

2004 Command and Control Research and Technology Symposium
"The Power of Information Age Concepts and Technologies"
June 15-17, 2004
Loews Coronado Bay Resort
San Diego, California

Topic: Information Operations / C2 Decisionmaking

Paper's Title: How to Represent the Content of Free-form Battlefield Reports

Author

Name: Dr. Matthias Hecking
Organization: FGAN/FKIE
Address: Neuenahrer Straße 20
53343 Wachtberg-Werthhoven
Germany
Phone: +49 228 9435 576
Fax: +49 228 9435 685
E-Mail: hecking@fgan.de

How to Represent the Content of Free-form Battlefield Reports

Dr. Matthias Hecking
FGAN/FKIE
Neuenahrer Straße 20
53343 Wachtberg-Werthoven
Germany
Phone: +49 228 9435 576
Fax: +49 228 9435 685
hecking@fgan.de

Abstract

The content extraction from free-form texts is necessary for the information operations in the NCW concept. Thus, the information domain of NCW needs knowledge-based technologies for the detection and analysis of the relevant information. Computational linguistics offers with the information extraction such a technology. This approach processes huge amount of texts of a specific sort. The result of the extraction is a formal description of the content of each natural language text. The design and implementation of this formal description is an important step in the development of the information operations. In this paper, a formal description of the meaning of free-form battlefield reports is presented. After an introduction to the information extraction technology and the description of the SOKRATES project, the approach to formally represent the battlefield reports is described in detail.

Introduction

The *processing of human language* is identified as a critical capability in many future military applications (cf. [Steeneken, 1996]). Especially the *content extraction* from free-form texts is important for any information operation of the NCW concept (s. [NCW, 2001], p. 5-15). *Information extraction* (IE) is an engineering approach for content extraction from free-form texts based on results of computational linguistics. IE allows to build systems that process huge amount of texts of a specific sort. Each IE system is tailored to a specific domain and task. IE uses a *shallow syntactic approach* (cf. [Hecking, 2003b]), i.e. that only parts of the sentences (so-called ‘chunks’) are processed with finite state automatons or transducers. These transducers identify the relevant information in the free-form text. A focal point in designing an IE system is the definition of the *templates*. These templates are the formal representation of the meaning of natural language texts.

In our project SOKRATES we apply IE to the analysis of German free-form battlefield reports. Our prototype is able to process simple battlefield reports. The templates were realized through *feature structures*. These feature structures describe all the objects, actions, etc., which can be found in the battlefield reports. The design of these representation structures is crucial to the success of the IE.

In this paper, we will first give a short introduction into the field of IE. Then, we describe our research project SOKRATES. The SOKRATES prototype is able to process battlefield reports in German. In the main part of the paper, we give a thorough description of the structures needed to represent the content of the military reports.

Information Extraction

During the *information extraction* (IE) relevant information about the Who, What, When, etc. in natural language texts is identified, collected, and normalized (cf. [Appelt, 1999], [Pazienza, 1999]). The relevant information is described through patterns called *templates*. These domain and task specific templates represent the meaning of the relevant information. During the IE task the templates are filled with the extracted information. Therefore, IE can be seen as the process of normalizing free-form text into a defined semantic structure.

To realize an IE system, language resources (lexicon, grammar) and appropriated parsing software are necessary. This software must be language-specific. Thus, the IE tools for the English language are not appropriated for analyzing German texts (due to the free-order of the language).

In order to achieve robust and efficient IE systems, domain knowledge must be integrated and shallow algorithms must be used. The domain knowledge is tightly integrated with the language knowledge, e.g., the name 'Leopard' in the lexicon has the categorical information 'tank'. This association between words and semantic information is domain-specific and has to be change for other applications.

The IE process itself is divided into sub steps. After tokenizing the text, the sentence boundaries must be identified. Then, the morphological component identifies the word stems, the abbreviation, and detects the syntactic information (e.g., grammar case and gender). After this, the chunk parsing with transducers selects parts of the natural language text that are relevant for the specific information extraction task. The chunks are then used to instantiate the templates, which represent the result of the IE process.

The instantiated templates are a *formal description of the meaning* of the texts. In our research project SOKRATES we use IE to analyze German battlefield reports. For this, we use typed feature structures as templates. For more details about information extraction from battlefield reports refer to [Hecking, 2003a] and [Hecking, 2004].

The Project SOKRATES

The overall objective of the SOKRATES project is to analyze written German battlefield reports. To do this, we apply the IE to free-form battlefield reports (cf. [Casals, 2004a], [Casals, 2004b], [Frey, 2004], [Hecking, 2001], [Hecking, 2002], [Hecking, 2003a], [Hecking, 2003b], [Hecking, 2004], [Schade, 2003a], [Schade, 2003b]). The result of the IE is semantically enhanced with the help of an ontology. The result of this semantic processing is stored in the LC2IEDM database (cf. [ATCCIS, 2002]). These stored results can be used for different purposes. One purpose is

that location changes of units initiate automatically changes of tactical symbols on the tactical map.

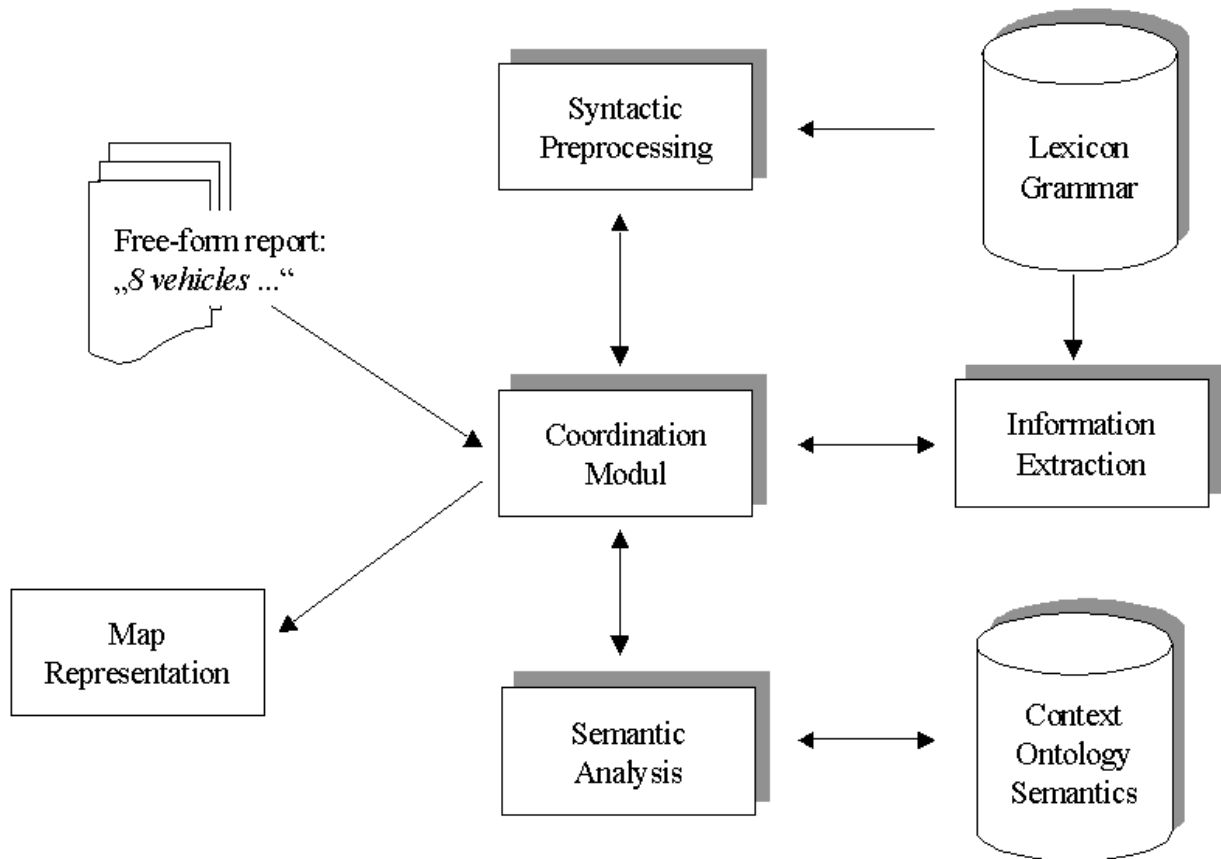


Figure 1: The architecture of the SOKRATES system

Our prototype is able to process simple battlefield reports like: “10.20 Uhr 18 Fahrzeuge marschieren bei Straßenkreuzung (CQ 072368) südlich MILESKIJ (CQ 0737) nach Norden.“ (“10:20 o'clock, 18 vehicles march at road crossing (CQ 072368) south of MILESKIJ (CQ 0737) to the north.”). The templates were realized through *feature structures*. These feature structures describe all the objects, actions, etc. that can be found in the natural language text of the battlefield reports. Feature structures are the chosen formalism to represent the natural language meaning.

The architecture of the SOKRATES prototype is shown in Figure 1 and a screenshot in Figure 2. The free-form reports are handed over to the *coordination module*, which is responsible for all the coordination in the system. In a first step, the *syntactic preprocessing* identifies the sentence boundaries. Next, the *information extraction* module uses the lexicon and the grammar transducers to identify and select the relevant parts in the natural language text. These parts are represented as typed feature structures and are coded in XML. The result of the information extraction is used by the *semantic analysis* component to deduce more information out of the extracted information with the help of an ontology and the context. After the semantic analysis the result is pushed into the LC2IEDM database and then it is used to alter automatically the position of tactical symbols on the map. For more details about the SOKRATES project see the above-cited reports.

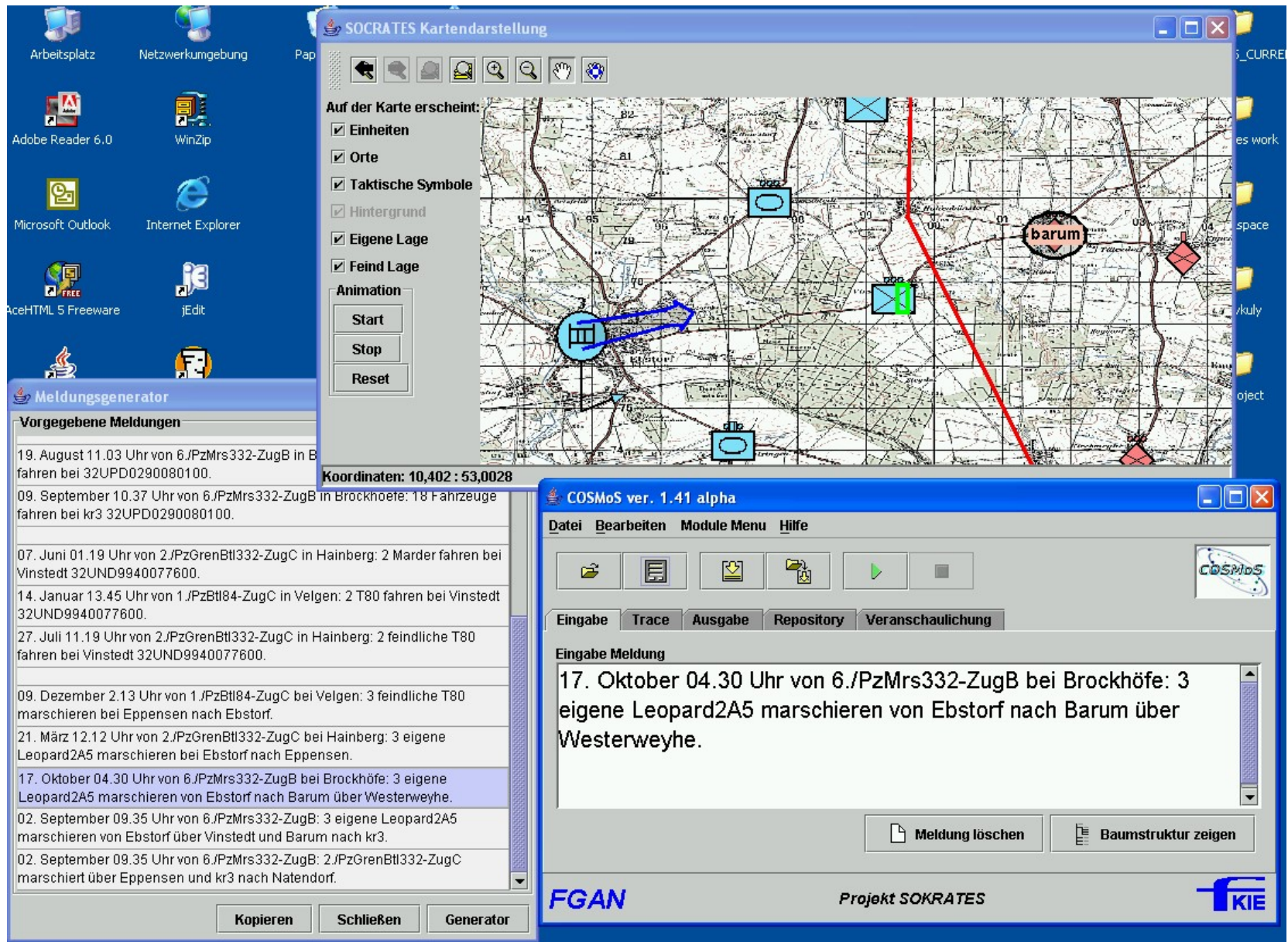


Figure 2: The SOKRATES prototype

The Representation of Meaning

In this section we describe in detail, how the meaning of free-form battlefield reports can be formalized. The meaning is represented by *typed feature structures*. The approach presented here was developed for the content extraction in the SOKRATES project. The general feature structure formalism is widely used in computational linguistics, e.g., in the HPSG approach (cf. [Pollard, 1994]).

Each feature structure consists of an unordered quantity of *features* and a *type declaration* (see Figure 3).

```
(
  (:TYPE . :MILITARY-ITEM)
  (:LOCATION
    (
      (:COORDINATES . "32upd0290080100")
      (:TYPE . :POINT)
    )
  )
  (:QUALIFIER . :AT)
)
```

Figure 3: Example of a feature structure

In Figure 3 the type declaration is given by (:TYPE . :MILITARY-ITEM). All other bracketed terms within the highest pair of brackets represent the set of the features. Each feature consists of a name (e.g., :QUALIFIER) and a value (e.g., :AT). In the simplest case the value is a number, a character string or an atomic type. In addition, a feature value can recursively be a feature structure, e.g., the value of the feature :LOCATION is a feature structure of the type :POINT.

If the IE processes a free-form battlefield report, the feature structures result from two sources. They can be part of a lexicon entry or they are constructed by the transducers during the IE process.

Application-specific Feature Structures

During the development of an IE system an important development step is the determination and definition of the used feature structures. These structures are used to represent the meaning of the natural language texts. Because these structures can be based on each other, they are arranged in a hierarchy. In Figure 4, a part of the feature structure hierarchy is shown that is used in the SOKRATES IE module (the whole hierarchy can be found in [Hecking, 2004]). The topmost feature structure is of the type **feature-structure**. This structure has no features. A possible subtype of **feature-structure** is the feature type **object** ("A → B" means "A is a supertype of B."). This type has also various subtypes; one of them is the type **equipment**. **equipment** includes the features that are listed in square brackets under the type name. For each feature the name (e.g., hostility) and the feature type of the possible feature values (e.g., *hostility-type*) are given. The features are inherited down the hierarchy. So the types **vehicle**, **weapon** and also **antitank** have the same features as **equipment**. Some of the types are atomic and have no features, e.g., all subtypes of **nationality-code**. If features are

defined for a subtype, these features extend the set of the inherited features. If the type of the feature value is enclosed by "{ ... }", the feature value consists of a set of feature structures of the indicated type. If the type is enclosed by "< ... >" the value consists of an ordered list of feature structures.

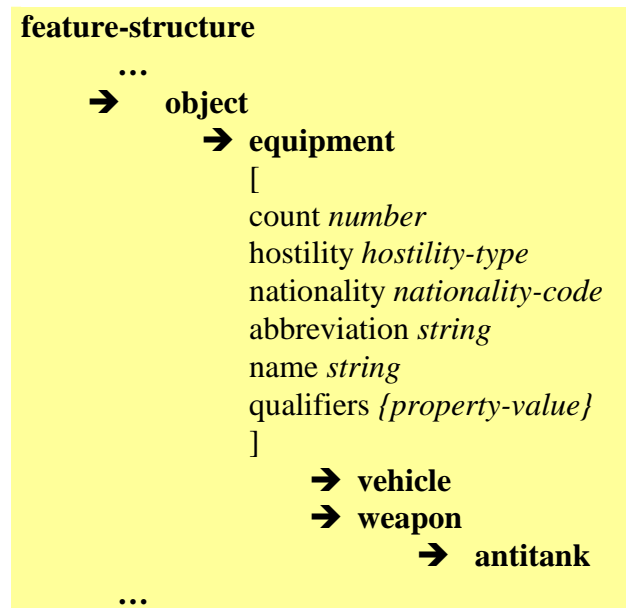


Figure 4: Part of the feature structure hierarchy

The whole military report is formalized by the type **report** (see Figure 5). In feature "addressee" the receiver of the report is described. For this, the feature value must be of the type *partner*. The feature "medium" shows how the report was transmitted (e.g., in written form). The actual content of the report is the value of the feature "message". This value can comprise a set of action descriptions. Who delivers the report is given by feature "speaker". When the report is given, is represented by feature "time" and "credibility" is a description of the credibility of the whole report.

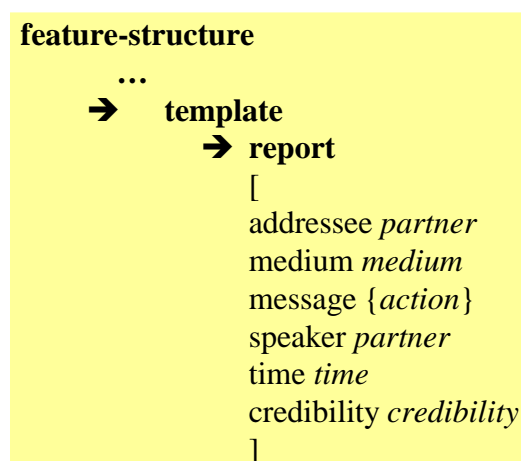


Figure 5: Feature structure for battlefield reports

The addressee and the speaker of the military report can have feature values of the type **partner** (see Figure 6). This type has the subtypes **multiple-partner** and **single-partner**. The subtypes **person** and **unit** of **single-partner** are used to represent single persons or units.

multiple-partner is necessary if both the reporting unit (feature "second-partner") and the superior unit (feature "first-partner") are mentioned.

The content of a report is represented by a feature structure that is one of the subtypes of the type **action**. In the current version of the application specific feature structures only the action type **move** (see Figure 7) is well-elaborated. This type is used to describe movements on the battlefield. The basic structure of **move** was adopted from the FrameNet project (cf. [Motion, 2003]) and adapted to the military requirements. During a movement, objects are moved. These objects are represented as feature values of feature "theme". The objects are moving from a starting point (source *military-item*) to an end point (goal *military-item*) using a path (path <*military-item*>). The objects can also be in an area (area {*military-item*}). The instrument (carrier) of the movement can also be given (but this is not totally worked out in the current version).

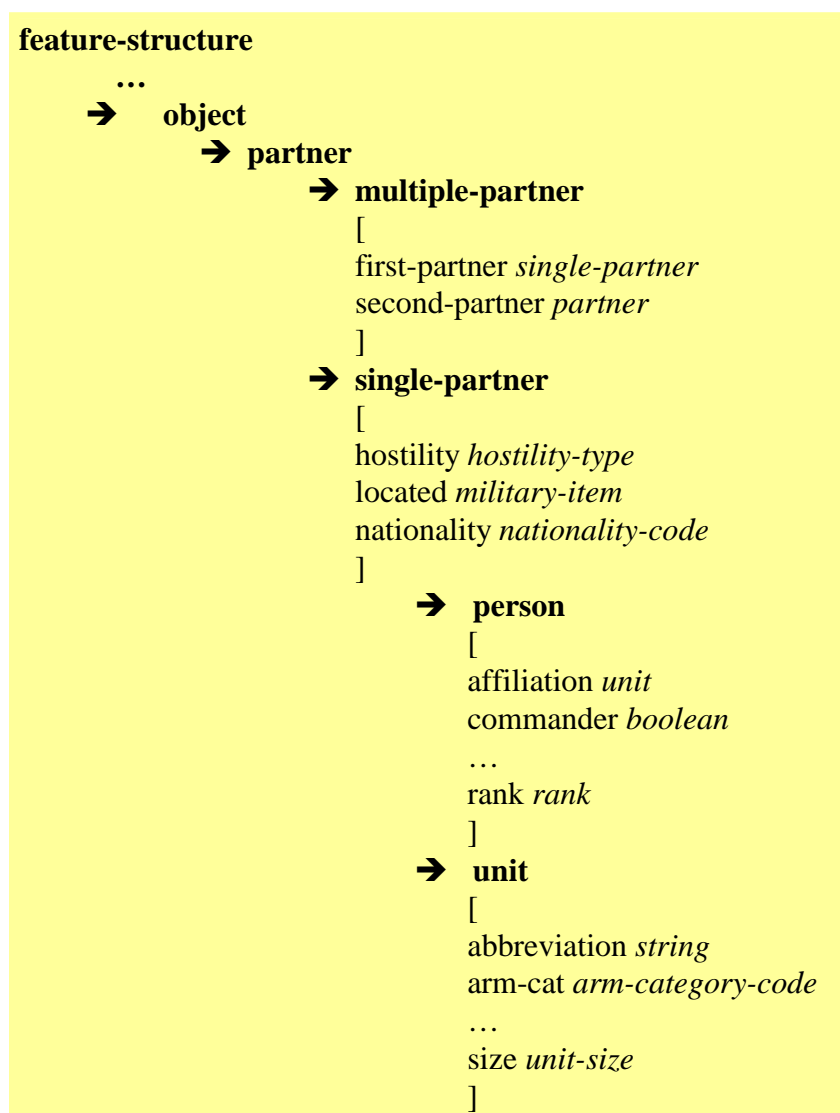


Figure 6: Feature structure 'partner'


```

move
[
  qualifiers {property}
  area {military-item}
  carrier undef
  distance number
  duration number
  goal military-item
  path <military-item>
  source military-item
  speed number
  start-time time
  theme theme
]

```

Figure 7: Feature structure 'move'

The feature "qualifiers" contains descriptions of adverbial phrases of the movement (e.g., "it moves *fast*"), if the descriptions are not contained in other features (e.g., in "goal <*military-item*>"). Subtypes of **verb-property-value** code the content of adverbial phrases.

```

feature-structure
  ...
  → object
    → equipment
      [
        count number
        ...
        qualifiers {property-value}
      ]
      → vehicle
      → weapon
      ...
    → military-item
      [
        ...
      ]
      → facility
        → airfield
        → barrier
      ...
    → partner
      → multiple-partner
        [
          first-partner single-partner
          second-partner partner
        ]
      → single-partner
      ...

```

Figure 8: Subtype of 'object'

In some of the feature values of the type **move**, structures of the subtypes of **object** are used. In Figure 8 these subtypes are shown. Ordnance can be found under **equipment**. Other military objects and concepts (e.g., airfield, barrier) are subtypes of **military-item**.

In order to describe movements, concepts of locations and directions are necessary. Directions are formalized by the type **direction** (see Figure 9), locations by **location** (see Figure 10) and its subtypes.

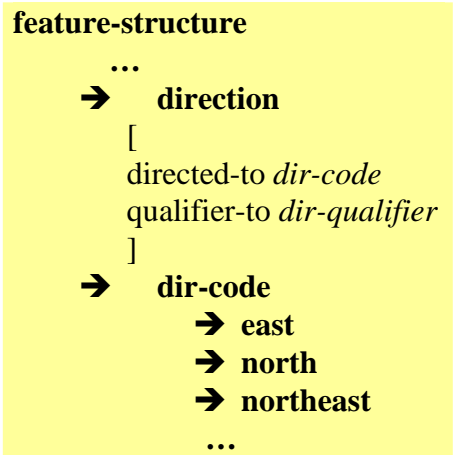


Figure 9: Formalization of directions

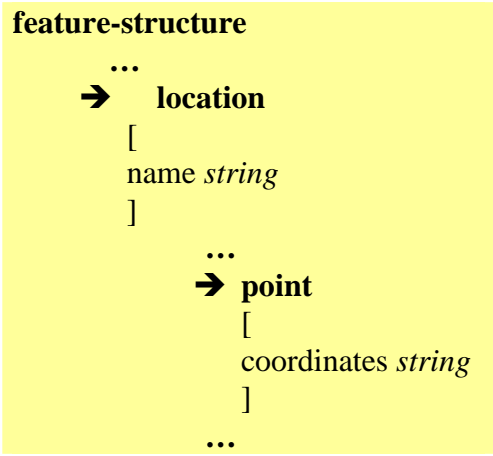


Figure 10: Formalization of locations

XML-format of the Feature Structure

The above examples of the feature structures are all given in the internal format. These structures can also be coded in XML. For this purpose, an XML schema was developed. A graphical representation of this schema is shown in Figure 11. The XML format allows any feature structure to be coded.

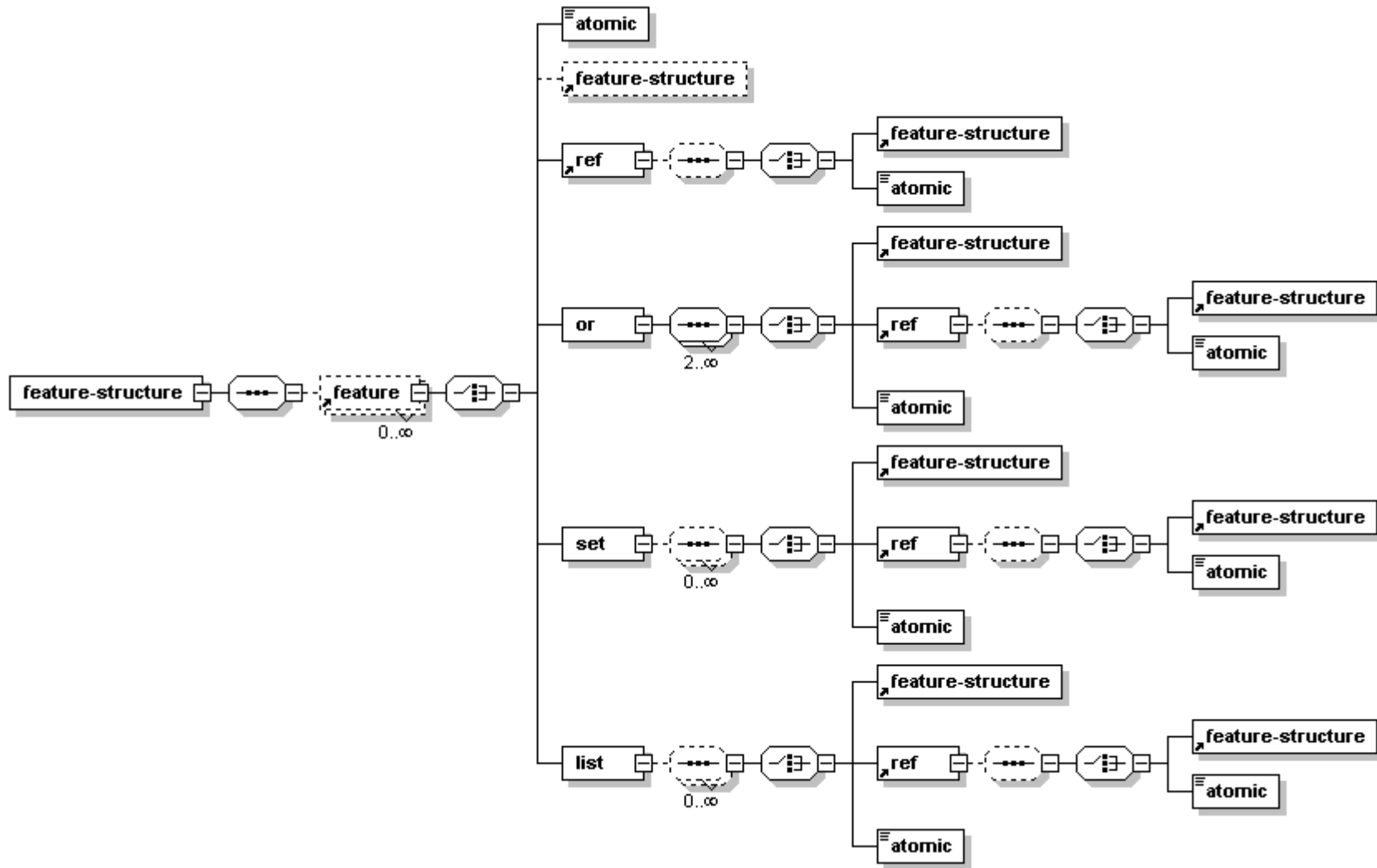


Figure 11: XML schema for coding the feature structures

Representation of a Battlefield Report

In the following, we show how a whole battlefield report is represented with the introduced feature structures. The report is:

"09.September 10.45 Uhr von VN - Militärbeobachtern in BIJELJINA: 2 BRDM 2 und 1 PT 76 durchfahren in rascher Fahrt auf Straße 14 – 1 die Ortschaft SULJIN HAN (CQ 5458) nach Westen."

("September 9, 10:45 o'clock, from UN military observer in BIJELJINA: 2 BRDM 2 and 1 PT 76 pass through the locality SULJIN HAN (CQ 5458) on the road 14 – 1 in fast travel to the west.")

Each report consists of the prolog

"09.September 10.45 Uhr von VN - Militärbeobachtern in BIJELJINA:"

and the actual matter

"2 BRDM 2 und 1 PT 76 durchfahren in rascher Fahrt auf Straße 14 – 1 die Ortschaft SULJIN HAN (CQ 5458) nach Westen."

In the tables of Figure 12, Figure 13 and Figure 14 the different feature structures are shown, which are produced from various parts of the text. In these tables is specified,

- which part of the report induces
- which syntactical phrase and
- the constructed feature structure (Feature structure A).

Also shown is:

- the syntactical function (not for the prolog) and
- the feature structure that will be part of the **move**-structure (Feature structure B).

The prolog has a fixed structure. After the date ("*09. September 10.45 Uhr*"), the speaker ("*von VN-Militärbeobachtern*") and the position of the speaker ("*in BIJELJINA*") are listed. The date results in a feature structure of the type :TIME, the speaker in a feature structure of the type :UNIT and the position in a structure of the type :TOWN. In the **move** feature structure the structure :TIME is stored as a value of the :TIME feature. The :TOWN structure is stored as the value of the :LOCATED feature in the :UNIT feature structure. The :UNIT becomes the value of the :SPEAKER feature in the **move** structure.

The part "*2 BRDM 2 und 1 PT 76*" is a nominal phrase (NP), which is constructed from two smaller phrases and a conjunction. For each of these phrases a feature structure of the type :TANK is produced. The whole NP is the subject of the sentence. Both **tank** feature structures are stored as values of the *objects* feature in the **move** structure.

Report part	<i>09.September 10.45 Uhr</i>	<i>von VN - Militärbeobachtern</i>	<i>in BIJELJINA:</i>
Phrases	Date	Reporting unit/person	Position of reporting unit/person
Feature structure A	((:TYPE . :TIME) (:MINUTE . 45) (:HOUR . 10) (:MONTH . 9) (:DAY . 9))	((:TYPE . :UNIT) (:NAME . "VN-Militärbeobachter") (:ABBREVIATION . "VN-MilBeob"))	((:TYPE . :TOWN) (:NAME . "bijeljina") (:QUALIFIER . :EXACTLY-AT))
Feature structure B	(:TIME ((:TYPE . :TIME) (:MINUTE . 45) (:HOUR . 10) (:MONTH . 9) (:DAY . 9)))	(:SPEAKER ((:TYPE . :UNIT) (:NAME . "VN-Militärbeobachter") (:ABBREVIATION . "VN-MilBeob") (:LOCATED ((:TYPE . :TOWN) (:NAME . "bijeljina") (:QUALIFIER . :EXACTLY-AT)))))	

Figure 12: Example – I

Report part	2 BRDM 2	und	1 PT 76	durchfahren	in rascher Fahrt
Phrases	NP ₁	Conj	NP ₂	VP	PP
	NP				
Feature structure A	((:TYPE . :TANK) (:ABB . "brdm-2") (:COUNT . 2))	:SET	((:TYPE . :TANK) (:ABB . "pt76") (:COUNT . 1))	(:TYPE . :MOVE)	((:VALUE . :FAST) (:TYPE . :PROPERTY))
Syntactical function	Subject			Predicate	Adverbial phrase
Feature structure B	(:OBJECTS (:SET ((:TYPE . :TANK) (:ABBREVIATION . "brdm-2") (:COUNT . 2)) ((:TYPE . :TANK) (:ABBREVIATION . "pt76") (:COUNT . 1))))			(:TYPE . :MOVE)	(:QUALIFIERS (:SET ((:VALUE . :FAST) (:TYPE . :PROPERTY))))

Figure 13: Example – II

Report part	<i>auf Straße 14 - 1</i>	<i>die Ortschaft SULJIN HAN (CQ 5458)</i>	<i>nach Westen</i>	.
Phrases	PP	NP	PP	
Feature structure A	((:TYPE . :WAY) (:NAME . "14-1") (:QUALIFIER . :ON))	((:LOCATION ((:COORDINATES . "cq5458") (:TYPE . :POINT))) (:NAME "suljin han") (:TYPE . :TOWN))	((:QUALIFIER . :TOWARDS) (:DIRECTED-TO . :WEST) (:TYPE . :DIRECTION))	
Syn. fct.	Adverbial phrase	Object	Adverbial phrase	
Feature structure A	(:AREA (:SET ((:TYPE . :WAY) (:NAME . "14-1") (:QUALIFIER . :ON)) ((:LOCATION ((:COORDINATES . "cq5458") (:TYPE . :POINT))) (:NAME "suljin han") (:TYPE . :TOWN)))))		(:GOAL ((:QUALIFIER . :TOWARDS) (:DIRECTED-TO . :WEST) (:TYPE . :DIRECTION)))	

Figure 14: Example – III

The verb *"durchfahren"* ("to pass through") forms the verbal phrase (VP). This VP is the predicate of the sentence. It is the feature type (:TYPE . :MOVE) of the VP, which is responsible for selecting the correct **action** subtype, in this example the **move** feature structure.

The prepositional phrase (PP) *"in rascher Fahrt"* ("in fast travel") produces a feature structure of the type :PROPERTY. This PP is an adverbial phrase, which specifies in more detail the VP (the movement is fast, not slow). The feature structure of this PP is stored in the *qualifiers* feature of the **move** structure.

The prepositional phrase *"auf Straße 14 - 1"* ("on the road 14 – 1") produces a feature structure of the type :WAY. It is also an adverbial phrase and it is stored in the *area* feature of the **move** structure.

The :TOWN-structure is also stored in the *area* feature. This structure results from the report part *"die Ortschaft SULJIN HAN (CQ 5458)"* ("the locality SULJIN HAN (CQ 5458)").

"nach Westen" ("to the west") produces a feature structure of the type :DIRECTION. This is stored in the *goal* feature.

The filled **move** feature structure is the feature value of the :MESSAGE-feature of the **report** structure. The whole produced feature structure for the example is shown in Figure 15.

```
(
  (:CREDIBILITY . :TRSTED)
  (:TIME
    (
      (:TYPE . :TIME)
      (:MINUTE . 45)
      (:HOUR . 10)
      (:MONTH . 9)
      (:DAY . 9)
    )
  )
  (:SPEAKER
    (
      (:TYPE . :UNIT)
      (:NAME . "VN-Militärbeobachter")
      (:ABBREVIATION . "VN-MilBeob")
      (:LOCATED
        (
          (:TYPE . :TOWN)
          (:NAME . "bijeljina")
          (:QUALIFIER . :EXACTLY-AT)
        )
      )
    )
  )
)
```



```

)
(:MESSAGE
  (:SET
    (
      (
        (:AREA
          (:SET
            (
              (
                (:TYPE . :WAY)
                (:NAME . "14-1")
                (:QUALIFIER . :ON)
              )
              (
                (:LOCATION
                  (
                    (:COORDINATES . "cq5458")
                    (:TYPE . :POINT)
                  )
                )
                (:NAME . "suljin han")
                (:TYPE . :TOWN)
              )
            )
          )
        )
      )
    )
  (:QUALIFIERS
    (:SET
      (
        (
          (:VALUE . :FAST)
          (:TYPE . :PROPERTY)
        )
      )
    )
  (:TYPE . :MOVE)
  (:GOAL
    (
      (:QUALIFIER . :TOWARDS)
      (:DIRECTED-TO . :WEST)
      (:TYPE . :DIRECTION)
    )
  )
  (:THEME
    (
      (:OBJECTS

```

```

      (:SET
      (
      (
      (:TYPE . :TANK)
      (:ABBREVIATION . "brdm-2")
      (:COUNT . 2)
      )
      (
      (:TYPE . :TANK)
      (:ABBREVIATION . "pt76")
      (:COUNT . 1)
      )
      )
      )
      (:TYPE . :THEME)
    )
  )
  (:MEDIUM . :LETTER)
  (:ADDRESSEE
  (
  (:TYPE . :UNIT)
  )
  )
  (:TYPE . :REPORT)
)

```

Figure 15: The feature structure of the example

For more information concerning the feature structures and the information extraction refer to [Hecking, 2004].

Conclusion

In this paper, we first gave a short introduction into the promising field of information extraction. Second, we described our research project SOKRATES. In the main part of the paper, we showed how German free-form battlefield reports are formalized by feature structures. For this, we described the hierarchy of feature structures and applied the structures to an example. The shown representation formalism can be used for other languages as well. Next, we will extend the hierarchy, so that more **action** subtypes (e.g., **attack**) are modelled.

References

- [ATCCIS, 2002] ATCCIS WP 5-7, Edition 5.0. *Overview of the Land C2 Information Exchange Data Model (LC2IEDM)*. NATO: ATCCIS PWG, 2002.
- [Appelt, 1999] Appelt, D. & Israel, D. *Introduction to Information Extraction Technology*. Stockholm: IJCAI-99 Tutorial, 1999, <http://www.ai.sri.com/~appelt/ie-tutorial/>.
- [Casals, 2004a] Casals Elvira, X. *Project SOKRATES: Interaction with the LC2IEDM Database*. FKIE-Bericht, Wachtberg: Forschungsgesellschaft für Angewandte Naturwissenschaft e.V., to appear, 2004.
- [Casals, 2004b] Casals Elvira, X. *Project SOKRATES: Processing of Headers for the Information Extraction Component*. FKIE-Bericht, Wachtberg: Forschungsgesellschaft für Angewandte Naturwissenschaft e.V., to appear, 2004.
- [Frey, 2004] Frey, M. L. & Schade, U. *Modular Framework for Military Report Processing*. FKIE-Bericht. Wachtberg: Forschungsgesellschaft für Angewandte Naturwissenschaft e.V., FKIE-Bericht Nr. 73, 2004.
- [Hecking, 2001] 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.
- [Hecking, 2002] Hecking, M. *Analysis of Spoken Input to C2 Systems*. In: Proceedings of the 7th International C2 Research and Technology Symposium (ICCRTS), Québec City, Kanada, 2002.
- [Hecking, 2003a] Hecking, M. *Information Extraction from Battlefield Reports*. In: Proceedings of the 8th International Command and Control Research and Technology Symposium (ICCRTS), Washington, DC, U.S.A., 2003.
- [Hecking, 2003b] Hecking, M. *Analysis of Free-form Battlefield Reports with Shallow Parsing Techniques*. Paper presented at the RTO IST Symposium on „Military Data and Information Fusion“, held in Prague, Czech Republic, October 20-22, 2003, Best Paper Award.
- [Hecking, 2004] Hecking, M. *Informationsextraktion aus militärischen Freitextmeldungen*. FKIE-Bericht, Wachtberg: Forschungsgesellschaft für Angewandte Naturwissenschaft e.V., FKIE-Bericht, to appear, 2004.
- [NCW, 2001] Department of Defense. *Network Centric Warfare – Report to Congress*. 27 July 2001.
- [Pazienza, 1999] Pazienza, M. T. (ed.) *Information Extraction*. Berlin, 1999.
- [Pollard and Sag, 1994] Pollard, C. & Sag, I. A. *Head-Driven Phrase Structure Grammar*. The University of Chicago Press, Chicago, London, 1994.
- [Schade, 2003a] Schade, U. *Ontologieentwicklung für Heeresanwendungen*. Forschungsgesellschaft für Angewandte Naturwissenschaften e.V. (FGAN), FKIE-Bericht Nr. 57, 2003.
- [Schade, 2003b] Schade, U. *Towards an Ontology for Army Battle C2 Systems*. In: Proceeding of the 8th ICCRTS, 2003.
- [Steeneken, 1996] Steeneken, H. J. M. *Potentials of Speech and Language Technology Systems for Military Use: an Application and Technology Oriented Survey*. NATO, Technical Report, AC/243(Panel 3)TP/21, 1996.