specifications. In Proc. of the Fifth IJCAI, Cambridge, MA, pp. 1045-1058, August 1977.