12th ICCRTS
"Adapting C2 to the 21st Century"

June 19-21, 2007
Naval War College
Newport, RI
U.S.A.

Paper's Title: The KFOR Text Corpus

Topic: C2 Technologies and Systems; C2 Concepts, Theory, and Policy

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

The KFOR Text Corpus

Dr. Matthias Hecking
FGAN/FKIE
Neuenahrer Straße 20
53343 Wachtberg-Werthhoven
Germany

hecking@fgan.de

Abstract

The new deployments of the German Federal Armed Forces necessitate analysis of large quantities of HUMINT reports. The realized ZENON system uses an information extraction approach for the (partial) content analysis of English HUMINT reports from the KFOR deployment of the Bundeswehr. More than 4,000 military reports from this deployment were used as a starting point for the realization of the ZENON prototype. From these reports 800 were manually annotated and form the *KFOR Text Corpus*. This corpus is a specialized micro-text corpus, which contains the syntactic and semantic annotations in different layers. In this paper, the KFOR Text Corpus and its use in the evaluation and the improvement of the ZENON system are presented. After a short introduction, an explanation is given why corpora are needed for the evaluation of natural language processing systems. In the main body of the paper, the KFOR Text Corpus and its use for the evaluation of the ZENON system is described in detail. First, the different annotation layers and annotation types are presented. The corpus structure is also explained. Finally, the use of the corpus to evaluate and improve the ZENON system is shown. Various examples are given.

1. Introduction

The *processing of human language* was identified as a critical capability in many future military applications (cf. [Steeneken, 1996]). Especially the *content analysis* of free-form texts is important for any information operation of the Network Centric Warfare (NCW) concept (s. [NCW, 2001], p. 5-15). The content analysis can be realized through *Information Extraction* (IE) which is a natural language processing technique (cf. [Appelt, 1999], [Hecking, 2004a]).

We set up the *research project ZENON*[1], in which the information extraction approach is used for the (partial) content analysis of English HUMINT reports from the KFOR deployment of the Bundeswehr (cf. [Hecking, 2006a], [Hecking, 2006b], [Hecking, 2005a], [Hecking, 2003a], [Hecking, 2003b], [Hecking, 2004a], [Hecking, 2004b]). The overall objective of this research is to create a *graphically navigatable Entity-Action-Network*. The information about the actions and named entities are identified from each sentence and the content of the sentences are formally represented. These formal representations can be combined and presented in the navigatable network.

After creating natural language processing systems the efficiency of such systems must be evaluated. According to the state of the art this is accomplished by a comparison of the *produced annotations* with the *expected annotations*. The expected annotations are given by

[1] according to: Zenon of Citium, 336 BC - 264 BC, philosopher, founder of the Stoicism

a *corpus*, i.e. a set of texts and associated annotations. Depending on the text sort and the analysis objectives different syntactic (e.g. part-of-speech, conjugation information, structure of nominal phrases...) and/or semantic (e.g. name of cities, rivers, countries...) annotations are needed.

To develop, to evaluate and to improve the information extraction components of the ZENON prototype the *KFOR Text Corpus* was realized (cf. [Hecking, 2006c]). Starting with 4,498 military reports (mostly in English) from the KFOR deployment of the German Federal Armed Forces, 800 of them were manually annotated and form the *KFOR Corpus*[2]. The performance of the ZENON information extraction is quantitatively evaluated relative to the KFOR corpus.

This paper is structured as follows. First, a short introduction into corpora and their use for evaluation is given. In the main part of the paper, the KFOR text corpus is described in detail. Various annotation types are presented. Finally, the use of the corpus during the evaluation of the ZENON system is shown.

2. Corpora for Computational Linguistics

A *corpus* is used to carry out empirical research on written or spoken texts (cf. [McEnery, 2001], p. 29ff). It consists of a set of annotated documents that is representative, of finite size and machine-readable, which is used as a standard reference.

Because we are not able to list all texts of a language variety (e.g. all HUMINT reports in English from 1980 to 2000) we have to build a sample of it. This sample must be maximally *representative* of the variety under examination. Most of the corpora are *finite* in size and very huge (e.g. American National Corpus [ANC, 2007], British National Corpus [BNC, 2007]). Nearly all corpora are in *machine-readable* form. This allows searching and manipulating in ways, which are simply not possible with non-machine-readable corpora. This is also the basis to enrich the text with additional information, the *annotations*. There is also often a tacit understanding that a corpus constitutes a *standard reference* for the language variety that it represents.

Corpora exist in two formats: *unannotated* (i.e. only a set of texts form the corpus) or *annotated* (i.e. texts enhanced with various types of linguistic information). Corpora used in computational linguistic research are annotated. E.g., in the sentence "The bomb did not ignite in the station of Koblenz." the additional information `city[40, 47, {name= Koblenz}]` forms the semantic annotation for the string "Koblenz", i.e. this string is the name of a city and the name starts in position 40 and ends in 47.

Depending on the intended use of the corpus, different classes of annotation are used in constructing corpora (cf. [McEnery, 2001], p. 39ff):

- *Textual* or *extra-textual*: Basic information about the nature of the text, e.g., author name, sex of the author, date the text was written, the variety of the language, broad subject domain, etc.
- *Part-of-speech (POS)*: The aim of the part-of-speech tagging is to assign to each token (word, number, punctuation mark) a code indicating its part of speech (e.g. past participle, noun, adjective). This forms the basis for further analysis such as syntactic parsing and semantic annotation.

[2] Since the KFOR corpus is classified, it is not freely available. The report [Hecking, 2006c] is *not* classified.

- *Lemma*: Each word is reduced to its respective lexeme – the word form that one would look up if one were looking for the word in a dictionary.
- *Parsing*: After annotating the part-of-speech it is possible to bring these categories into higher-level syntactic relationships with one another. These descriptions are often called 'treebanks' and constitute the annotation.
- *Semantics*: Two different types of semantic annotation can be identified. The first one describes a semantic relationship between items in the text, e.g. the agent of a particular action. The second one marks semantic features of words, especially word senses.
- *Phonetic transcription*: Corpora of spoken language can also be transcribed using a form of phonetic transcription.
- *Prosody*: Prosodic annotation tries to capture in a written form the suprasegmental features of spoken language – primarily stress, intonation, and pauses.

A corpus can be used for various purposes:

- A part of the corpus can be used for the *construction* of the natural language processing components, e.g. information extraction components. The lexicon and the grammars are optimized towards this part of the corpus.
- The performance of natural language processing components can be quantitatively *evaluated* relative to the parts of corpus, which were not used during construction.
- A corpus can be used for *other research objectives* (e.g., complexity of nominal phrases, word sense disambiguation, machine learning of grammatical structures, etc.).

The KFOR text corpus is a micro-text corpus consisting of HUMINT reports in English from the KFOR deployment of the Bundeswehr. In the KFOR corpus various syntactic and semantic annotation types are used. The corpus is used for constructing and evaluating the information extraction component of the ZENON system.

3. The KFOR Text Corpus

4,498 military reports (mostly in English) from the KFOR deployment of the German Federal Armed Forces were used for the realization of the ZENON prototype. From these reports 800 were manually annotated and form the *KFOR Corpus*. This corpus is a specialized micro-text corpus (cf. [McEnery, 2001]). The corpus covers 886,000 tokens and contains the annotations in different layers (cf. [Hecking, 2006c]). The following layers are available:

- *Original markups*: In this layer those parts of the message are annotated that are already formatted (e.g. addressee, topic, source).
- *Token*: This layer contains the annotations about words, numbers, etc. The part-of-speech information and the lemma are also given.
- *Gazetteer*: In this layer those expressions are annotated that were identified over lists of names (e.g., first names, city names).
- *Sentence*: These annotations refer to sentences and begin and end markers of comments.
- *Named entities*: City, Company, Coordinates, Country, CountryAdj, Currency, Date, GeneralOrg, MilitaryOrg, Number, Percent, Person, PoliticalOrg, Province, Region, River, Time and Title.
- *Verb group*: The verbal phrases are annotated.

- *Thematic roles*: The syntactic and semantic function of expressions in sentences is annotated (cf. [Kremer, 2006a]).

During the creation of the corpus a first version of the annotations was produced automatically. These annotations were then checked manually and corrected. For both working-steps GATE (cf. [Cunningham, 2002]) was used. The corpus is represented in

- the GATE-specific format,
- the GATE-specific format in XML,
- the ANC (American National Corpus) stand-off annotation format, and
- the TIGER-XML format.

The corpus contains both syntactic and semantic annotations. These are arranged into different *annotation layers*. The Figure 1 indicates, which annotation layers and annotation types are present, whether they are syntactic or semantic annotation types, and which of the annotation types were manually corrected.

Syntactical/ semantical	**Annotation layer**	**Annotation type**	**Checked manually**
syntactical	Original markup	DocID, DTGMeldung, Einsatz, Empfaenger, Hauptthema, Koordinate, Meldung, Meldungstyp, Ort, Quelle, Sachverhalt, Schlagworte, Titel, Unterthema	no
syntactical	Token	Token, SpaceToken	no
semantical	Gazetteer	Lookup	no
syntactical	Sentence	Sentence Comment Split	yes yes no
semantical	NE	City, Company, Coordinates, Colour, CountryAdj, Currency, Date, DocumentID, GeneralOrg, MilDateTime, MilitaryOrg, Number, Percent, Person, PoliticalOrg, Province, Region, River, Time, Title	yes
syntactical	VG	VG	yes
semantical	ThematicRole	ThRo	yes

Figure 1: Annotation layers and annotation types

For each annotation the type, the layer (Set), the start- and the end-position and a set of annotation-specific features are given. Each feature consists of a name and a value. A feature appears only, if a value is present.

In the example

```
City NE xxx yyy {name=BERLIN}
```

the annotation is of type `City`. It belongs to the annotation layer `NE`. The string to which the annotation refers begins in position `xxx` and ends with position `yyy`. The annotation possesses a feature with the name `name` and the value 'BERLIN'.

In contrast to the syntactic annotation delimitation problems can occur with the semantic annotation. A classification of linguistic material according to semantic criteria is not always clearly feasible. Words can to be polysem (more than one meaning) and different language users connote (positive or negative) them differently. E.g., the abbreviation "KPC" (Kosovo Protection Corps) can refer to an organization, which is a kind of THW (German Federal Agency for Technical Relief). This is official view of the political institutions (cf. [UNMIK, 2006]). Thus the abbreviation would be annotated as `PoliticalOrg` in the KFOR corpus. However, according to another opinion, the KPC (as a successor of the Kosovo Liberation Army (KLA)) is a terrorist organization. Under this reading the KPC must be annotated as a `MilitaryOrg`. If there are these cases of doubt, the official view of the political institutions forms the basis for the annotation in the KFOR corpus. If this official view is not assignable, the view that is used by most language users forms the basis for the annotation. Even if this view is not assignable, the annotator decides according to a set of defined rules (cf. [Hecking, 2006c, appendix B]).

Several annotation types and the structure of the annotations are described in the following. The complete description of all annotation types can be found in [Hecking, 2006c].

3.1 Token

The first annotation layer contains the basic building blocks, i.e. information about words, numbers, etc. Two annotation types belong to this layer: `SpaceToken` and `Token`. The structure of the annotation type `SpaceToken` is:

Annotation type	Feature name	Feature value
`SpaceToken`	kind	space control
	length	String
	string	String

and the structure of `Token` is:

Annotation type	Feature name	Feature value
`Token`	affix	String
	category	CC, CD, DT, EX, FW, IN, JJ, JJR, JJS, JJSS, -LRB-, LS, MD, NN, NNP, NNPS, NNS, NP, NPS, PDT, POS, PP, PRPR$, PRP, PRP$, RB, RBR, RBS, RP, STAART, SYM, TO, UH, VBD, VBG, VBN, VBP, VB, VBZ, WDT, WP$, WP, WRB, ’::’, ’,’, ’$’, ’-’, ’(’, ’.’, ’#’, ’)’, ””
	kind	word, number, symbol, punctuation
	length	Number
	orth	allCaps, lowercase, mixedCaps, upperInitial
	position	startpunct, endpunct
	root	String
	string	String

`SpaceToken` is used to mark spaces between words, number, etc. and control sequences, e.g. carriage return, line feed, etc.

Each `Token` annotation contains among other things information about the part-of-speech (`category`) and the lemma (`root` and `affix`). The part-of-speech values are explained in [Hecking, 2006c, p. 53ff].

The string "KFOR" is annotated as a token with `Token Token xxx yyy {category= NNP, kind=word, length=4, orth=allCaps, root=kfor, string= KFOR}`.

3.2 Gazetteer

In this annotation layer those expressions are marked, which were identified through lists of names (so-called gazetteers). In the annotation type `Lookup` the two features `majorType` and `minorType` are used. The following listed values for them are used in the KFOR corpus:

Annotation type	majorType	minorType
`Lookup`	colour	<no>
	country_adj	<no>
	date	day, month
	location	city, country, province, river, region
	number	<no>
	organization	general, military, political, company
	person_first	female, male
	time	ampm, hour, zone
	title	civilian, police, military, male, female

The `Lookup` annotations are used for the production of other annotations.

The string "BERLIN" is identified through a list of city names and is annotated as `Gazetteer Lookup xxx yyy {majorType=location, minorType=city}`.

3.3 Named Entities (NE)

The annotation layer of the **Named Entities (NE)** is the most extensive. During the creation of these annotations different rules were used (cf. [Hecking, 2006c, appendix B]).

National, supra-national and non-governmental military entities are treated as **military organizations**. Camps (e.g. "Camp BONDSTEEL") are not handled as military organizations. The structure of this annotation type is simple:

Annotation type	Feature name	Feature value
`MilitaryOrg`	name	String

For example, "NATO" is annotated with `MilitaryOrg NE xxx yyy {name=NATO}` and "116. NLA brigade" with `MilitaryOrg NE xxx yyy {name=116. NLA brigade}`.

The structure of the **military date-time-expression** annotation type is:

Annotation type	Feature name	Feature value
MilDateTime	year	String
	month	1, ..., 12
	day	1, ..., 31
	hour	1, ..., 24
	minute	1, ..., 60
	timeZone	UTC, ...

For example, "091100Bjul01" is annotated with `MilDateTime NE xxx yyy {year=01, month=7, day=9, hour=11, minute=0, timeZone=B}`.

3.4 Verbal Group

The verbal expressions occurring in English are also marked. The structure of the **verb group** annotation type is:

Annotation type	Feature name	Feature value
VG	adverb	String
	adverbPost	String
	infinitive	String
	negation	yes
	special	HadBetter, SupposedTo, BeTo, HaveTo, GotTo, GoingTo, AbleTo, UnableTo, UsedTo
	tense	BeVBG, BeVBN, FutCon, FutPer, FutPerCon, HaveVBG, HaveVBN, HaveBeenVBG, Inf, Pas, PasCon, PasPer, PasPerCon, Per, PerCon, Pre, PreCon, PrePer, PrePerCon, SimFut, SimPas, SimPre
	type	FVG, MODAL, NFVG, PART, SPECIAL
	voice	active, passive

`adverb` and `adverbPost` are used to handle adverbs in the verb complex. Special verbs (e.g., to have to, to be able to, …) are handled with the help of the `special` feature. For these verbs the `type`-feature has the value `SPECIAL`. In "... who are able to speak ALBANIAN language..." the verb complex "are able to speak" is annotated with `VG xxx yyy {infinitive=speak, special=AbleTo, tense=Pre, type=SPECIAL}`. The tense-feature describes the time dimension of the verbal group, e.g., `PasPerCon` means "Past Perfect Continuous" and `HaveBeenVBG` abbreviates "have been + Gerund or Present Participle". The `type`-feature is used for sub-classifying the verb complex. The values are: `FVG` (Finite Verb Group), `MODAL` (Modal Verb), `NFVG` (Non Finite Verb Group), `PART` (Participle) and `SPECIAL`.

The verbal complex in the sentence "CPC *can no more tolerate* this ladys behavior."[3] is annotated with `{adverb=more, infinitive=tolerate, modal=can, neg=yes, type=MODAL, voice=active}`.

The annotation of the verb complex reveals different problems, e.g.,

[3] This is an original sentence from the KFOR corpus, therefore typos were not corrected.

- Verb constituents can also be part of a nominal phrase (NP). In the NP "a hardware check-up of the planned test" the part "check-up" is annotated with `VG xxx yyy {infinitive=check-up, tense=Inf, type=PART}`.

- Non-native English speakers use words which can't be found in the dictionary but which are intelligible. E.g., the verb "to unclarify" can't be found in [LEO, 2006] or [Sinclair, 2001]. These words are annotated in the KFOR corpus. For the example: `VG xxx yyy {infinitive=unclarify, tense=Inf, type=NFVG, voice=active}`.

- Words not belonging to the verb complex can be part of the complex, e.g., the subject in questions. Through this, the verb complex is divided into parts. Should these parts be handled as separate verb complexes or not? In the KFOR corpus the parts are handled separately, e.g., in "Should they have a coalition?" the two parts "Should" and "have" are annotated with `VG xxx yyy {modal=should, type=MODAL}` and `VG xxx yyy {infinitive=have, tense=Inf, type=NFVG, voice=active}`.

4. Evaluation of the ZENON System

A first version of the ZENON system was built (cf. [Hecking, 2006a, Hecking, 2006b]). The KFOR corpus was used to evaluate the realized information extraction component. For the evaluation the GATE "Corpus Benchmark Evaluation Tool" was used (cf. [Cunningham, 2002]). For the evaluation the following three metrics are used (definitions are from [GATE, 2007, chapter 11]).

- *Precision P* measures the number of correctly identified items as a percentage of the number of all items identified. In other words, it measures how many of the items that the system identified were actually correct, regardless of whether it also failed to retrieve correct items. The higher the precision, the better the system is at ensuring that *what is identified is correct*.

- *Recall R* measures the number of correctly identified items as a percentage of the total number of correct items. In other words, it measures how many of the items that should have been identified actually were identified, regardless of how many spurious identifications were made. The higher the recall rate, the better the system is at *not missing correct items*.

- The *F-measure* is often used as a weighted average of the Precision and Recall. It is defined as: F = (2*P*R)/(P+R).

The Toolset GATE offers the possibility to compare two different sets of annotations on the same documents with the help of the "Corpus Benchmark Evaluation Tool". This is used to compare the annotations produced by the ZENON system with the respective annotations of the KFOR corpus. The corpus then serves as a "measuring stick" to quantitatively judge the performance of the information extraction component of the ZENON system. Figure 2 shows an example. All NE annotations, which were produced by the information extraction of the ZENON system were compared with the NE annotations of the KFOR corpus. This comparison was made with all 800 documents of the corpus. The result represents a snapshot (December 2006) and is used to improve the information extraction of the ZENON system continuously.

Statistics

Annotation Type	Correct	Partially Correct	Missing	Spurious	Precision	Recall	F-Measure
City	3094	147	144	649	0.8142673521850899	0.9357459379615952	0.8707903780068728
Company	68	15	185	11	0.8031914893617021	0.28171641791044777	0.4171270718232044
Coordinates	2220	61	23	205	0.9052695092518102	0.9767795138888888	0.9396659707724425
Colour	21	2	1	0	0.9565217391304348	0.9166666666666666	0.9361702127659574
CountryAdj	602	77	2095	39	0.8920612813370473	0.23089401586157174	0.3668384879725086
Currency	40	9	101	0	0.9081632653061225	0.2966666666666667	0.44723618090452266
Date	310	56	531	1	0.9209809264305178	0.37681159420289856	0.5348101265822786
DocumentID	903	21	7	0	0.9886363636363636	0.981203007518797	0.9849056603773584
GeneralOrg	2	1	381	0	0.8333333333333334	0.006510416666666667	0.012919896640826874
MilDateTime	0	0	40	0	0.0	0.0	0.0
MilitaryOrg	988	300	159	1345	0.43220660843144704	0.7864547339322737	0.557843137254902
Number	4648	120	208	1416	0.7613195342820182	0.9461414790996785	0.8437275985663083
Percent	36	7	17	0	0.9186046511627907	0.6583333333333333	0.7669902912621358
Person	358	120	1289	23	0.8343313373253493	0.23655913978494625	0.36860670194003525
PoliticalOrg	1921	406	712	280	0.8147295742232451	0.6989141164856861	0.7523910733262488
Province	1	0	0	0	1.0	1.0	1.0
Region	11	2	232	0	0.9230769230769231	0.04897959183673469	0.09302325581395349
River	3	1	3	0	0.875	0.5	0.6363636363636364
Time	14	13	106	0	0.7592592592592593	0.15413533834586465	0.25625
Title	504	286	147	161	0.6803364879074658	0.6905016008537886	0.6853813559322034

Overall average precision: 0.8669906178545418
Overall average recall: 0.671384245583424
Overall average fMeasure : 0.6360350718350036

Figure 2: Evaluation results for information extraction of the ZENON system

5. Conclusion

The *KFOR text corpus* is a micro-text corpus consisting of HUMINT reports in English from the KFOR deployment of the Bundeswehr. It was used for constructing and evaluating the information extraction component of the ZENON system.

In this paper, the KFOR corpus was presented. After introducing the concept of corpora for computational linguistics research and development, the KFOR text corpus was described in detail. Various annotation types were presented. In the last section, the use of the corpus for the evaluation of the information extraction component of the ZENON system was shown.

References

[ANC, 2007] *American National Corpus*. http://www.americannationalcorpus.org/about.html (15.1.2007).

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

[BNC, 2007] *British National Corpus*. http://www.natcorp.ox.ac.uk/ (15.1.2007).

[Cunningham, 2002] H. Cunningham, D. Maynard, K. Bontcheva, V. Tablan. *GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications*. Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02). Philadelphia, July 2002.

[GATE, 2007] *Developing Language Processing Components with GATE Version 4 (a User Guide)*. http://www.gate.ac.uk/sale/tao/#x1-29300011 (15.1.2007).

[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.

[Hecking, 2004a] Hecking, M. *Informationsextraktion aus militärischen Freitextmeldungen*. Forschungsgesellschaft für Angewandte Naturwissenschaft e.V. (FGAN), Wachtberg, Germany, FKIE-Bericht Nr. 74, 2004.

[Hecking, 2004b] Hecking, M. *How to Represent the Content of Free-form Battlefield Reports*. In: Proc. of the 2004 Command and Control Research and Technology Symposium (CCRTS) "The Power of Information Age Concepts and Technologies", June 15-17, 2004, San Diego, California.

[Hecking, 2005a] Hecking, M. *Domänenspezifische Informationsextraktion am Beispiel militärischer Meldungen*. In: A.B. Cremers, R. Manthey, P. Martini, V. Steinhage (Hrsg.) "INFORMATIK 2005”, Band 2, Lecture Notes in Informatics, Volume P-68, Bonn, 2005.

[Hecking, 2006a] Hecking, M. *Content Analysis of HUMINT Reports*. In: Proc. of the 2006 Command and Control Research and Technology Symposium (CCRTS) "THE STATE OF THE ART AND THE STATE OF THE PRACTICE", June 20-22, 2006, San Diego, California.

[Hecking, 2006b] Hecking, M. *Navigation through the Meaning Space of HUMINT Reports*. In: "Proceedings of the 11th International Command and Control Research and Technolgy Symposium”, September 26-28, 2006, Cambridge, UK.

[Hecking, 2006c] Hecking, M. *Das KFOR-Korpus als Ergebnis semantisch annotierter militärischer Meldungen*. Forschungsgesellschaft für Angewandte Naturwissenschaft e.V. (FGAN), Wachtberg, Germany, FKIE-Bericht Nr. 124, 2006.

[Kremer, 2006a] Kremer, C. *Eine Untersuchung von Bewegungsverben im KFOR-Korpus im Vergleich zu FrameNet*. Forschungsgesellschaft für Angewandte Naturwissenschaft e.V. (FGAN), Wachtberg, Germany, FKIE-Bericht Nr. 117, 2006.

[LEO, 2006] LEO (2006). *Deutsch-Englisches-Wörterbuch*, http://dict.leo.org (5.10.2006).

[McEnery, 2001] McEnery, T., Wilson, A.. *Corpus Linguistics*. Edinburgh University Press, Edinburgh, 2nd edition, 2001.

[NCW, 2001] Department of Defense. *Network Centric Warfare – Report to Congress*. 27 July 2001.

[Sinclair, 2001] Sinclair, J., Ed. (2001). *Collins COBUILD English Dictionary for Advanced Learners*. HarperCollins Publishers, 3rd edition edition.

[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.

[UNMIK, 2006] UNMIK (2006). *Kosovo Protection Corps*, http://www.unmikonline. org/1styear/ kpcorps.htm (17.10.2006).