Exploring context issues within natural language information

Kellyn Rein
Command and Control Systems
Information Analysis
Fraunhofer FKIE
Fraunhoferstr. 20, 53343 Wachtberg
GERMANY

kellyn.rein@fkie.fraunhofer.de

“Hard data” vs “soft data”

Hard Data is defined as

> “data in the form of numbers or graphs, as opposed to qualitative information. In the world of Big Data and the Internet of Things (IoT), Hard Data describes the types of data that are generated from devices and applications, such as phones, computers, sensors, smart meters, traffic monitoring systems, call detail records, bank transaction records, among others. This information can be **measured, traced, and validated**….”

Soft Data [is defined] as

> “human intelligence, data that **is full of opinions, suggestions, interpretations, contradictions and uncertainties**.”

Your context is not my context

Oxford Dictionaries offers the following two-pronged definition:

1. the circumstances that form the setting for an event, statement, or idea, and in terms of which it can be fully understood.
2. the parts of something written or spoken that immediately precede and follow a word or passage and clarify its meaning.

dictionary.com likewise gives two variations on context:

1. the parts of a written or spoken statement that precede or follow a specific word or passage, usually influencing its meaning or effect.
2. the set of circumstances or facts that surround a particular event, situation, etc.

Your context is not my context

Oxford Dictionaries offers th

1. the circumst idea, and in terms
2. the parts of so nd follow a word

dictionary.com like

1. the parts of a wri ment that precede or follow a specific word or passage, ..y influencing its meaning or effect.
2. the set of circumstances or facts that surround a particular event, situation, etc.

Within the sensor data fusion community, "context" indicates knowledge about the immediate environment of the sensors. For example, Schilit and Theimer [9] use context to indicate the location, identities of nearby people and objects, and changes to those objects.

Thus: definition 1 from Oxford and definition 2 from dictionary.com

Your context is not my context

Oxford Dictionaries offers the following two-pronged definition:

1. the circumstances that form the setting for an event, statement, or idea, and in terms of which it can be fully understood.
2. the parts of something written or spoken that immediately precede and follow a word or passage and clarify its meaning.

dictionary.com likewise gives two variations on context:

1. the parts of a written or spoken statement that precede or follow a specific word or passage, usually influencing its meaning or effect.
2. the set of circumstances or facts that surround a particular event, situation, etc.

Your context is

Oxford

1. the
   term
2. the p
   a wo

dictionary

1. the pa
   word o
2. the set

At the data level, the soft data community is generally concerned with the context of words and phrases within the text.

At a higher level "context" may include political orientation of the writer, overt or covert intent in the writing, social and cultural background and other rather difficult factors to measure.

Furthermore, context may vary depending upon the expectations, knowledge, prejudices, experience of the information recipient.

Thus: both definitions from Oxford and definition 1 from dictionary.com

Your context is not my context

Oxford Dictionaries offers the following two-pronged definition:

1. the circumstances that form the setting for an event, statement, or idea, and in terms of which it can be fully understood.
2. the parts of something written or spoken that immediately precede and follow a word or passage and clarify its meaning.

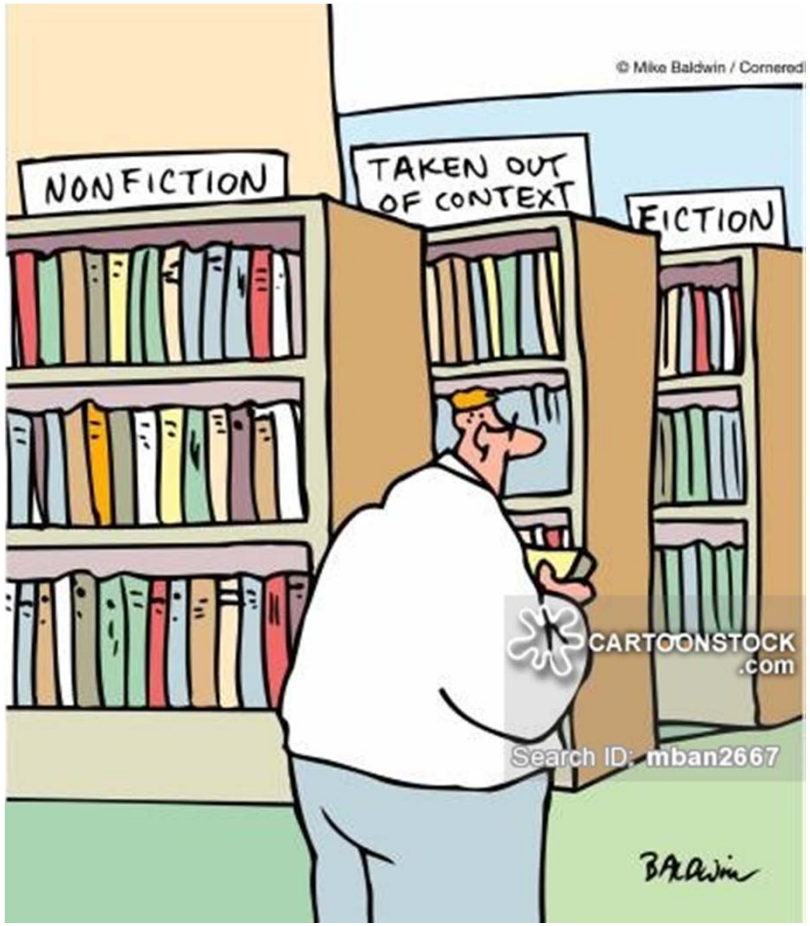
dictionary.com likewise gives two variations on context:

1. the parts of a written or spoken statement that precede or follow a specific word or passage, usually influencing its meaning or effect.
2. the set of circumstances or facts that surround a particular event, situation, etc.

So, we can say that the meaning of " context" depends upon…..

So, we can say that the meaning of " context" depends upon…..the context!

Nonfiction, taken out of context, fiction.



Situation awareness vs intelligence

According to Endsley, situation awareness is:

"the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the *projection of their status in the near future*," (italics added).

Continuous updating of important environmental elements in the area of interest such as locations of military units (both friendly and hostile), movements of personnel and equipments, locations and conditions of facilites, structures, etc.

Also information on non-military or paramilitary activities, political climate, tribe coalitions.

Often displayed on maps and C2 systems.

Situation awareness vs intelligence

According to Endsley, situation awareness is:

"the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and [illegible] *of their status in the near fu[illegible]*

Continuous updatin[illegible] of interest such as loca[illegible] movements of perso[illegible] facilites, structures, e[illegible]

Also information on nc[illegible] paramilitary activities, political climate, tribe coalitions.

Often displayed on maps and C2 systems.

In general, both the geographical area and the timeline are (relatively) restricted.

"Small footprint, shorter timelines"

Situation awareness vs intelligence

Intelligence requires careful and systematic collection of information with the goal of detecting patterns of behavior being used by the enemy in order to disrupt threatening activities.

In contrast to SA, sense-making for intelligence purposes often involves timelines which are much longer, covering weeks, months or years rather microseconds, minutes or hours. Furthermore, the geographical area covered may be very extensive.

> For example: the current fight against ISIS involves information-gathering on several continents, and that information is to a very great extent text-based.

The data collected may include focused reports from intelligence assets and analyses from various agencies, but also may include many types of open sources including news sources, government documents and research results.

Situation awareness vs intelligence

Intelligence requires careful and systematic collection of information with the goal of detecting patterns of behavior being used by the enemy in order to disrupt threatening activities.

In contrast to SA, sense-maki… …ves timelines which ar… …ther microseconds, min… covered may be ver…

For example: th… gathering on se… …great extent text-base…

In general, both the geographical area of interest and the timeline are for all intents and purposes unrestricted.

"large footprint, long timelines"

The data collected may include focused reports from intelligence assets and analyses from various agencies, but also may include many types of open sources including news sources, government documents and research results.

Situation awareness vs intelligence

SA: concerned with safety and protection of assets, “here and now“

Intelligence: longer term analysis of patterns of behavior: movements of ships over time, communications with other port authorities, etc.


How deep
is the mud?
Depends on
who you ask.
We all go through the
same stuff differently.

Processing natural language data

An information extraction pipeline typically consists of the following elements:

1. **Tokenizer** determines individual tokens of text (words, numbers, abbreviations, punctuation marks)

Processing natural language data

An information extraction pipeline typically consists of the following elements:

1. **Tokenizer** determines individual tokens of text (words, numbers, abbreviations, punctuation marks)
2. **Gazetteer** compares tokens to elements of lists containing the names of various types such as person names, organizations, towns, landmarks, etc. Often very domain-specific (particularly geographical elements)

Processing natural language data

An information extraction pipeline typically consists of the following elements:

1. **Tokenizer** determines individual tokens of text (words, numbers, abbreviations, punctuation marks)
2. **Gazetteer** compares tokens to elements of lists containing the names of various types such as person names, organizations, towns, landmarks, etc. Often very domain-specific (particularly geographical elements)
3. **Sentence splitter** determines the boundaries of sentences (beginning and end). Rules must take into account things such as titles (*Dr.*, *Mrs.*, etc.) or abbreviations such as *i.e.*, *e.g.*, *etc.*, and so on.

Processing natural language data

An information extraction pipeline typically consists of the following elements:

1. **Tokenizer** determines individual tokens of text (words, numbers, abbreviations, punctuation marks)
2. **Gazetteer** compares tokens to elements of lists containing the names of various types such as person names, organizations, towns, landmarks, etc. Often very domain-specific (particularly geographical elements)
3. **Sentence splitter** determines the boundaries of sentences (beginning and end). Rules must take into account things such as titles (*Dr.*, *Mrs.*, etc.) or abbreviations such as *i.e.*, *e.g.*, *etc.*, and so on.
4. **Part-of-speech tagger** identifies elements as noun, verb, preposition, etc., based upon the definition of the word as well as its context within the sentence.

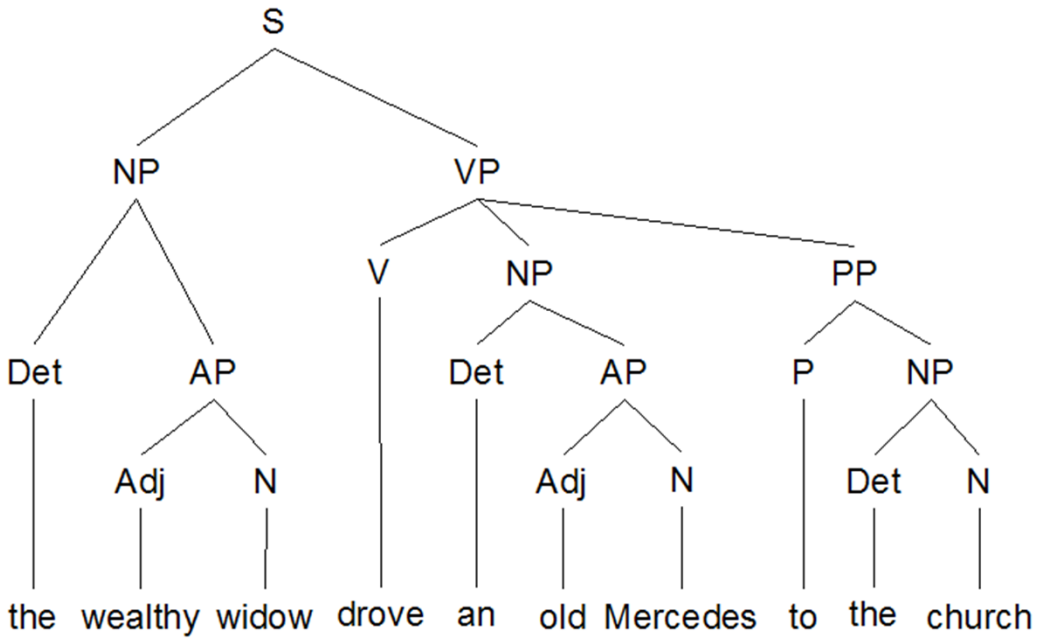
Processing natural language data

An information extraction pipeline typically consists of the following elements:

1. **Tokenizer** determines individual tokens of text (words, numbers, abbreviations, punctuation marks)
2. **Gazetteer** compares tokens to elements of lists containing the names of various types such as person names, organizations, towns, landmarks, etc. Often very domain-specific (particularly geographical elements)
3. **Sentence splitter** determines the boundaries of sentences (beginning and end). Rules must take into account things such as titles (*Dr.*, *Mrs.*, etc.) or abbreviations such as *i.e.*, *e.g.*, *etc.*, and so on.
4. **Part-of-speech tagger** identifies elements as noun, verb, preposition, etc., based upon the definition of the word as well as its context within the sentence.
5. **Named-entities transducer** combines elements from the gazetteers above:

   example: for "Dr. Mohammed el-Baradei", the gazetteer will provide the annotations *title* for "Dr.", *male forename* for "Mohammed" and *surname* for "el-Baradei"

Processing natural language data



The complete parse tree of "The wealthy widow drove an old Mercedes to the church."[jenge et al]

Processing natural language data

Semantic role labelling links word meanings to sentence meaning by exploiting syntactic, lexical, and semantic information.

In English, syntactic information is based upon word order information:

"dog bites man" vs "man bites dog"

who is doing the biting and who is being bitten is determined by who appears before the verb and who appears after.

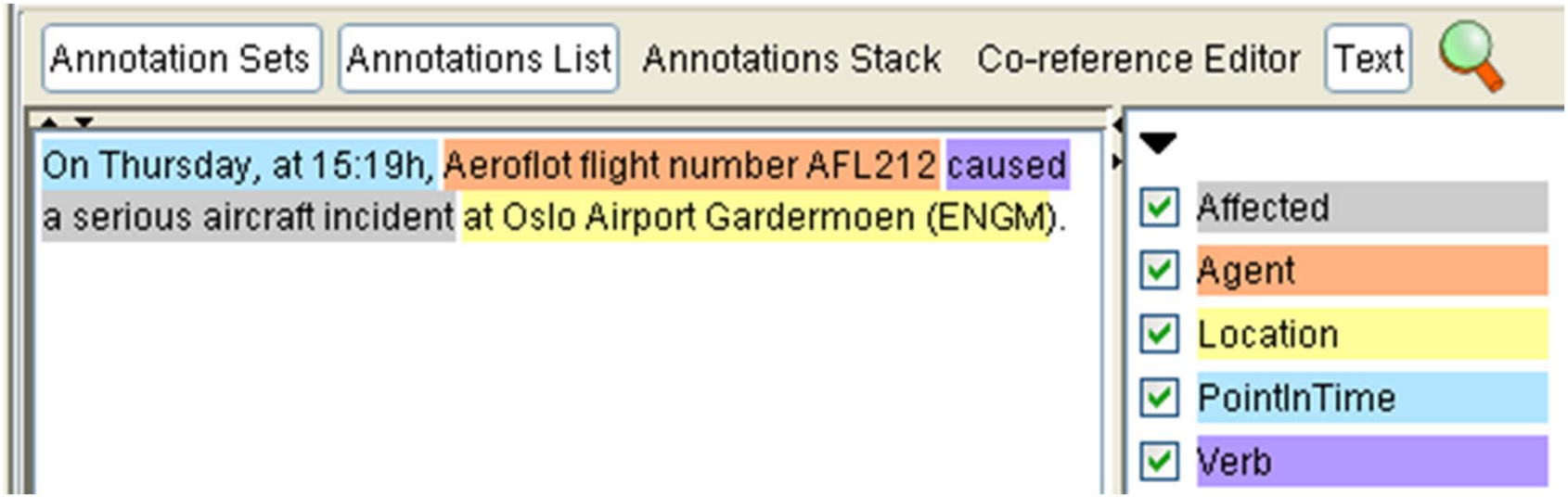
In German the role is determined by case endings:

"Der Hund beißt den Mann" vs "Den Mann beißt der Hund"

Lexical information is provided mostly by verbs and prepositions.

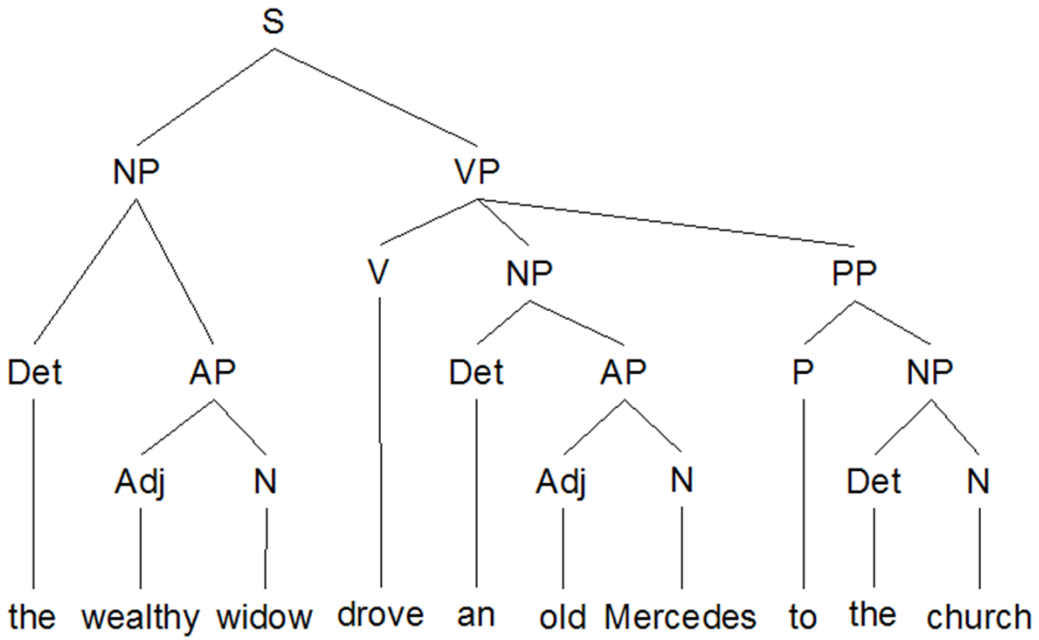
-- the preposition "*at*" normally signals either a *location*
"*at the townhall*" or *point in time* (e.g., "*at one o'clock*").

Processing natural language data



Preliminary labelling of semantic role information as calculated by MIETER developed by Fraunhofer FKIE.

Processing natural language data



The complete parse tree of "The wealthy widow drove an old Mercedes to the church."[jenge et al]

Processing natural language data

As can be seen from the preceding examples, a single sentence may contain a myriad of individual pieces of data:

- the widow is wealthy,
- she drove a Mercedes,
- she can drive,
- the car is old,
- she went to the church for some reason

- the aircraft incident was serious,
- it happened on a Thursday,
- it happened at 15:19,
- the aircraft involved belonged to Aeroflot,
- its flight number was AFL212

Text analytics

A variety of techniques for analyzing natural language text and retrieving certain types of information from the documents at hand using analysis techniques based upon lexical and grammatical patterns in the language. Among these are:

- **Document classification**: using linguistic and statistical analysis, documents may be classified (type of content, language, etc.), summarized, clustered (based upon predefined or learned classification.

Text analytics

A variety of techniques for analyzing natural language text and retrieving certain types of information from the documents at hand using analysis techniques based upon lexical and grammatical patterns in the language. Among these are:

- **Document classification**: using linguistic and statistical analysis, documents may be classified (type of content, language, etc.), summarized, clustered (based upon predefined or learned classification.
- **Named-entity recognition/pattern recognition**: identification of names of individuals, places, organizations, etc., as well as patterns such as telephone numbers, email addresses, etc.

Text analytics

A variety of techniques for analyzing natural language text and retrieving certain types of information from the documents at hand using analysis techniques based upon lexical and grammatical patter

- **Document classificat** nents may be classified (type of co ed upon predefined or learne
- **Named-entity recogn** s of individuals, places, organizations ers, email addresses, etc.

Question for the non-Americans in the audience:

What does this pattern represent?

123 - 45 - 6789

Named-entity recognition would likely have some problems with this one!

An 82-year-old Georgia woman named Serpentfoot is trying to change her name to a 101-word articulation of her philosophy.

Nofoot Allfoot-69-mouth-tail-solids-liquids-gases-animals-vegetable-mineral-all-predators-and-prey-that-consume-and-move-with-feet-fins-wings-wheels-canes-roots-limbs-vines-landslides-dust-wind-water-fire-ice-gravity-vacuums-black-holes-going-over-under-around-and-through-Our-Greater-Self-our-habitat-the-cosmos-of-which-we-are-but-part-and-where-all-life-feeds-upon-other-life-from-the-smallest-atoms-or-bacteria-to-the-great-black-holes-and-dog-eat-dog-and-"Last-Suppers"-where-we-are-what-we-eat-or-consume-and-each-lives-on-in-the-other…∞ Serpentfoot

Text analytics

A variety of techniques for analyzing natural language text and retrieving certain types of information from the documents at hand using analysis techniques based upon lexical and grammatical patterns in the language. Among these are:

- **Document classification**: using linguistic and statistical analysis, documents may be classified (type of content, language, etc.), summarized, clustered (based upon predefined or learned classification.
- **Named-entity recognition/pattern recognition**: identification of names of individuals, places, organizations, etc., as well as patterns such as telephone numbers, email addresses, etc.
- **Coreference identification**: identifying alternate names for the same object.
  - "Barack Obama", "President Obama", "the US president", "the 44th president", "44"

Text analytics

A variety of techniques for analyzing natural language text and retrieving certain types of information from the documents at hand using analysis techniques based upon lexical and grammatical patterns in the language. Among these are:

- **Document classification**: using linguistic and statistical analysis, documents may be classified (type of content, language, etc.), summarized, clustered (based upon predefined or learned classification.
- **Named-entity recognition/pattern recognition**: identification of names of individuals, places, organizations, etc., as well as patterns such as telephone numbers, email addresses, etc.
- **Coreference identification**: identifying alternate names for the same object.
  - "Barack Obama", "President Obama", "the US president", "the 44th president", "44"
- **Sentiment analysis**: uses lexical clues such as specific words or phrases buried within the text to determine prevailing sentiment, emotion or opinion.

Text analytics

A variety of techniques for analyzing natural language text and retrieving certain types of information from the documents at hand using analysis techniques based upon lexical and grammatical patterns in the language. Among these are:

- **Document classification**: using linguistic and statistical analysis, documents may be classified (type of content, language, etc.), summarized, clustered (based upon predefined or learned classification.
- **Named-entity recognition/pattern recognition**: identification of names of individuals, places, organizations, etc., as well as patterns such as telephone numbers, email addresses, etc.
- **Coreference identification**: identifying alternate names for the same object.
  - "Barack Obama", "President Obama", "the US president", "the 44th president", "44"
- **Sentiment analysis**: uses lexical clues such as specific words or phrases buried within the text to determine prevailing sentiment, emotion or opinion.
- **Relationship and event extraction**: identifying relationships between objects in text
  - "Susan works at ABC Company", "Jane is the sister of Bob", "Mozart died in 1791"

Structuring natural language data

Extracted text-based information is often stored in structured formats for further processing and simplified access. Currently, the most widely structures for storage of text-based information for automatic processing generally fall into two categories: ontologies, and databases / triple stores, the latter of which are a special kind of database.

Each of these has its strengths and weaknesses for sense-making, which we will discuss in this section.

Structuring natural language data

Ontologies contain information about the characteristics of and relationships between different classes of objects within a specific domain, that is, a definition of a shared concept of the objects in the domain.

For domain “humans”:

- a “parent” is a (human) object who has at least one instance of an object called “child”,
- a “mother” is a special subclass of parent with the extra characteristic that she also has the gender “female” and so on.

Structuring natural language data

Ontologies contain information about the characteristics of and relationships between different classes of objects within a specific domain, that is, a definition of a shared concept of the objects in the domain.

For domain “humans”:

- a “parent” is a (human) object who has at least one instance of an object called “child”,
- a “mother” is a special subclass of parent with the extra characteristic that she also has the gender “female” and so on.

Then we know some things about entities :

“Mary must be female because she is a mother”

and relationships between objects

“If Mary is Susan’s mother, then Susan is Mary’s child”).

Structuring natural language data

Ontologies have the advantage that we have defined in advance exactly what each class of objects is and how it relates to all other objects within our domain of interest.

Structuring natural language data

Ontologies have the advantage that we have defined in advance exactly what each class of objects is and how it relates to all other objects within our domain of interest.

However, ontologies are classification systems, and in the process of building the ontology we must ***make a priori decisions as to what things belong together***.

Structuring natural language data

Databases are useful for storing large amounts of often complex information about specific instances of objects within the domain of interest.

The information contained within a relational database is stored in a series of files containing objects (records) of similar structures, which can be represented as tables.

Structuring natural language data

Databases are useful for storing large amounts of often complex information about specific instances of objects within the domain of interest.

The information contained within a relational database is stored in a series of files containing objects (records) of similar structures, which can be represented as tables.

In order to retrieve information, one must have exact knowledge about the structures.

Structuring natural language data

However, determining the structure ahead of time means that the analysts have made ***a priori decisions*** *as to what information is needed and what information belongs together*.

Later changes to the structures within the database are possible, but not always easy to effect.

Structuring natural language data

A triple store is a potential solution to some of the complexity issues of a relational database.

Rather than records inside of more complexly structured file a triple is a three-part data entity in the form

subject-predicate-object:

Structuring natural language data

A triple store is a potential solution to some of the complexity issues of a relational database.

Rather than records inside of more complexly structured file a triple is a three-part data entity in the form

subject-predicate-object:

“1-800-555-1234 is-a telephone number”
“Susan Smith works-at ABC Company”
“ABC Company produces widgets”

Out of context, out of mind

Intelligence requires careful and systematic collection of information with the goal of detecting patterns of behavior being used by the enemy in order to disrupt threatening activities.

Over time the enemy learns from past mistakes and modify their behavior to again escape detection.

This means that the ***threat models and behavioral expectations which are created today may well be outdated tomorrow.***

Out of context, out of mind

Intelligence requires careful and systematic collection of information with the goal of detecti… f behavior being used by th… atening activities

Over time … d modify their beha…

This mean… expectatio… utdated tomorrow.

This also means that ***information which we find unimportant today may be highly significant tomorrow.***

Out of context, out of mind

Intelligence re… …nation with the goal… …in order to disru…

Over… …behavior to again…

This m…

This als… …y may be highly s…

Additior… …omplex over time

Additionally, we may not always know in advance what we are looking for.

Therefore, to make decisions in advance about what is and what isn't important may prove to be a mistake.

Out of context, out of mind

Extracting and storing isolated pieces of information out of the context in which they were stated may result information loss.

Elaine flew from London to Stockholm via Amsterdam on 17 November.

Out of context, out of mind

Extracting and storing isolated pieces of information out of the context in which they were stated may result information loss.

Elaine flew from London to Stockholm via Amsterdam on 17 November.

From this we can, of course, extract triples such as “Elaine flew to Stockholm”, “Elaine flew via Amsterdam” and “Elaine flew on 17 November.”

Out of context, out of mind

Extracting and storing isolated pieces of information out of the context in which they were stated may result information loss.

Elaine flew from London to Stockholm via Amsterdam on 17 November.

From this we can, of course, extract triples such as “Elaine flew to Stockholm”, “Elaine flew via Amsterdam” and “Elaine flew on 17 November.”

However, the fact that Elaine flew via Amsterdam ***on that particular date*** (perhaps because another person of interest also was at Amsterdam airport on that day) – may be of most interest .

Out of context, out of mind

Extracting and storing isolated pieces of information out of the context in which they were stated may result information loss.

Elaine flew from London to Stockholm via Amsterdam on 17 November.

From this we can, of course, extract triples such as “Elaine flew to Stockholm”, “Elaine flew via Amsterdam” and “Elaine flew on 17 November.”

However, the fact that Elaine flew via Amsterdam ***on that particular date*** (perhaps because another person of interest also was at Amsterdam airport on that day) – may be of most interest .

Thus the context (day, time, from where, to where, etc.,) may be key to understanding the meaning of Elaine’s travel.

Out of context, out of mind

Extracting and storing isolated pieces of information out of the context in which they were s

Elaine flew fro ***vember.***

From this we can, Stockholm", "Elai vember."

However, if we ar that the most interesting i ***hat*** ***particular date*** (p as at Amsterdam airpo to reconstruct unles

Solution:

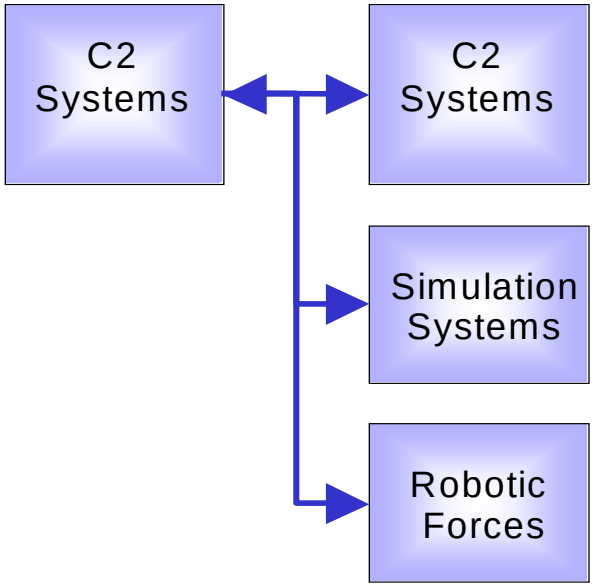
A structured machine-processable format which preserves content and context, such as

Battle Management Language

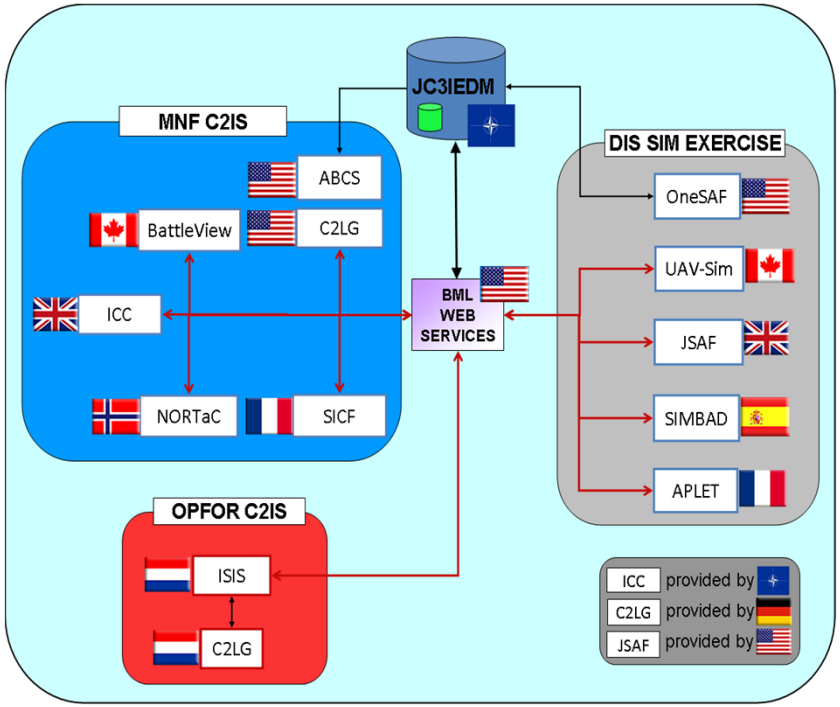
Thus the context (day, time, from where, to where, etc.,) may be key to understanding the meaning of Elaine's travel.

Battle Management Language: a common basis for communication

- Started under SISO project group Coalition Battle Management Language
- Later also under aegis of NATO RTO MSG-048 (Modeling and Simulation) and MSG-085
- Terms and values from NATO standard data model JC3IEDM serve as lexical elements of BML
- Defines terms for war operations as well as non-war operations such as disaster relief



Coalition BML has proven successful for communicating between command and control systems of multiple nations



Architecture of successful experiment for NATO RTO MSG.048 “Coalition BML” in Manassas, Virginia, November 2009

Potential of BML approach for fusing high and low level data

BML foresees representation for and processing of

- HUMINT/OSINT (text) information
- Results of sensor data processing
- Results of other fusion algorithms

Formal Grammar

- C2LG is a Context Free Grammar
-
- Grammar $G = \langle \Phi, \Sigma, R, S \rangle$
- Φ = Non Terminal Symbols
- Σ = Terminal Symbols
- Production rules $R \subseteq \Gamma^* \times \Gamma^* \mid \Gamma = \Phi \cup \Sigma$ (written $\alpha \rightarrow \beta$)
- Start symbol $S \in \Phi$

- Context Free Grammar
- $\alpha \rightarrow \beta \mid \alpha \in \Phi \wedge \beta \in \Gamma^*$

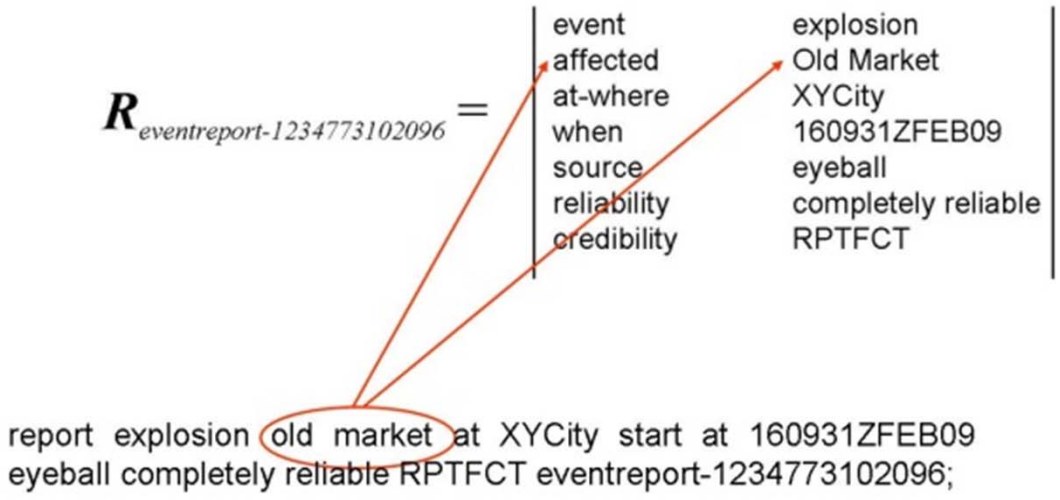
"Coalition patrol reports a bomb was set off in the old market in XYCity about half an hour ago."

"Coalition patrol reports a bomb was set off in the old market in XYCity about half an hour ago."

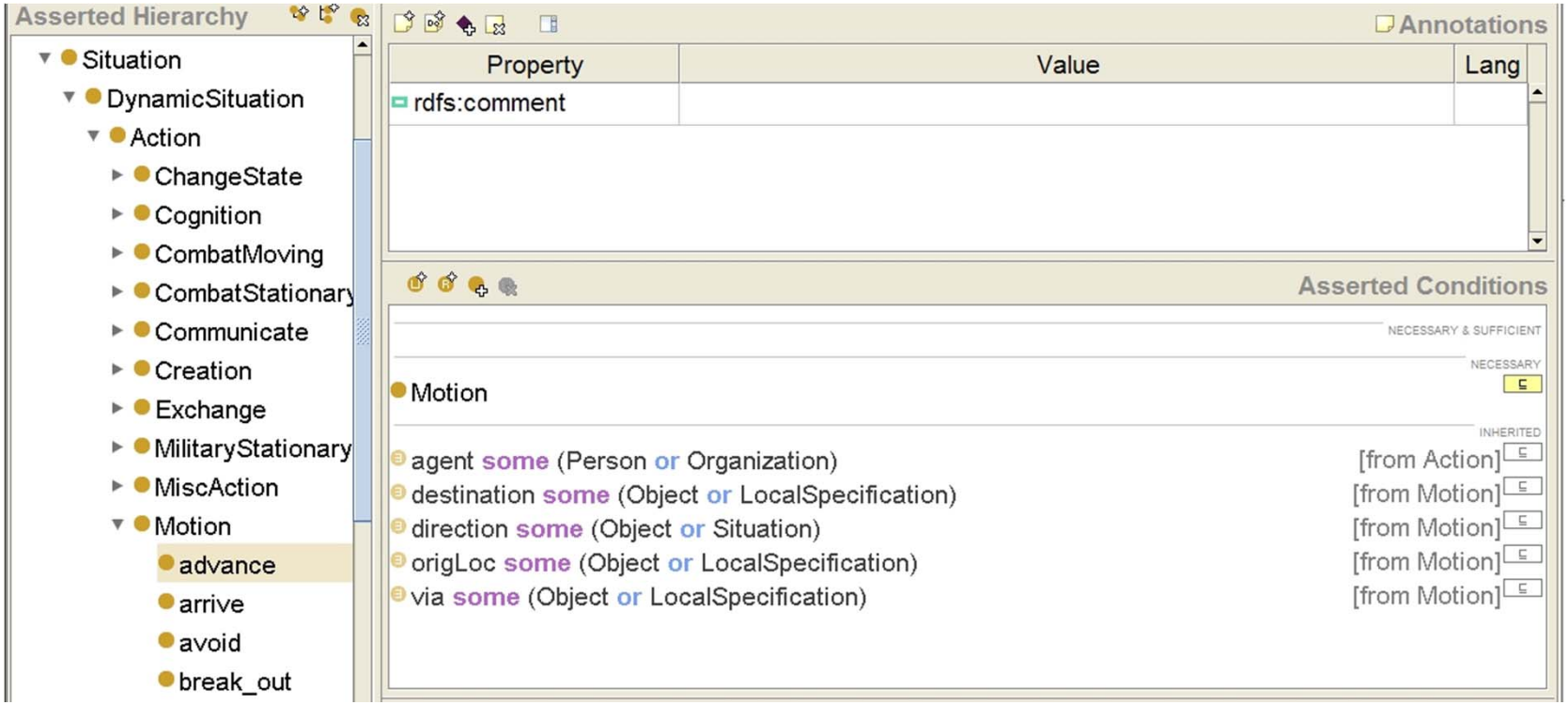
report explosion old market at XYCity start at 160931ZFEB09 eyeball completely reliable RPTFCT eventreport-1234773102096;

BML for structuring natural language data

Representation of the report "Coalition forces report the detonation of a bomb at the Old Market in XYCity at shortly past 4 p.m. today" as a BML string (bottom) and a feature-value (structured) matrix.



BML for structuring natural language data



BML for structuring natural language data

Among the rules for verbs of motion in BML is the non-terminal “RouteWhere” which can be expanded in the following three ways:

a) RouteWhere → ***along*** RouteName
b) RouteWhere → ***towards*** Location | ***towards*** Bearing
c) RouteWhere → (***from*** Location) ***to*** Location (***via*** Location*)

BML for structuring natural language data

Among the rules for verbs of motion in BML is the non-terminal “RouteWhere” which can be expanded in the following three ways:

a) RouteWhere → ***along*** RouteName

b) RouteWhere → ***towards*** Location | ***towards*** Bearing

c) RouteWhere → (***from*** Location) ***to*** Location (***via*** Location*)

In a) “RouteWhere” can be expanded by the keyword “***along***” followed by the unique name (“RouteName”) of a route which is already known (i.e., is stored in the database).

BML for structuring natural language data

Among the rules for verbs of motion in BML is the non-terminal “RouteWhere” which can be expanded in the following three ways:

a) RouteWhere → ***along*** RouteName

b) RouteWhere → ***towards*** Location | ***towards*** Bearing

c) RouteWhere → (***from*** Location) ***to*** Location (***via*** Location*)

In a) “RouteWhere” can be expanded by the keyword “***along***” followed by the unique name (“RouteName”) of a route which is already known (i.e., is stored in the database).

In b) only the direction of the movement is known, so “RouteWhere” is expanded by the keyword “***towards***” followed by either a location (such as a city or landmark) or a bearing (i.e., cardinal point such as “north” or degrees between 0 and 360).

BML for structuring natural language data

Among the rules for verbs of motion in BML is the non-terminal "RouteWhere" which can be expanded in the following three ways:

a) RouteWhere → ***along*** RouteName

b) RouteWhere → ***towards*** Location | ***towards*** Bearing

c) RouteWhere → (***from*** Location) ***to*** Location (***via*** Location*)

In a) "RouteWhere" can be expanded by the keyword "***along***" followed by the unique name ("RouteName") of a route which is already known (i.e., is stored in the database).

In b) only the direction of the movement is known, so "RouteWhere" is expanded by the keyword "***towards***" followed by either a location (such as a city or landmark) or a bearing (i.e., cardinal point such as "north" or degrees between 0 and 360).

In c) "RouteWhere" can be expanded by a sequence of three spatial constituents, namely an optional starting point (also called origin) that is preceded by the keyword "***from***", a mandatory destination preceded by the keyword "***to***", and an optional path identified by the keyword "***via***". In the case of the path constituent it is possible to list more than one location following the keyword "***via***", i.e., the path between origin and destination need not be a straight line.

BML for structuring natural language data

Among the rules for verbs of motion in BML is the non-terminal “RouteWhere” which can be expanded in the following three ways:

a) RouteWhere → ***along*** RouteName

b) RouteWhere → ***towards*** Location | ***towards*** Bearing

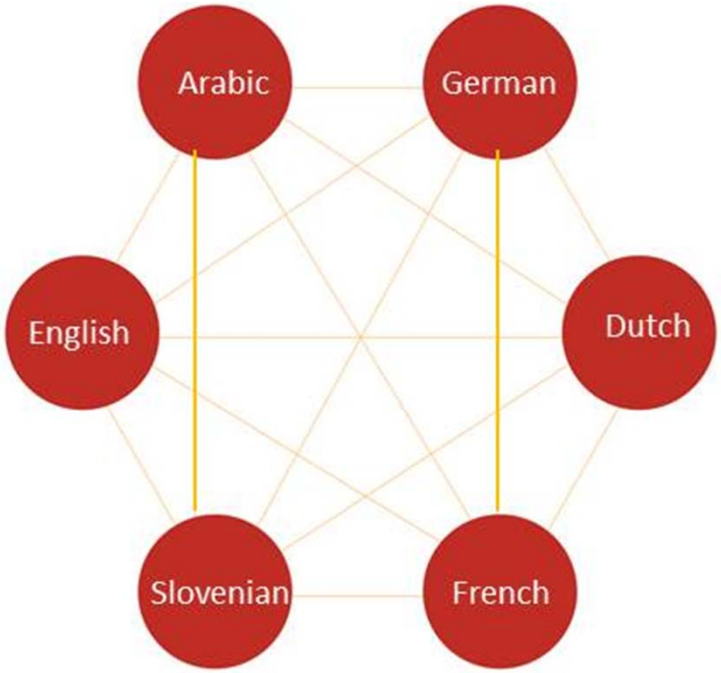
c) RouteWhere → (***from*** Location) ***to*** Location (***via*** Location*)

In a) “RouteWhere” can be expanded by the keyword “***along***” followed by the unique name (“RouteName”) of a route which is already known (i.e., is stored in the database).

In b) only the direction of the movement is known, so “RouteWhere” is expanded by the keyword “***towards***” followed by either a location (such as a city or landmark) or a bearing (i.e., cardinal point such as “north” or degrees between 0 and 360).

In c) “RouteWhere” can be expanded by a sequence of three spatial constituents, namely an optional starting point (also called origin) that is preceded by the keyword “***from***”, a mandatory destination preceded by the keyword “***to***”, and an optional path identified by the keyword “***via***”. In the case of the path constituent it is possible to list more than one location following the keyword “***via***”, i.e., the path between origin and destination need not be a straight line.

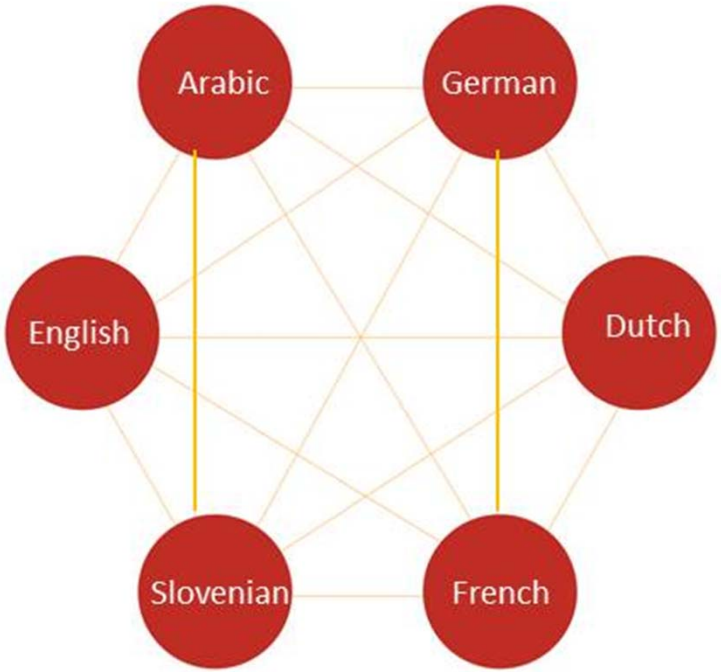
BML as a lingua franca for fusion



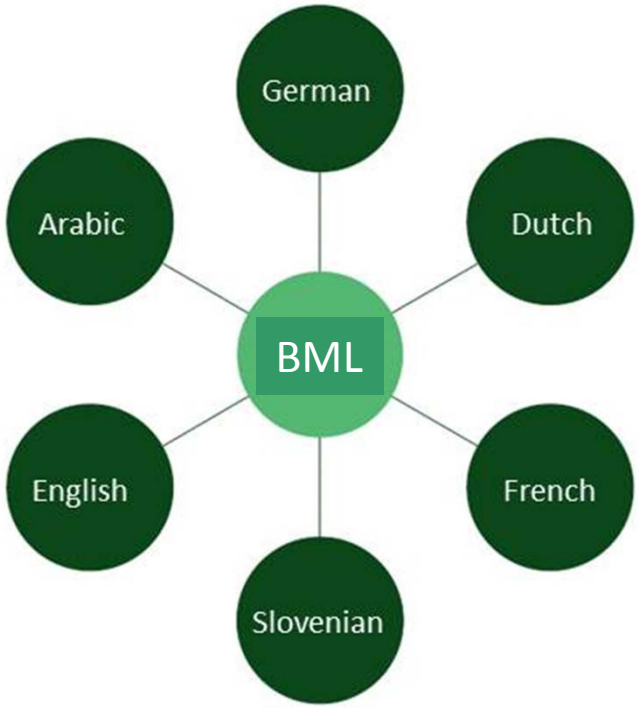
In international operations there may be a multitude of languages being used by various players (e.g.,coalition partners). As a result, various pieces of information about the area of interest may be presented in different languages.

Fusion of these varous puzzle pieces requires translation from one language to another.

BML as a lingua franca for fusion

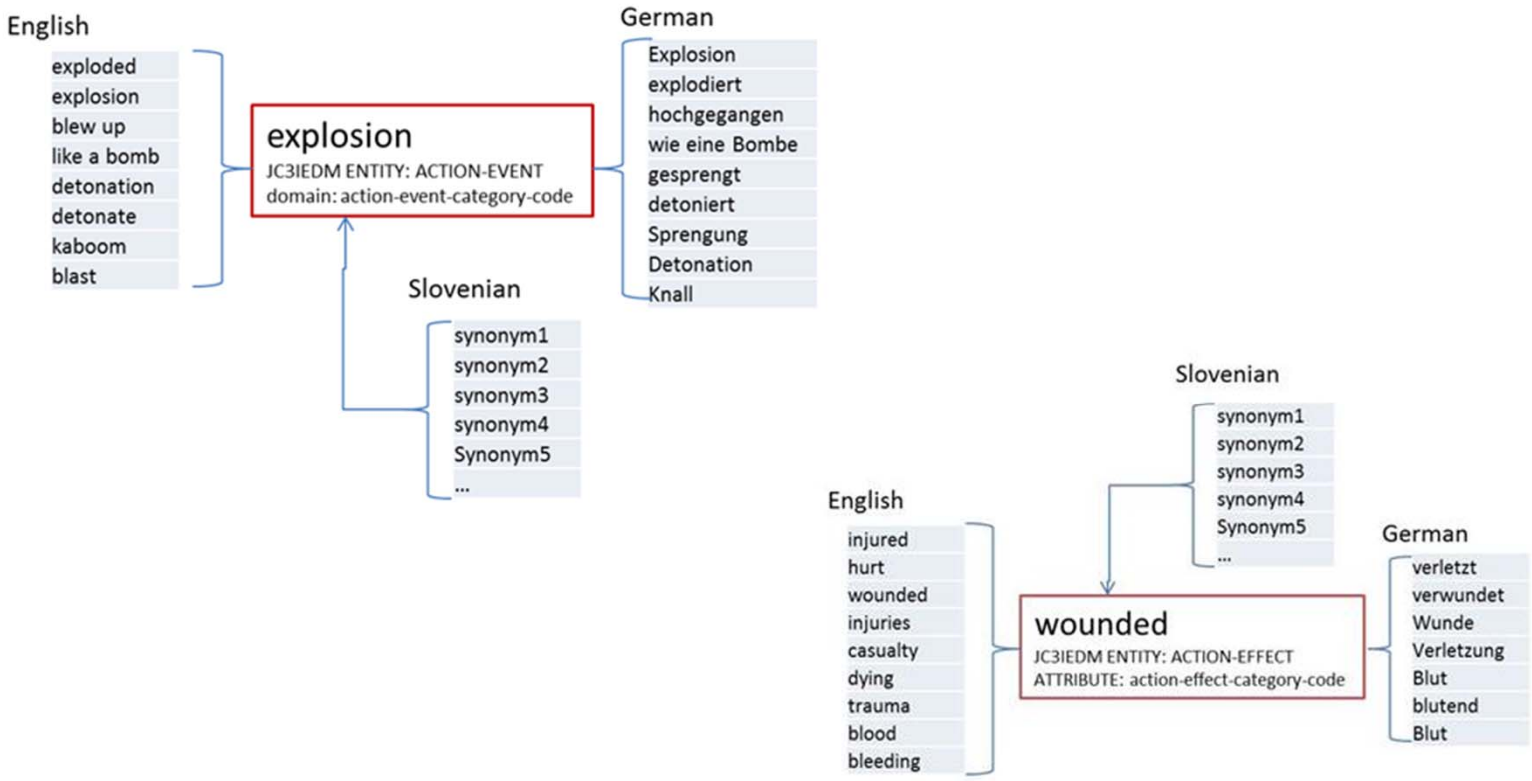


Alternative: BML in the center.



BML as a lingua franca for fusion

Reducing synonymy



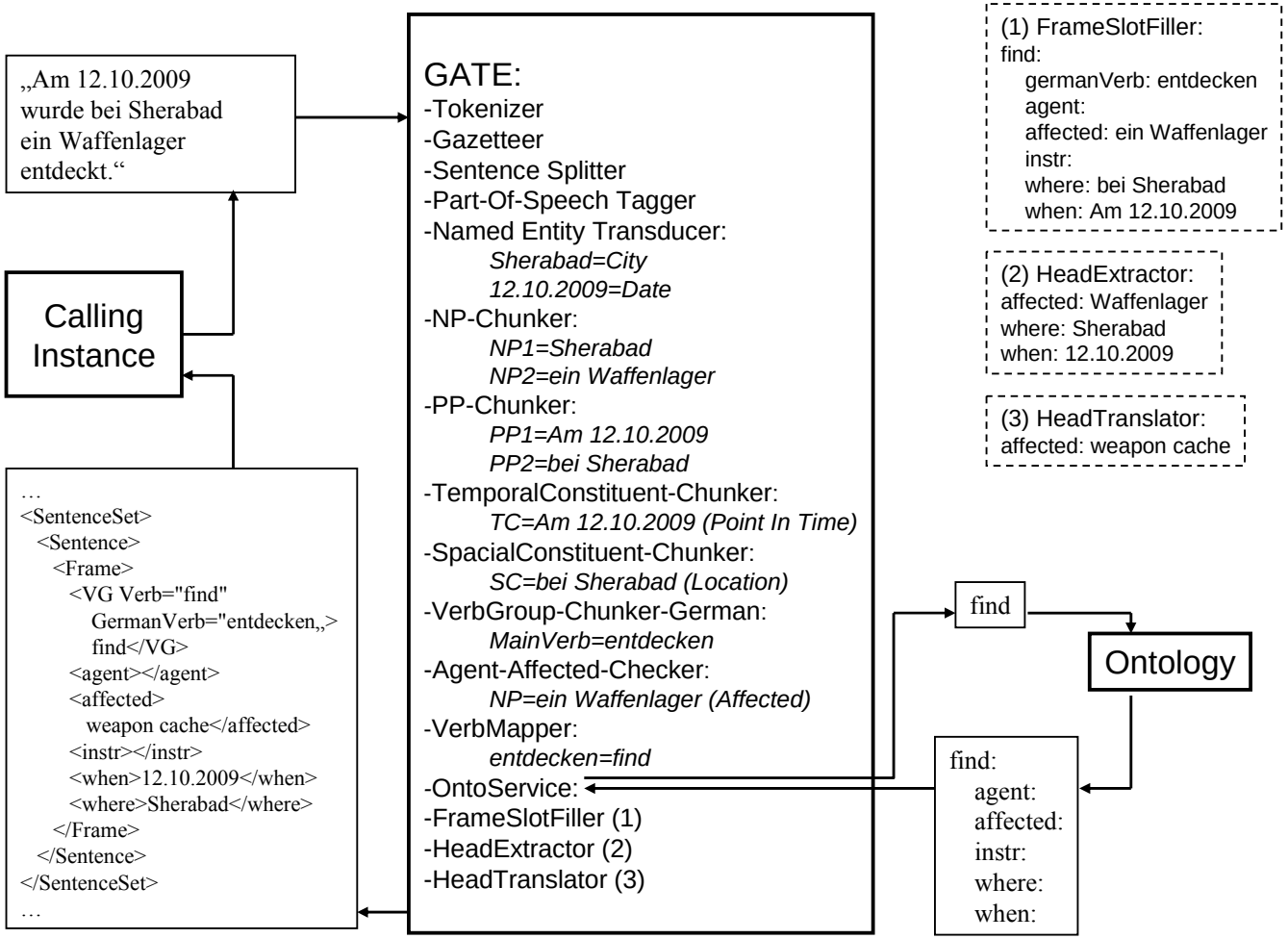
BML as a lingua franca for fusion

Conversion of German to BML

Am 12.10.2009 wurde ein Waffenlager bei Sherabad entdeckt.

BML as a lingua franca for fusion

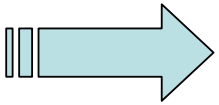
Conversion of German to BML



BML as a lingua franca for fusion

Conversion of German to BML

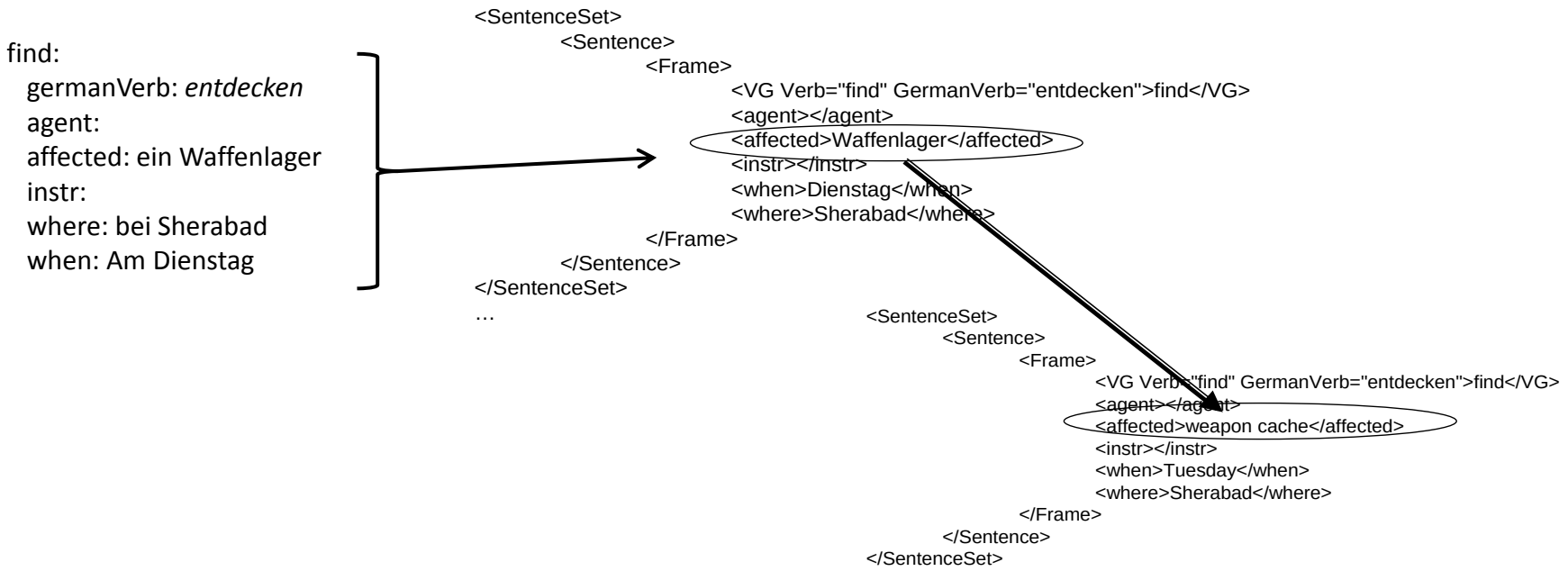
find:
germanVerb: *entdecken*
agent:
affected: ein Waffenlager
instr:
where: bei Sherabad
when: Am Dienstag



```
<SentenceSet>
    <Sentence>
        <Frame>
            <VG Verb="find" GermanVerb="entdecken">find</VG>
            <agent></agent>
            <affected>ein Waffenlager</affected>
            <instr></instr>
            <when>Am Dienstag</when>
            <where>bei Sherabad</where>
        </Frame>
    </Sentence>
</SentenceSet>
…
```

BML as a lingua franca for fusion

Conversion of German to BML



BML as a lingua franca for fusion

BML as a lingua franca for fusion

Longcross Chain

- Weight: ~450kg
- 20 km/h
- 200 kg Payload

RUAG „Garm“ Chain

- Weight: ~500kg
- 20 km/h
- 200 kg Payload

BML as a lingua franca for fusion

RB → Hostility Regarding (Identification Status-Value) At- Where When Certainty Label

The rules for an „Information Report“ are a specialized case of that rule:

RB → Hostility Phenomenon Identification MeasuredValue At-Where When Certainty Label
MeasuredValue → ValueOfMeasure UnitOfMeasurement

Example:
[information report] neutral Temperature Weather-Sensor0815 16.5 degree at [Point A] ongoing at 20101211124322.456 RPTFCT UGS Weather-Sensor0815-measure0154;

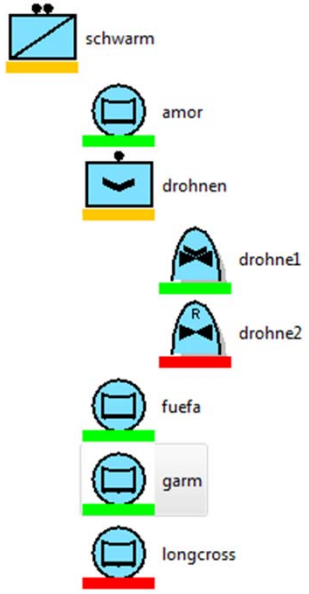
In the example, a robot reports that its sensor “Weather-Sensor0815” has measured a value of 16.5 degrees for the phenomenon “Temperature”. This measurement was taken at Point A and was done at the point in time following the “ongoing at” keywords. It also says that this measurement is reported as fact (RPTFCT) and that its source in an unattended ground sensor (UGS). This report was labelled “Weather-Sensor0815-measure0154”.

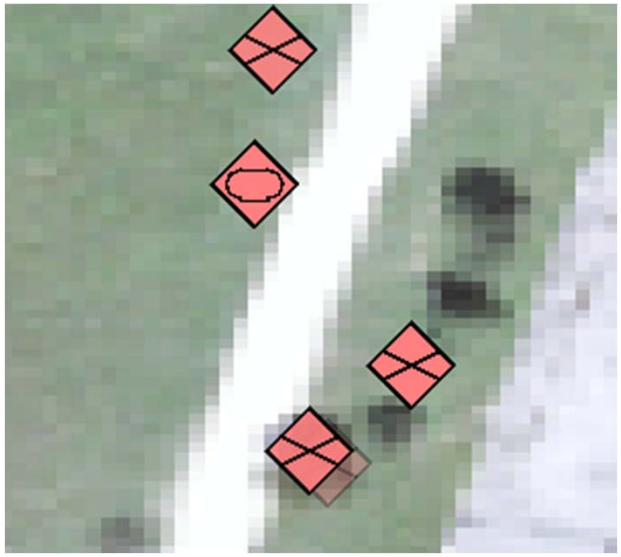
This kind of reports allows also the exchange of information which is not measured by sensors but have a similar format. This can be e. g. the remaining fuel of a battalion.

Example:
[information report] own fuel 3InfBtl 50 percent ongoing at 20101211124322.456 RPTFCT info-report0145;

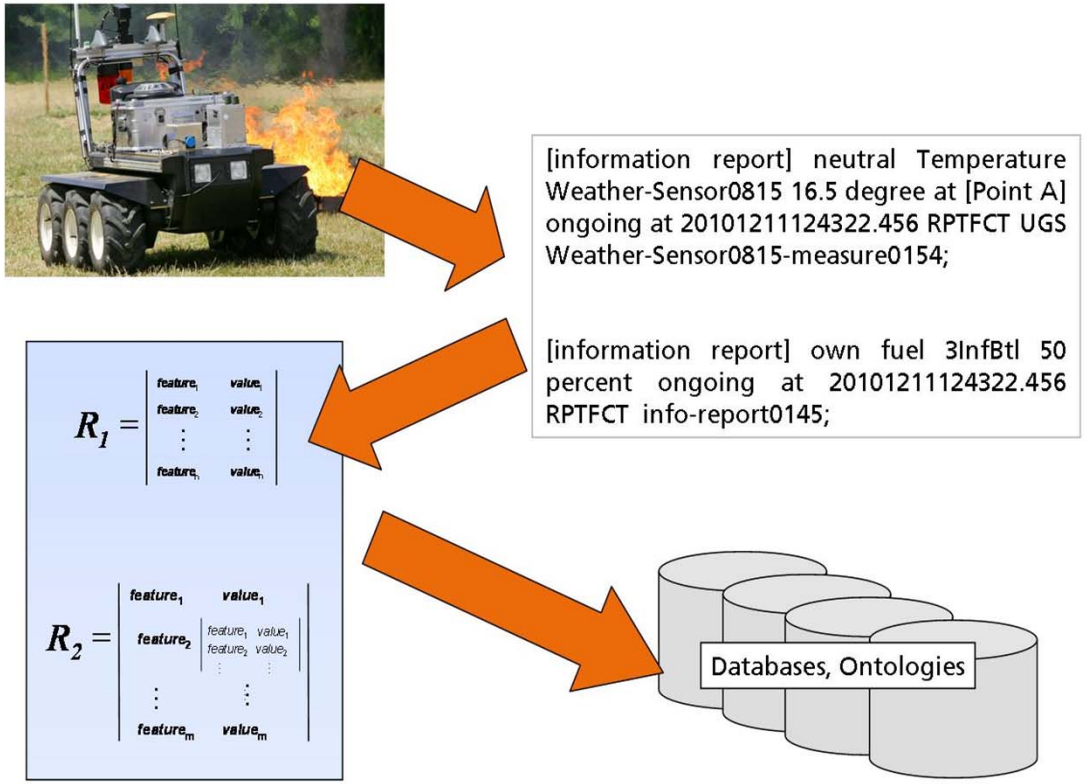
BML as a lingua franca for fusion Reports from robot swarm

- Reports are also expressed on “high-level”.
- Aggregate data to produce high-level information.
  - Examples: Robot status, Red-Force Tracking



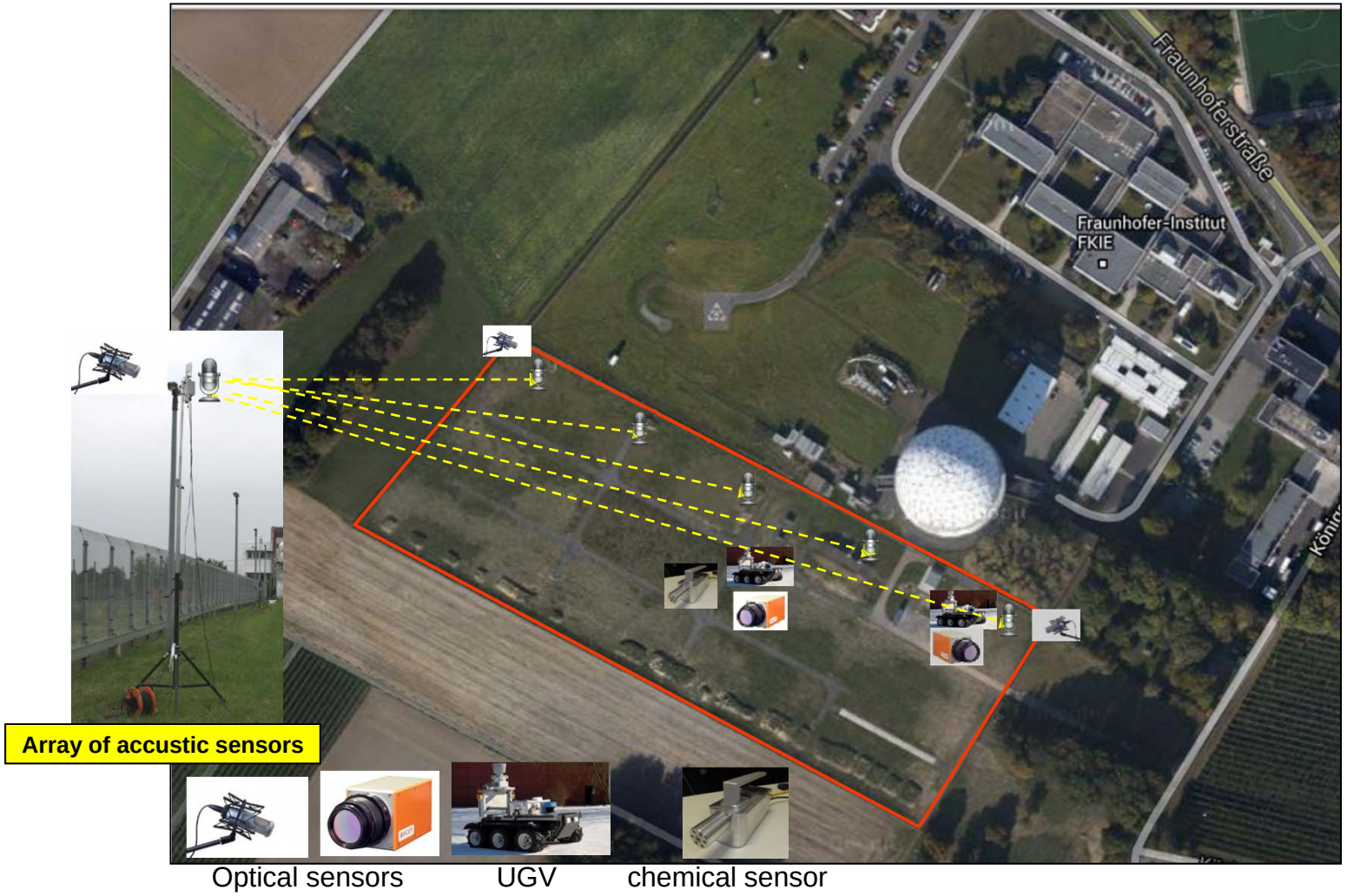


BML as a lingua franca for fusion



Robot sensor readings are reported as BML statements, stored as feature value matrices.

Experimental area at Fraunhofer FKIE



IST - 106 Live Experiment

Application domain: camp / border / infrastructure protection

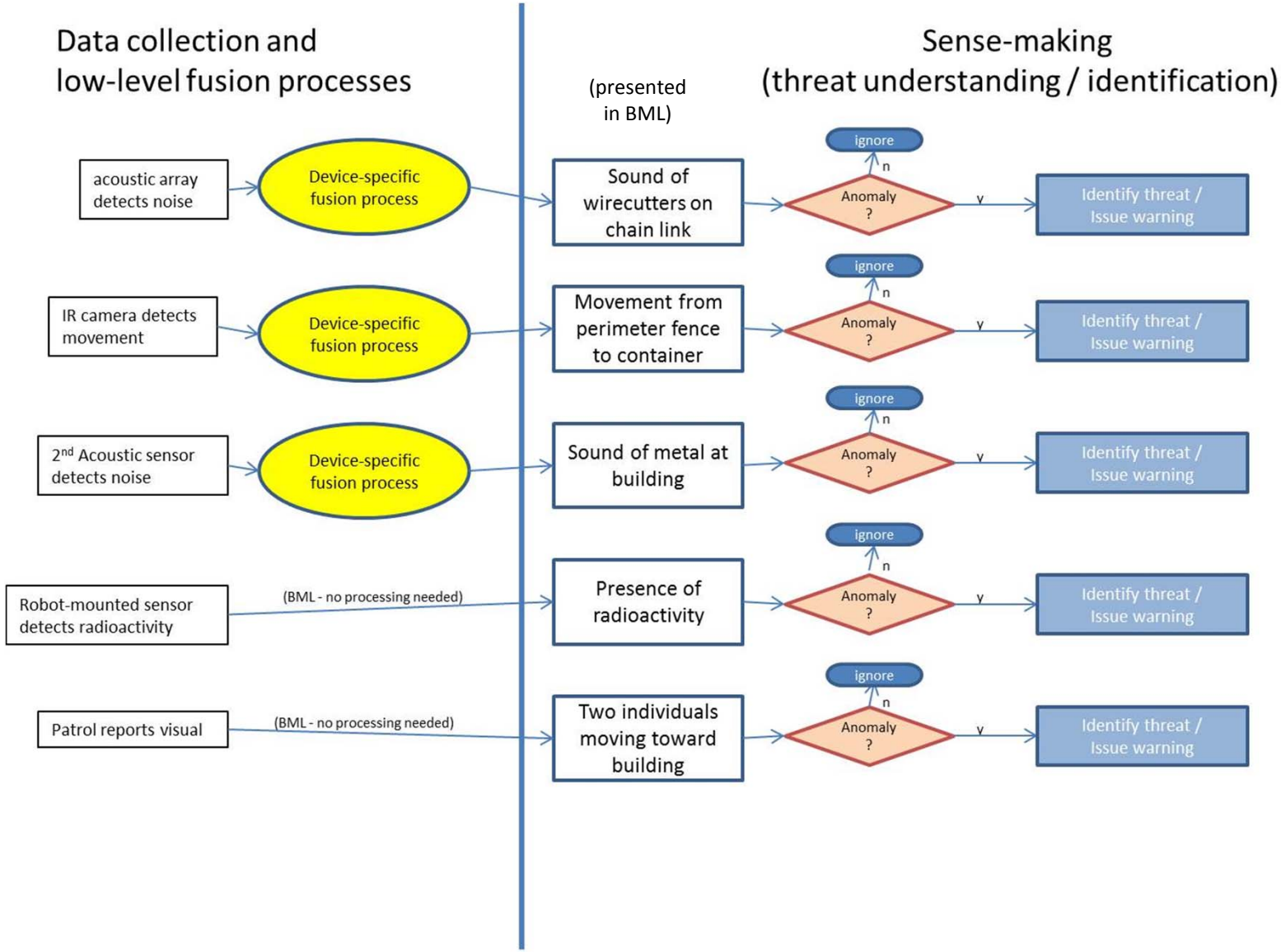
Scenario: Successful breach through a fence, intrusion and position of bomb

General fusion functionality / processing modules /exploitation steps

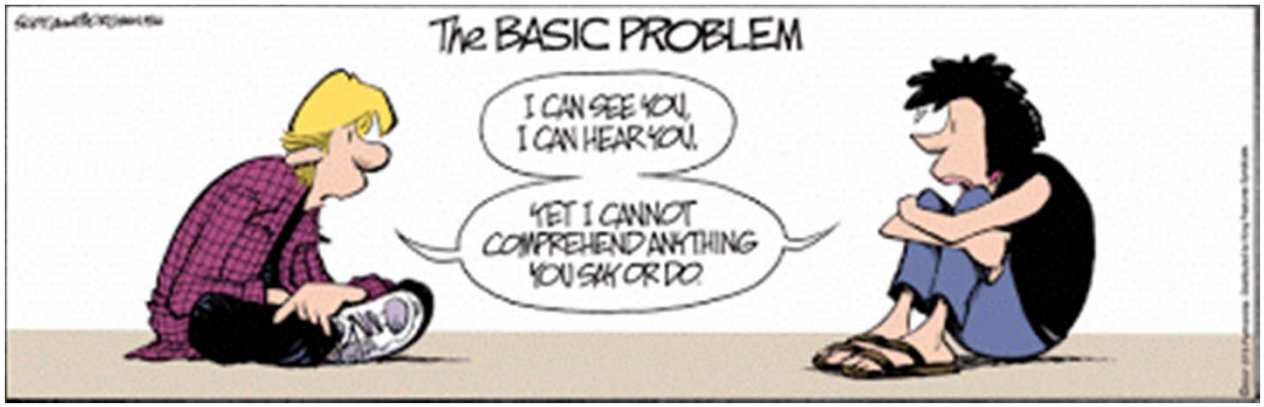
- Detection, localization, classification and tracking of the sources
- Fusion: network of acoustic sensors, fusion of AcINT and ImINT and HUMINT
- Resource management: direct imagery sensor, send out UGV (robot)
- Display situational information

Live Experiment: Aspects for perimeter surveillance

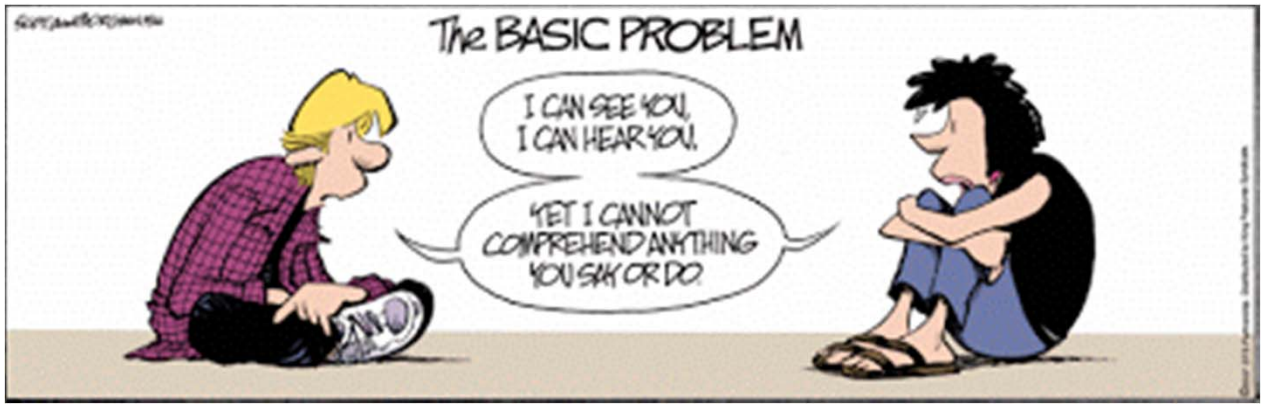
- Acoustic event detection
- Camp protection, UGV patrols
- Anomalous event, unpermitted approach
- Network /array of audio sensors
- localization and tracking
- event detector and classifier (optical sensors)
- detection of hazardous material (chemical sensor)


Data collection and
low-level fusion processes
(presented
in BML)
Sense-making
(threat understanding / identification)
acoustic array detects noise
Device-specific fusion process
Sound of wirecutters on chain link
ignore
n
Anomaly ?
y
Identify threat / Issue warning
IR camera detects movement
Device-specific fusion process
Movement from perimeter fence to container
ignore
n
Anomaly ?
y
Identify threat / Issue warning
2nd Acoustic sensor detects noise
Device-specific fusion process
Sound of metal at building
ignore
n
Anomaly ?
y
Identify threat / Issue warning
Robot-mounted sensor detects radioactivity
(BML - no processing needed)
Presence of radioactivity
ignore
n
Anomaly ?
y
Identify threat / Issue warning
Patrol reports visual
(BML - no processing needed)
Two individuals moving toward building
ignore
n
Anomaly ?
y
Identify threat / Issue warning

A major goal of the presentations on issues dealing with natural language processing for fusion was to increase understanding between the fusion communities…



A major goal of the presentations on issues dealing with natural language processing for fusion was to increase understanding between the fusion communities…



Did we succeed??

- Questions ?