An Extension of the TIGER Query Language for Treebanks with Frame Semantics Annotation

Thesis Presentation for M.Sc. LST

Torsten Marek <tmarek@coli.uni-saarland.de>

Saarland University

February 9th, 2009

Outline

Motivation

Linguistic Queries

Find all sentences where the role TOPIC *in the frame* STATEMENT *is realized by a PP with the preposition "über".*

Motivation

Linguistic Queries

Find all sentences where the role TOPIC *in the frame* STATEMENT *is realized by a PP with the preposition "über".*

SALSA Corpus (Burchardt et al., 2006)

- Annotations:
  - Syntax/Phrase Structure (from the TIGER corpus)
  - Frame Semantics

Motivation

Linguistic Queries

Find all sentences where the role TOPIC *in the frame* STATEMENT *is realized by a PP with the preposition "über".*

SALSA Corpus (Burchardt et al., 2006)

- Annotations:
  - Syntax/Phrase Structure (from the TIGER corpus)
  - Frame Semantics

[Hotels und Gaststätten]$_{\text{SPEAKER}}$ klagen$_{\text{STATEMENT}}$ [über knauserige Gäste]$_{\text{TOPIC}}$.

[Hotels and restaurants]$_{\text{SPEAKER}}$ complain$_{\text{STATEMENT}}$ [about stingy guests]$_{\text{TOPIC}}$.

Motivation II

Searching with TIGER

- *. . . PP with the preposition "über"*

```
[cat="PP"] >AC [word="über"]
```

- *. . . role* TOPIC *in the frame* STATEMENT *. . .*

  Not expressible with TIGER

Motivation II

Searching with TIGER

- *. . . PP with the preposition "über"*

```
[cat="PP"] >AC [word="über"]
```

- *. . . role* TOPIC *in the frame* STATEMENT *. . .*

Not expressible with TIGER

Thesis Goal

Develop and implement an extension of TIGER that allows searching for:

- frame instances, roles and targets
- the combination of semantic annotation with syntactic structure

TIGER Primer (Lezius, 2002)

Syntactic Structures

Language for description of partial syntactic structures with:

- node descriptions: local features

  surface string, POS tag, lemma, phrasal category

- relations: phrase structure

  dominance, siblings, precedence

TIGER Primer (Lezius, 2002)

Syntactic Structures

Language for description of partial syntactic structures with:

- node descriptions: local features

  surface string, POS tag, lemma, phrasal category

- relations: phrase structure

  dominance, siblings, precedence

Query

```
#h:[word="Haus" & pos="NN"]
    & [cat="NP"] > #h
```

Find all graphs that contain an NP with the noun "Haus".

TIGER Primer (Lezius, 2002)

Syntactic Structures

Language for description of partial syntactic structures with:

- node descriptions: local features

  surface string, POS tag, lemma, phrasal category

- relations: phrase structure

  dominance, siblings, precedence

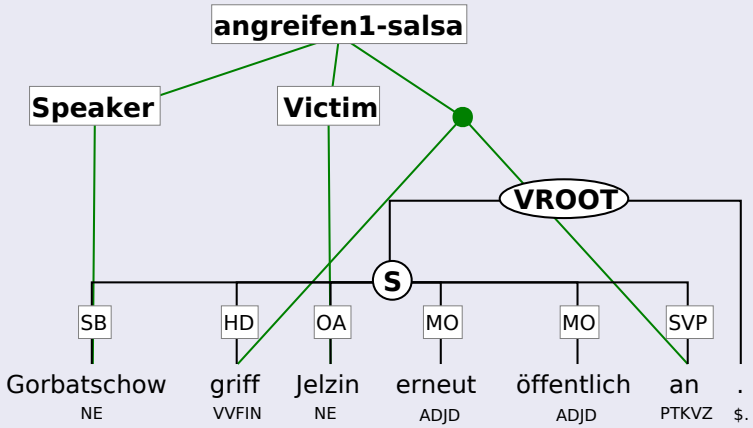
Query

```
#h:[word="Haus" & pos="NN"]
      & [cat="NP"] > #h
```

Find all graphs that contain an NP with the noun "Haus".

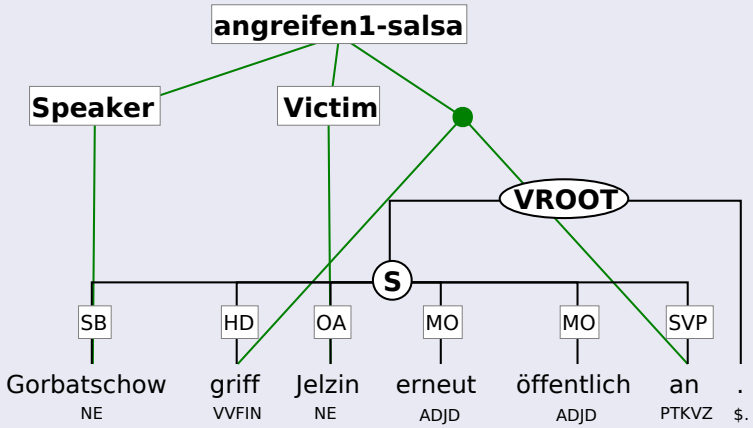
Structure of Frame Semantic Annotation

Sample Graph from SALSA

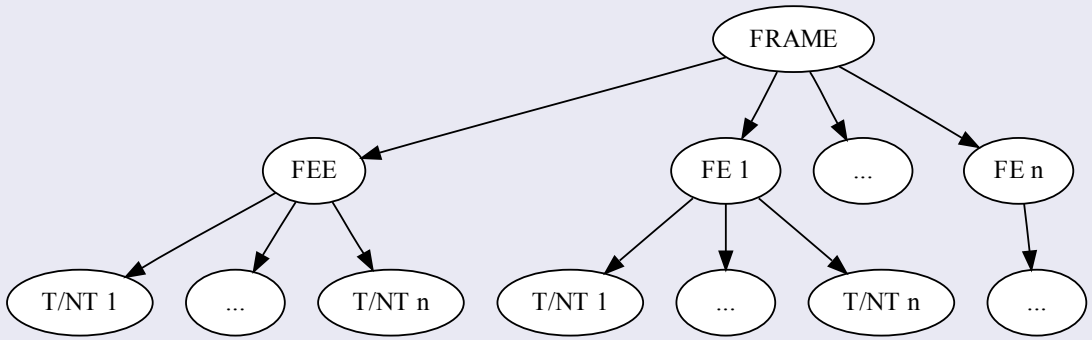


Structure of Frame Semantic Annotation

Sample Graph from SALSA



Structure of Frame Annotation



Design Guidelines

1. No interference with syntax-only queries
   - Queries for syntactic elements only may not be influenced by frame semantic annotation
2. Seamless integration with the existing query language
   - Easy to learn and understand
   - Minimal additions to query language constructs
   - As few changes to existing elements as possible
3. Stateless implementation
   - No change in processing behavior given the presence of frame semantics annotation should be needed

New Node Types

Original TIGER Node Types

- each node is a feature record (FREC)
- terminals (T) and nonterminals (NT)

New Node Types

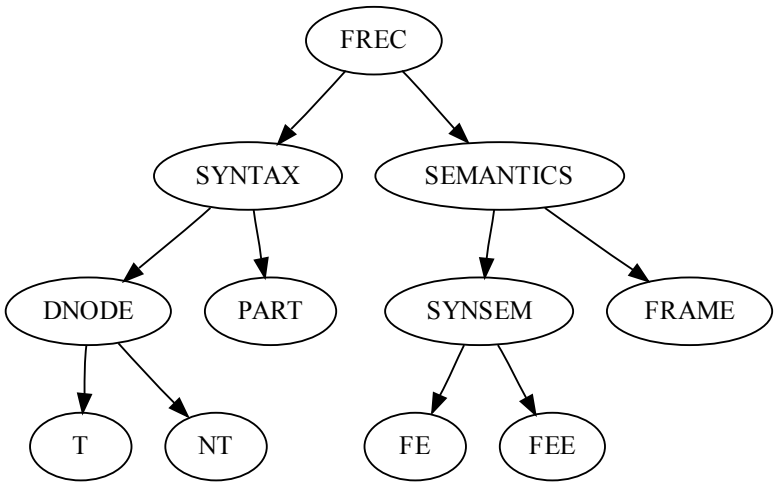
Original TIGER Node Types

- each node is a feature record (FREC)
- terminals (T) and nonterminals (NT)

New Elements

- frame instances
- targets/frame-evoking elements
- roles/frame elements

Revised Type Hierarchy



Remarks

- All node types are feature records
- DNODE: nodes in the syntactic dominance hierarchy
- SYNSEM: frame members, connect frame instances to syntactic material
- SYNTAX: nodes that can be referenced from SYNSEMs
- SEMANTICS: all elements of frame semantic annotation

Query Syntax

Node Descriptions

- Features for frames, roles and targets are fixed by the annotation format
  - FEE: *lemma*
  - FE: *role*, *semtype*
  - FRAME: *frame*
- Conflicts between syntax and semantics nodes are possible

Query Syntax

Node Descriptions

- Features for frames, roles and targets are fixed by the annotation format
  - FEE: *lemma*
  - FE: *role*, *semtype*
  - FRAME: *frame*
- Conflicts between syntax and semantics nodes are possible

T or FEE?

```
[lemma="Haus"]
```

Violation of the guidelines!

Query Syntax

Node Descriptions

- Features for frames, roles and targets are fixed by the annotation format
  - FEE: *lemma*
  - FE: *role*, *semtype*
  - FRAME: *frame*
- Conflicts between syntax and semantics nodes are possible

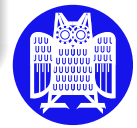
T or FEE?

```
[lemma="Haus"]
```

Violation of the guidelines!

Solution

Node descriptions for SEMANTICS nodes are enclosed between { and }



Frame-to-Frame Relations

Relations in FrameNet

- Inheritance
- Subframe
- Using
- CoreSet
- ...

Important:
Defined between abstract frames, not concrete frame instances.

(Ruppenhofer et al., 2006)

Frame-to-Frame Relations

Relations in FrameNet

- Inheritance
- Subframe
- Using
- CoreSet
- ...

Important:
Defined between abstract frames, not concrete frame instances.

(Ruppenhofer et al., 2006)

New Syntax Elements

`{frame=[Event]}`

Find all instances of frames that inherit from EVENT.

Frame-to-Frame Relations

Relations in FrameNet

- Inheritance
- Subframe
- Using
- CoreSet
- ...

Important:
Defined between abstract frames, not concrete frame instances.

(Ruppenhofer et al., 2006)

New Syntax Elements

`{frame=[Event]}`

Find all instances of frames that inherit from EVENT.

Type Literal

- similar to `/.../`
- modifies matching feature values against literals

1. Introduction
   - Motivation
   - TIGER
2. Design
3. Extensions
   - Nodes & Features
   - Relations
   - Predicates
4. Results & Future Work

Features

Remarks

- All features are defined for a type (the *domain*)
- Fixed by TIGER/SALSA XML (Erk and Padó, 2004)
- Some features are not part of the annotation, but added from the frame descriptions during corpus preparation
- Allowable feature values are taken from frame database

Frames: FRAME

- *frame*
  Name of the frame

Features II

Targets: FEE

- *lemma*
  The lemma of the word or phrase evoking the frame
- *head*
  The head of the lemma

Roles: FE

- *role*
  The name of the role
- *semtype*
  The ontological filler type.

  Present on all nodes, but in SALSA only used for FE
- *coretype*
  The role coreness: Core, Peripheral, Extra-Thematic

Relations: Frame Members

Frame Members

FRAME > SYNSEM

Check if a SYNSEM node is part of a frame instance

Example

```
{frame="Commerce_sell"} > {role="Buyer"}
```

In den USA ist es gesetzlich verboten, [Tabakwaren]$_{\text{GOODS}}$ [an Personen unter 18 Jahren]$_{\text{BUYER}}$ zu verkaufen$_{\text{COMMERCE_SELL}}$.

In the US, selling$_{\text{COMMERCE_SELL}}$ [tobacco products]$_{\text{GOODS}}$ [to people under 18 years]$_{\text{BUYER}}$ is outlawed.

Relations: Frame Siblings

Frame Siblings

SYNSEM $ SYNSEM

Check if two SYNSEM nodes are part of the same frame instance.

Example

```
{role="Factory"} $ {role="Product"}
```

Zu DDR-Zeiten wurden [dort]$_{\text{FACTORY}}$ hauptsächlich [Röntgenfilme]$_{\text{PRODUCT}}$ produziert$_{\text{MANUFACTURING}}$.

At the time that the GDR existed, mainly [X-ray films]$_{\text{PRODUCT}}$ were produced$_{\text{MANUFACTURING}}$ [there]$_{\text{FACTORY}}$.

Relations: Syntactic Material

Frame Siblings

SYNSEM > SYNTAX

Check if a SYNTAX node is referenced by a SYNSEM.

Example

```
{role="Topic"} > #pp:[cat="PP"] &
#pp >AC [word="über"]
```

[Hotels und Gaststätten]$_{\text{SPEAKER}}$ klagen$_{\text{STATEMENT}}$ [über knauserige Gäste]$_{\text{TOPIC}}$.

[Hotels and restaurants]$_{\text{SPEAKER}}$ complain$_{\text{STATEMENT}}$ [about stingy guests]$_{\text{TOPIC}}$.

Relations: Underspecification

Underspecification

$$\text{FE} \sim \text{FE} | \text{FRAME} \sim \text{FRAME}$$

Two nodes are part of the same underspecification block.

Example

$$\{\texttt{frame="Removing"}\} \sim \{\texttt{frame="Taking"}\}$$

Gott hat gegeben und [Gott]$_{\text{AGENT}}$ hat genommen$_{\text{REMOVING,TAKING}}$.

God gives and [God]$_{\text{AGENT}}$ takes away$_{\text{REMOVING,TAKING}}$.

Predicates

Core Sets

`has_coreset / no_coreset`

Can be used to test if at least one core set of roles is instantiated completely on a FRAME node.

Predicates

Core Sets

`has_coreset / no_coreset`

Can be used to test if at least one core set of roles is instantiated completely on a FRAME node.

Non-local References

`has_external / no_external`

Checks for the presence of non-local references on roles in the original annotation.

Result

Solution for Initial Question

```
{frame="Statement"} > #r:{role="Topic"}
        & #r > #pp:[cat="PP"]
        & #pp >AC [word="über"]
```

Find all sentences where the role TOPIC *in the frame* STATEMENT *is realized by a PP with the preposition "über".*

Result

Solution for Initial Question

```
{frame="Statement"} > #r:{role="Topic"}
        & #r > #pp:[cat="PP"]
        & #pp >AC [word="über"]
```

Find all sentences where the role TOPIC *in the frame* STATEMENT *is realized by a PP with the preposition "über".*

Implementation

- Extensions are fully implemented
- Query browser front-end exists, to be set up for CoLi

Non-local References

Non-local References

Roles may contain syntactic material from other graphs.

Non-local References

Non-local References

Roles may contain syntactic material from other graphs.

[Das Unternehmen]$_{\text{MANUFACTURER}}$ produzierte einschließlich Fremdfertigung mehr als 20 000 Fahrzeuge. Im Jahr zuvor waren [19 348 Autos]$_{\text{PRODUCT}}$ vom Band gerollt$_{\text{MANUFACTURING}}$.

[The company]$_{\text{MANUFACTURER}}$ produced 20 000 cars, including external production. In the previous year, [19 348 cars]$_{\text{PRODUCT}}$ left the assembly line$_{\text{MANUFACTURING}}$.

s13732–s13733 of the SALSA corpus

Non-local References

Non-local References

Roles may contain syntactic material from other graphs.

[Das Unternehmen]$_{\text{MANUFACTURER}}$ produzierte einschließlich Fremdfertigung mehr als 20 000 Fahrzeuge. Im Jahr zuvor waren [19 348 Autos]$_{\text{PRODUCT}}$ vom Band gerollt$_{\text{MANUFACTURING}}$.

[The company]$_{\text{MANUFACTURER}}$ produced 20 000 cars, including external production. In the previous year, [19 348 cars]$_{\text{PRODUCT}}$ left the assembly line$_{\text{MANUFACTURING}}$.

s13732–s13733 of the SALSA corpus

Problems

- SYNSEM > SYNTAX can introduce non-local references
- adjacent graphs have to be considered
- unsupported for now, outside the scope of this work

Future Work

Short Term

- Better web front-end
- Speed improvements
- Full coverage test suite

Future Work

Short Term

- Better web front-end
- Speed improvements
- Full coverage test suite

Long Term

- Queries on parallel corpora with aligned frame annotations
- Quantitative statistics on result sets (like TIGERSearch)
- Meta-queries for result set manipulation
- Generic mechanism for non-local references

The End

Thank you for your attention.

Questions?

References


Burchardt, A., Erk, K., Frank, A., Kowalski, A., Padó, S., and Pinkal, M. (2006).
The SALSA corpus: A german corpus resource for lexical semantics.
In *Proceedings of LREC 2006*, Genoa, Italy.


Erk, K. and Padó, S. (2004).
A powerful and versatile XML format for representing role-semantic annotation.
In *Proceedings of LREC 2004*, Lisbon, Portugal.


Lezius, W. (2002).
Ein Suchwerkzeug für syntaktisch annotierte Korpora.
PhD thesis, IMS, University of Stuttgart, Stuttgart, Germany.


Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., and Scheffczyk, J. (2006).
FrameNet II: Extended Theory and Practice.

Alternative Example for Non-local References

[Der US-Delegationsleiter John Kornblum]$_{\text{SPEAKER}}$ reagierte$_{\text{STATEMENT}}$ mit den Worten: "[Wir sind nicht bereit, 100 Tage zu warten, absolut nicht]$_{\text{MESSAGE}}$."

[US delegation chief John Kornblum]$_{\text{SPEAKER}}$ reacted$_{\text{STATEMENT}}$, saying: "[We are not willing to wait 100 days, absolutely not.]$_{\text{MESSAGE}}$."