Semantic Analysis of Military Relevant Texts for Intelligence Purposes

Sandra Noubours
Dr. Matthias Hecking
Fraunhofer Institute for Communication,
Information Processing and Ergonomics FKIE
Neuenahrer Straße 20
53343 Wachtberg-Werthhoven

sandra.noubours@fkie.fraunhofer.de
matthias.hecking@fkie.fraunhofer.de

© Fraunhofer FKIE

Fraunhofer FKIE

Outline

- Introduction
- The ZENON System
  and its information extraction functionalities
- A Semantic Role Labeling application
  for ZENONs Semantic Analysis
- Conclusion

© Fraunhofer FKIE

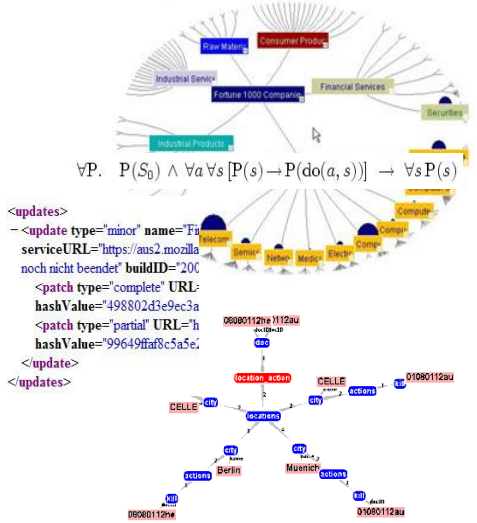
1. Introduction

Military intelligence

For military intelligence, large quantities of intelligence reports and other documents need to be analyzed.

content analysis

What?
Who? Where?
When? …?



Text from different sources can contain military relevant information (HUMINT, OSINT, …).

Information need to be extracted.

© Fraunhofer FKIE

1. Introduction

NLP for military intelligence

Natural language processing (NLP) can be applied to efficiently handle analysis of textual data.

➡ We set up the research project **ZENON**.

- ZENON realizes a (prototypical) **information extraction (IE) system** for the (partial) content analysis of English HUMINT reports.
- The system has further been extended for **multilingual information extraction**, i.e., processing Dari and Tajik texts.
- Here, we present the **improvement** of ZENON's English **semantic analysis** by **semantic role labeling (SRL)**.

© Fraunhofer FKIE

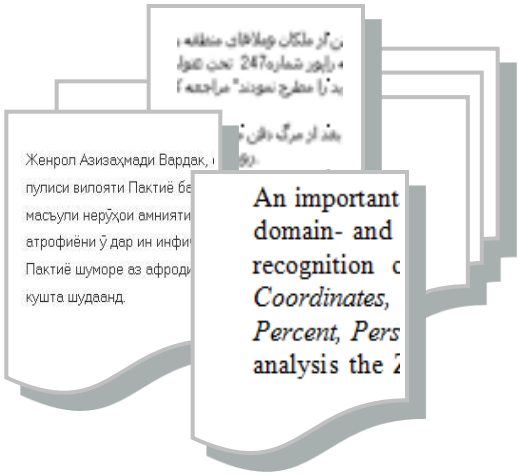
THE ZENON SYSTEM

© Fraunhofer FKIE

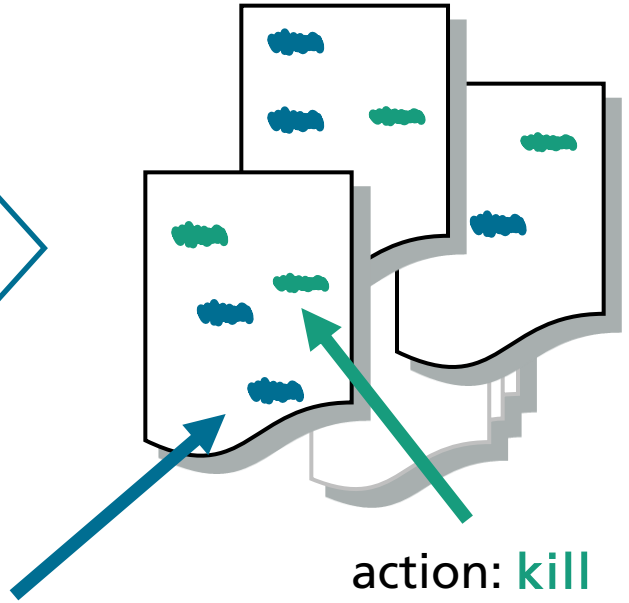
2. The ZENON System

Information extraction

Information extraction (IE) means the (partial) **content analysis** of free-form text. **Relevant information** about a specific entity and/or action in natural language texts is **identified, extracted and represented ...**



identification of relevant information



action: **kill**

entity: **MUELLER**

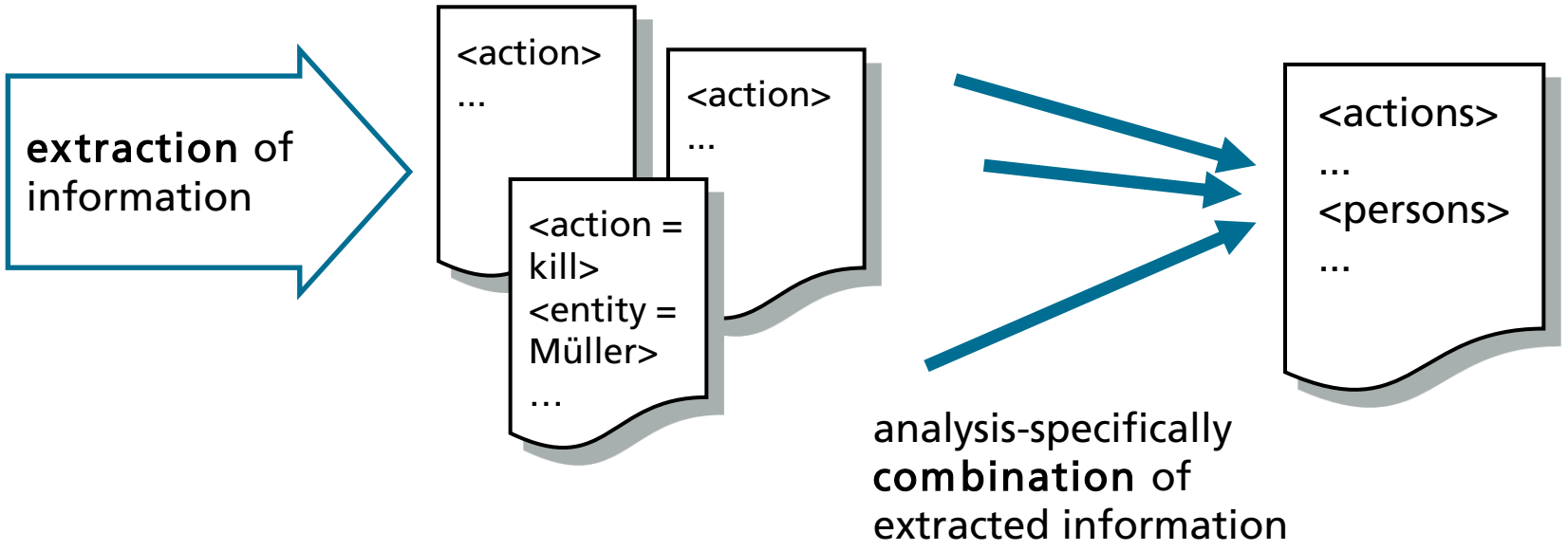
Knowledge about **actions** and **entities** is distributed over many reports.

© Fraunhofer FKIE

2. The ZENON System

Information extraction

Information extraction (IE) means the (partial) **content analysis** of free-form text. **Relevant information** about a specific entity and/or action in natural language texts is **identified, extracted and represented ...**



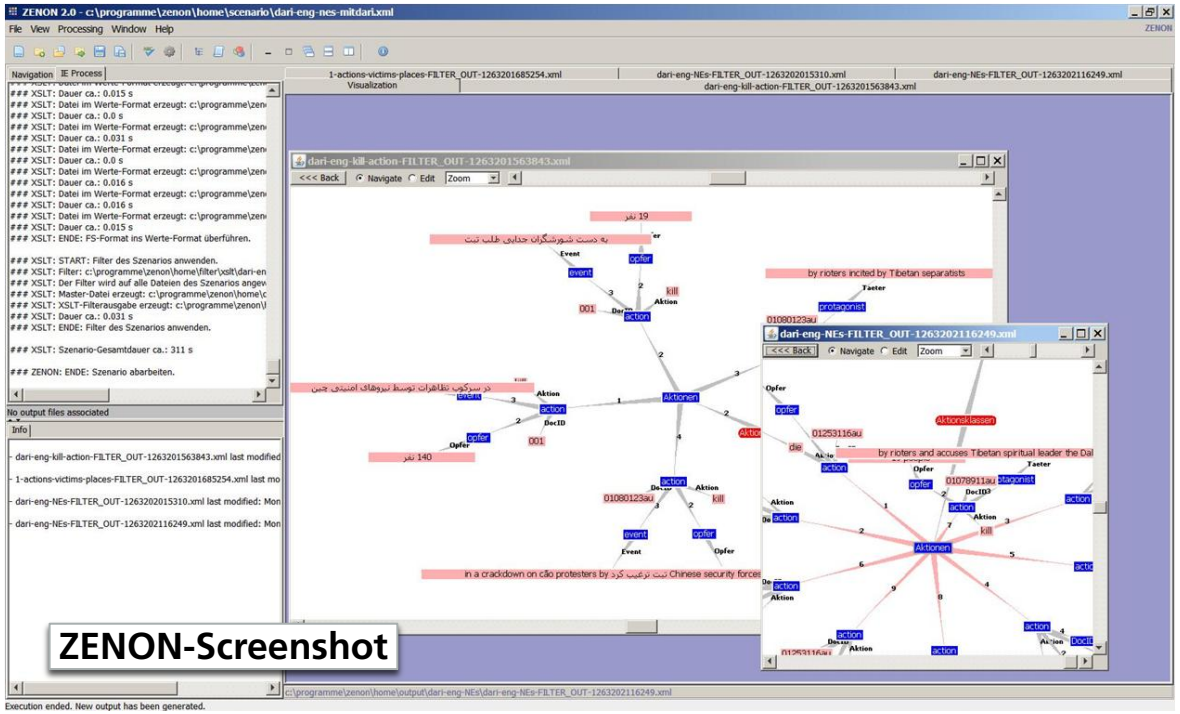
© Fraunhofer FKIE

2. The ZENON System

Information extraction

Information extraction (IE) means the (partial) **content analysis** of free-form text. **Relevant information** about a specific entity and/or action in natural language texts is **identified, extracted and represented ...**

graphical **representation**



© Fraunhofer FKIE

2. The ZENON System

Architecture

implemented using GATE software

Input: intelligence reports

Structural analysis

- Token Annotation
- POS Tags
- Sentence Annotation
- Named Entities
- Morphological Information & Lemma

Detection of verb phrases

- Verbal Chunks

Semantic Analysis

Recognize action types
→ **Infer semantic frames**

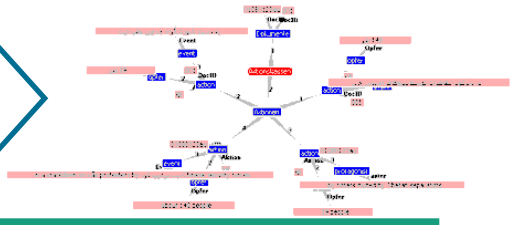
Extract relevant information
→ **Fill semantic frames**

- Semantic Frames (feature structures)

Select extracted information

Combine extracted information

Visualization



© Fraunhofer FKIE

2. The ZENON System

Original semantic processing

ZENON's **original semantic analysis** is based on **semantic frames** that are inferred from **action types**.

Example: *John Mueller was killed in an explosion incident.*

a) semantic analysis follows structural analysis and identification of verb phrases
[*John Mueller*] [*was killed*] [*in*] [*an explosion incident*].

b) the system is able to deduce an **action type** from certain text passages
… *was killed* … ➡ action type *KILL*: **A** *was killed in* **B**

c) the recognized action type determines the **semantic frame**
A *was killed in* **B** ➡ semantic frame: **VICTIM** *was killed in* **CAUSE**

d) the semantic frame is **filled with the identified entities**
[*John Mueller*]VICTIM was killed *in* [*an explosion incident*]CAUSE.

© Fraunhofer FKIE

2. The ZENON System

Original semantic processing: Problems

Action types & **semantic frames** have to be **explicitly defined**.

- manually written grammar rules specify in what textual context an action type can occur
- a semantic frame (that is inferred from a recognized action type) has to be manually encoded

Such a **manual approach is time-consuming and inefficient**.

➡ has been realized only for a small selection of English verbs and semantic frames

➡ **low coverage**

To **improve ZENON's semantic analysis**, we realized a **semantic role labeling** (**SRL**) application.

© Fraunhofer FKIE

Improving ZENONs semantic analysis by an

SRL APPLICATION

© Fraunhofer FKIE

3. SRL for ZENON

Semantic roles

Semantic roles are a form to represent the meaning of a text.
In the **context of a specific action**, **each entity** involved has a **certain semantic role**.

[*The policemen*]AGENT arrested [*the suspect*]PATIENT.

[*John Mueller*]VICTIM was killed *in* [*an explosion incident*]CAUSE.

© Fraunhofer FKIE

3. SRL for ZENON

Semantic Role Labling

Semantic role labeling (**SRL**) is the process of automatically indentifying semantic roles in a text. For each action encoded in a text the participating entities are identified and labeled with semantic roles.

In the course of **Information Extraction**, SRL can be useful as semantic roles constitute further knowledge about actions and entities.

© Fraunhofer FKIE

3. SRL for ZENON

Approach

Different SRL approaches exist, many of them are applying **machine learning** (statistical approach) where the system is trained on an annotated corpus.

- the training corpus should be domain specific
- there is no such corpus annotated with semantic roles existent for the military domain

➡ we implemented a **non-statistical approach** that makes use of a lexical resource

© Fraunhofer FKIE

3. SRL for ZENON

Approach

Our SRL approach is based on a **syntax-semantic-mapping**

- we derive **semantic roles from structural knowledge** about the clause
- for this purpose we apply information from the **lexical resource VerbNet** (VN)
  - VerbNet is an online lexicon that provides syntactic and semantic information for more than 3700 English verbs
  - semantic roles are defined in the context of syntactic structures
  - in this way **VerbNet describes mappings from syntax to semantic**, i.e., from syntactic frame to semantic roles, for each verb

© Fraunhofer FKIE

3. SRL for ZENON

Processing

For each action encoded in the text the verb and the participating entities are extracted (*syntactic analysis: a&b*) and their semantic roles are identified (*semantic annotation: c&d*). [➡ *syntax-semantic-mapping*]

a) **structural analysis** of the text is performed:
applying a statistical syntactic parser to generate a formalized syntactic representation of each sentence

b) results are used to identify the **main verb and its argument structure** for each clause

c) for each recognized main verb and its verb argument structure, the system **extracts matching VerbNet information** which also include semantic roles

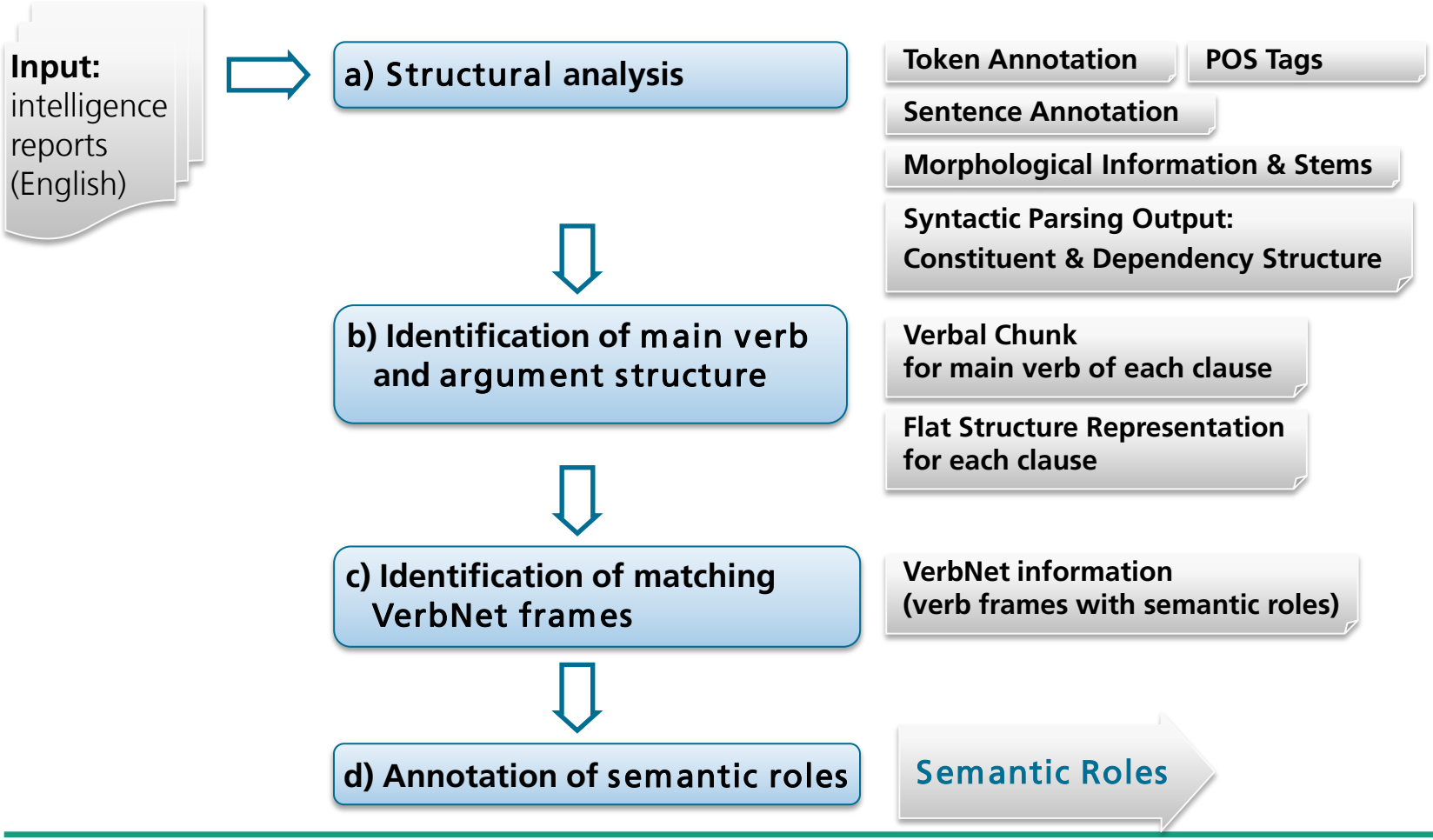
d) finally, the clause (i.e., the identified verb argument structure) is annotated with the **semantic roles**

© Fraunhofer FKIE

3. SRL for ZENON

Architecture

implemented using GATE software



© Fraunhofer FKIE

3. SRL for ZENON

Processing step a) structural analysis

Tokenization, **sentence splitting** and **morphological analysis**…

Example: *The suspect is following the politician into a public building .*

Token – Annotation:

The	suspect	is	following	the	politician	into	a	public	building	.

Sentence – Annotation:

The suspect is following the politician into a public building .

Lemma – Annotation:

The	suspect	be	follow	the	politician	into	a	public	building	.

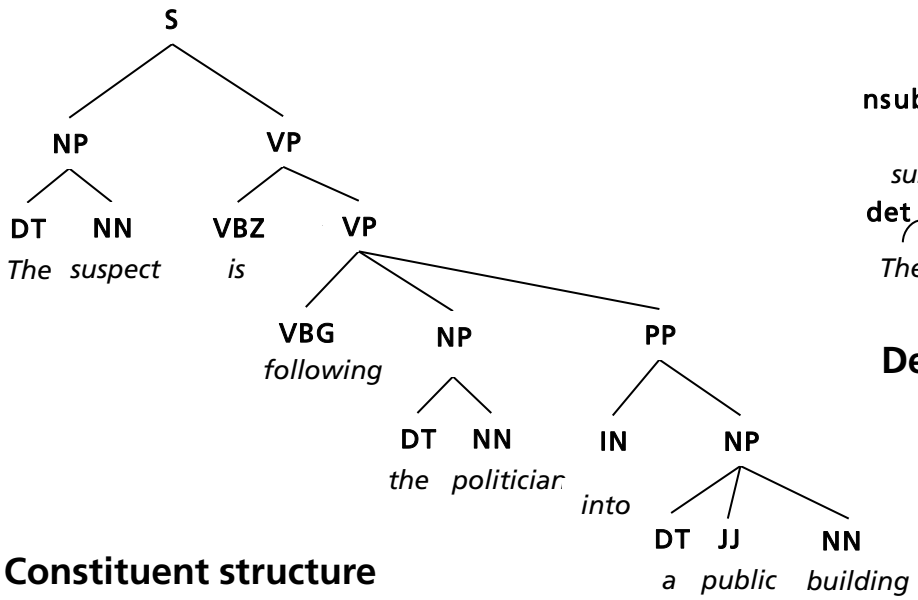
© Fraunhofer FKIE

3. SRL for ZENON

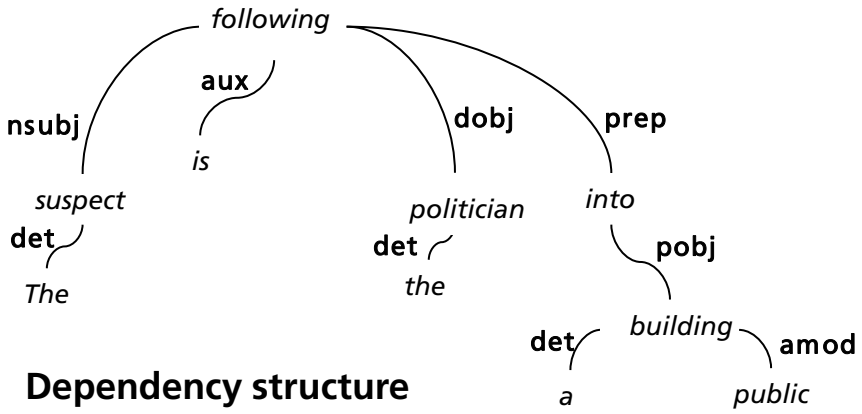
Processing step a) structural analysis

… **syntactic parsing** (Stanford Parser) of
constituent and dependency structure for each sentence.

Example: *The suspect is following the politician into a public building .*



Constituent structure



Dependency structure

© Fraunhofer FKIE

3. SRL for ZENON

Processing step b) identification of argument structure

Based on information from structural analysis:
Identification of **main verb** and **argument structure** for each clause.

Example: *The suspect is following the politician into a public building .*

Verbal Chunk – Annotation:

is following

chunk of main verbal expression
(main verb + auxiliaries + modal verbs + …)

Phrase Structure – Annotation:

NP	Verbal Chunk	NP	IN	NP

flat syntactic representation of argument structure
(Tags: NP = nominal phrase, Verbal Chunk = chunk of main verb, IN = preposition)

© Fraunhofer FKIE

3. SRL for ZENON

Processing step c) extraction of VerbNet information

Extraction of matching **VerbNet information**.

Example: *The suspect is following the politician into a public building .*

- get **lemma** for main verb

Verbal Chunk – Annotation: is following

Lemma – Annotation: follow

- convert representation of recognized argument structure of each clause into format of **VerbNet syntactic frames**

Phrase Structure – Annotation:

NP	Verbal Chunk	NP	IN	NP

Phrase Structure (in VerbNet format):

NP	Verb	NP	Preposition	NP

© Fraunhofer FKIE

3. SRL for ZENON

Processing step c) extraction of VerbNet information

Extraction of matching **VerbNet information**.

Example: *The suspect is following the politician into a public building .*

Lemma of main verb: **follow**

+

Phrase Structure (VerbNet format):

NP **Verb** **NP** **Preposition** **NP**

VerbNet

matching VerbNet verb frame (VerbNet class chase-51.6)**:**

[NP]AGENT **Verb**: **follow** **[NP]THEME** **Preposition** **[NP]LOCATION**

VerbNet – Semantic Roles:

AGENT **THEME** **LOCATION**

© Fraunhofer FKIE

3. SRL for ZENON

Processing step d) annotation with semantic roles

Annotation of verb-argument-structure with **VerbNet semantic roles**.

Example: *The suspect is following the politician into a public building .*

Verb-Argument-Structure (Phrase Structure – Annotation):

NP	Verbal Chunk	NP	IN	NP

\+

VerbNet – Information:

[NP]AGENT Verb: follow [NP]THEME Preposition [NP]LOCATION

AGENT	THEME	LOCATION

Semantic Role – Annotation:

AGENT	THEME	LOCATION

[*The suspect*]AGENT *is following* [*the politician*]THEME *into* [*a public building*]LOCATION .

ZENON

© Fraunhofer FKIE

CONCLUSION

© Fraunhofer FKIE

4. Conclusion

- **Processing of human language** is a critical capability in many future **military applications**.
- **ZENON** is a prototypical **information extraction system** for the partial content analysis of free-form texts. We expect that systems like ZENON will **increase productivity of the intelligence analyst**.
- We implemented a SRL application to improve ZENON's semantic analysis. This is expected to **improve the all-over performance of the ZENON system**.

© Fraunhofer FKIE

Reference

- S. Noubours. *Annotation semantischer Rollen in HUMINT-Meldungen basierend auf dem statistischen Stanford Parser und der lexikalischen Ressource VerbNet*. FKIE-Bericht Nr. 195. Wachtberg, Germany: Fraunhofer-FKIE, 2010.
- M. Hecking and T. Sarmina-Baneviciene. *A Tajik Extension of the Multilingual Information Extraction System ZENON*. In: Proceedings of the 15th International Command and Control Research and Technolgy Symposium (ICCRTS), Santa Monica, CA, USA, June 2010.
- M. Hecking. *System ZENON – Semantic Analysis of Intelligence Reports*. In: Proceedings of the LangTech 2008, February 28-29, 2008, Rome, Italy
- M. Hecking. *Multilinguale Textinhaltserschließung auf militärischen Texten*. In: M. Wunder and J. Grosche (Ed.). Verteilte Führungsinformationssysteme. Heidelberg, Germany: Springer, 2009.

© Fraunhofer FKIE

Thank you for your attention!

Questions?

© Fraunhofer FKIE