

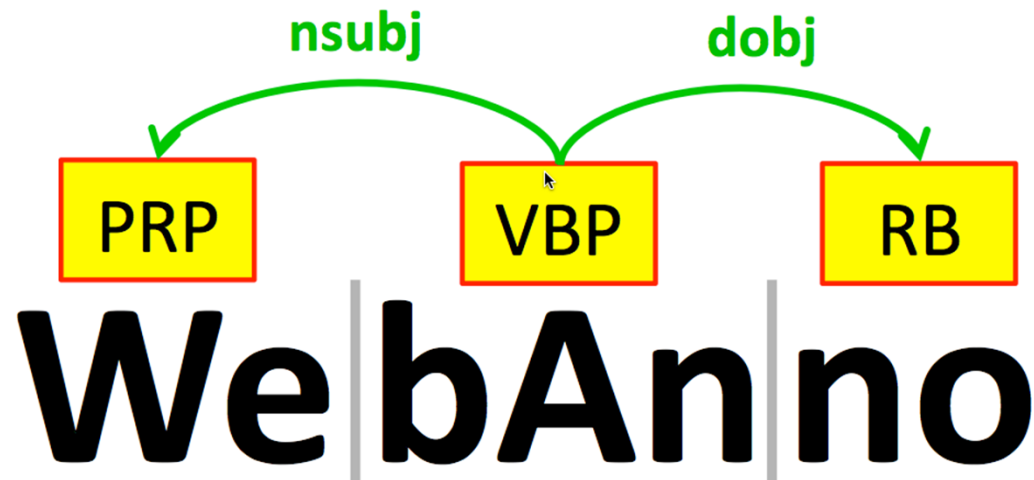


TECHNISCHE
UNIVERSITÄT
DARMSTADT

WebAnno

for less resourced and historical data annotation

Seid Muhie Yimam



Outline

-
- Introduction
 - What is WebAnno
 - Annotation UI
 - Basic Annotation layers
 - Advanced Features
 - Keyboard based annotations, Show tag description
 - Semantic annotation, Constraints
 - Automation and correction
 - Curation
 - Agreement
 - Import Export
 - WebAnno for different languages

Introduction

-
- Corpora annotated with **linguistic information** are required to develop robust statistical NLP systems.
 - Building such corpora, however, is an expensive, **labour-intensive**, and time-consuming work.
 - An annotation tool that will facilitate development of such corpora is needed.
 - Previous annotation tools
 - mostly **specific** to one annotation layer and specific language
 - mostly platform-dependent
 - mostly **single-user**, no administrative support

What is WebAnno?



-
- General purpose **web-based** annotation tool
 - Covers a wide range of linguistic annotations including various layers of **morphological, syntactic, and semantic** annotations
 - Custom annotation layers can be defined, allowing WebAnno to be used also for **non-linguistic annotation tasks**

What is WebAnno?



-
- **Multi-user** tool, also different roles such as **annotator**, **curator**, and **project manager**
 - Progress and quality of annotation projects can be **monitored** and measured in terms of **inter-annotator agreement**
 - Multiple annotation projects can be conducted in **parallel**

What is WebAnno?



-
- Different modes of annotation:
 - a **correction** mode to review externally pre-annotated data
 - **automation** mode in which WebAnno learns and offers annotation suggestions
 - **Curation** mode to adjudicate annotation disagreements
 - Fully web-based, a modern web-browser is sufficient
 - After installation on a web-server, all settings can be reached through the **browser**
 - **Open-source**

Main menu



[Annotation](#)

- Annotate texts from scratch



[Correction](#)

- Review and correct previously annotated documents



[Automation](#)

- Employ integrated machine learning capabilities



[Curation](#)

- Compare annotations from different annotators and merge them



[Monitoring](#)

- Assign workload to annotators and monitor their progress



[Projects](#)

- Create new projects

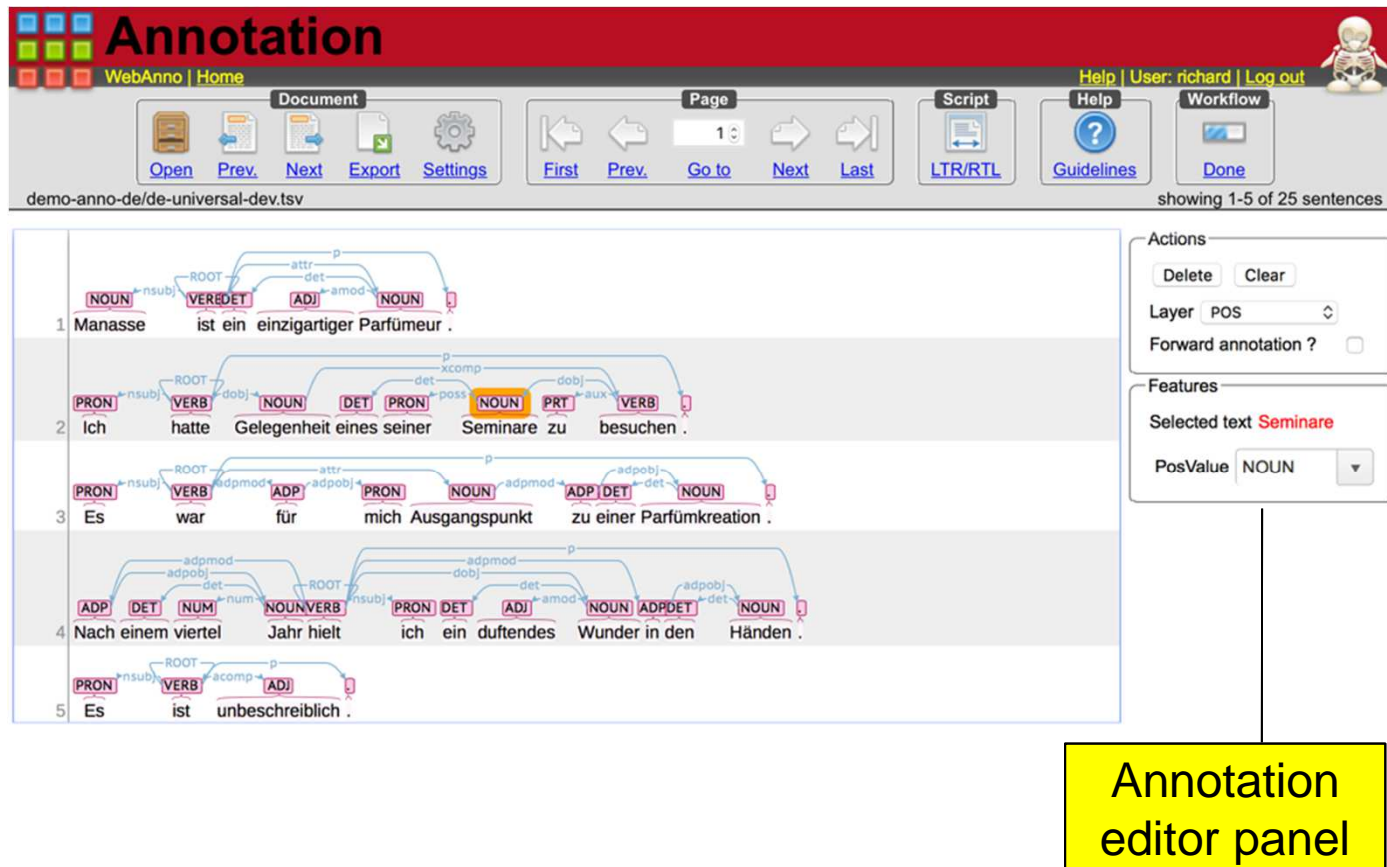


[Manage users](#)

- Create new user accounts

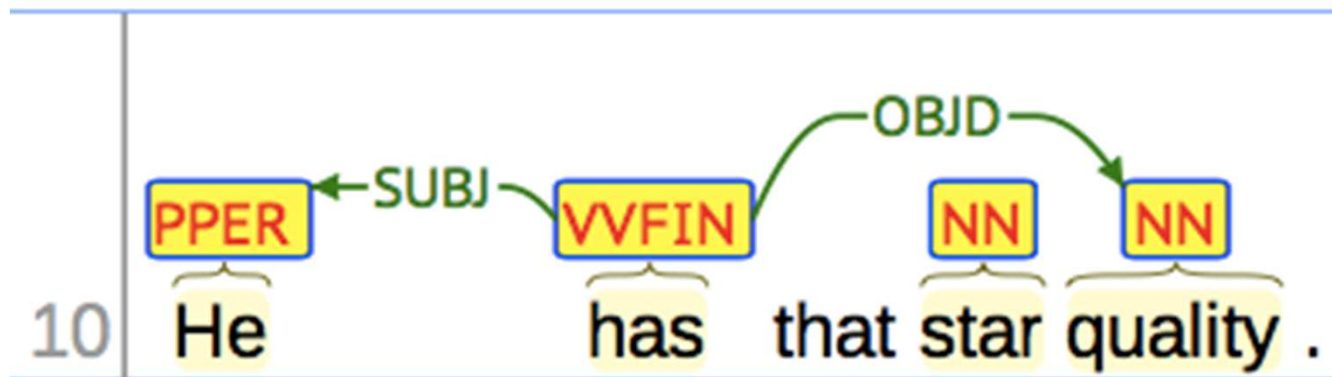
Annotation interface

- Editing elements always visible; changes take effect immediately



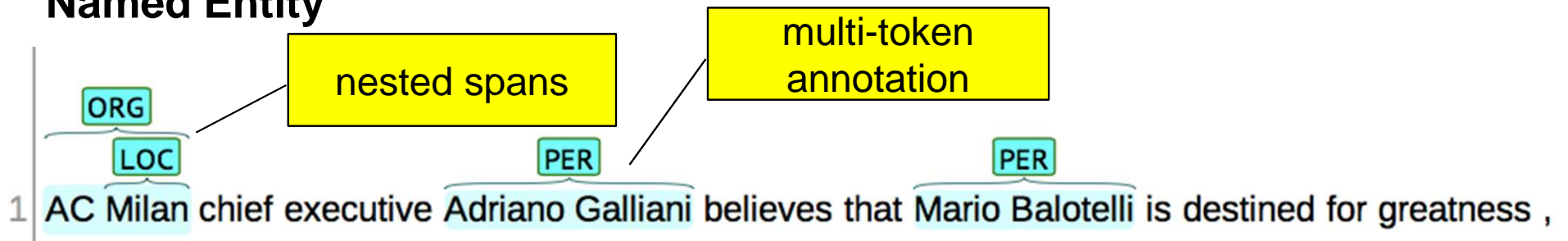
The screenshot displays the WebAnno annotation interface. At the top, there is a navigation bar with tabs for Document, Page, Script, Help, and Workflow. The Document tab is active, showing options like Open, Prev, Next, Export, and Settings. The Page tab shows navigation controls (First, Prev, Go to, Next, Last) and a page number of 1. The Script tab has LTR/RTL options. The Help tab has Guidelines. The Workflow tab has Done. Below the navigation bar, the document path is shown as 'demo-anno-de/de-universal-dev.tsv' and it indicates 'showing 1-5 of 25 sentences'. The main area shows five sentences with their corresponding POS tags and syntactic tree structures. The second sentence, 'Ich hatte Gelegenheit eines seiner Seminare zu besuchen.', is highlighted. The annotation editor panel on the right shows 'Actions' (Delete, Clear), 'Layer' set to POS, 'Forward annotation?' checkbox, 'Features' (Selected text 'Seminare'), and 'PosValue' set to NOUN. A yellow box labeled 'Annotation editor panel' points to this panel.

POS and dependency parsing

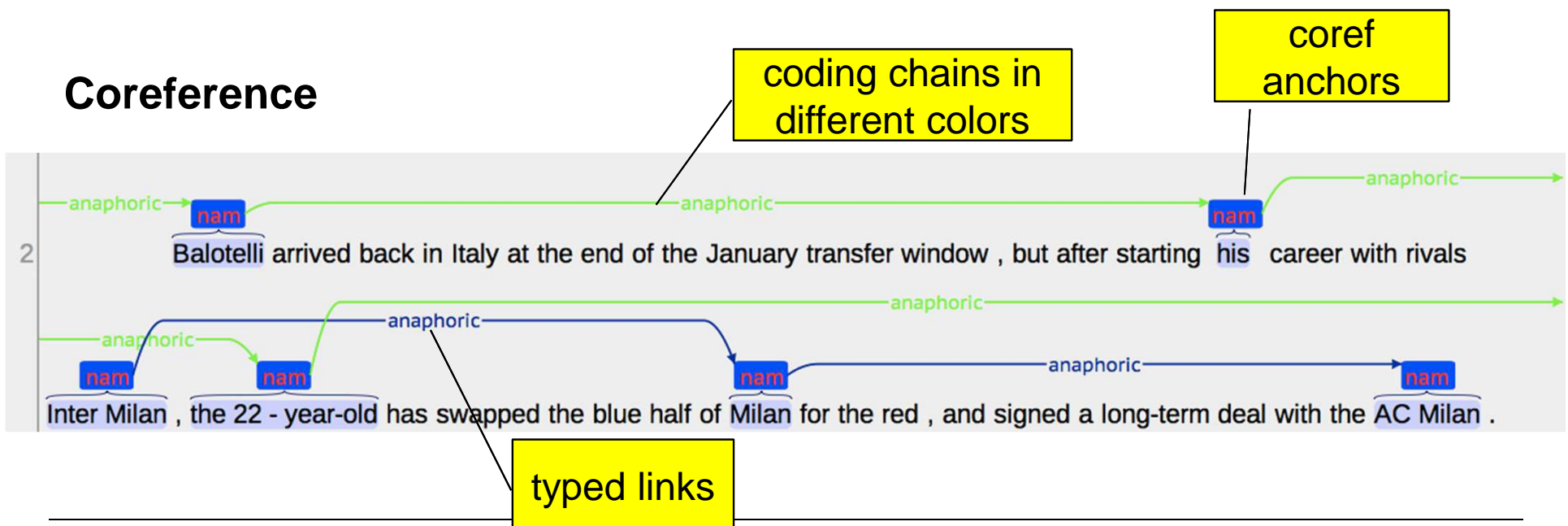


Named entity and co-reference

Named Entity

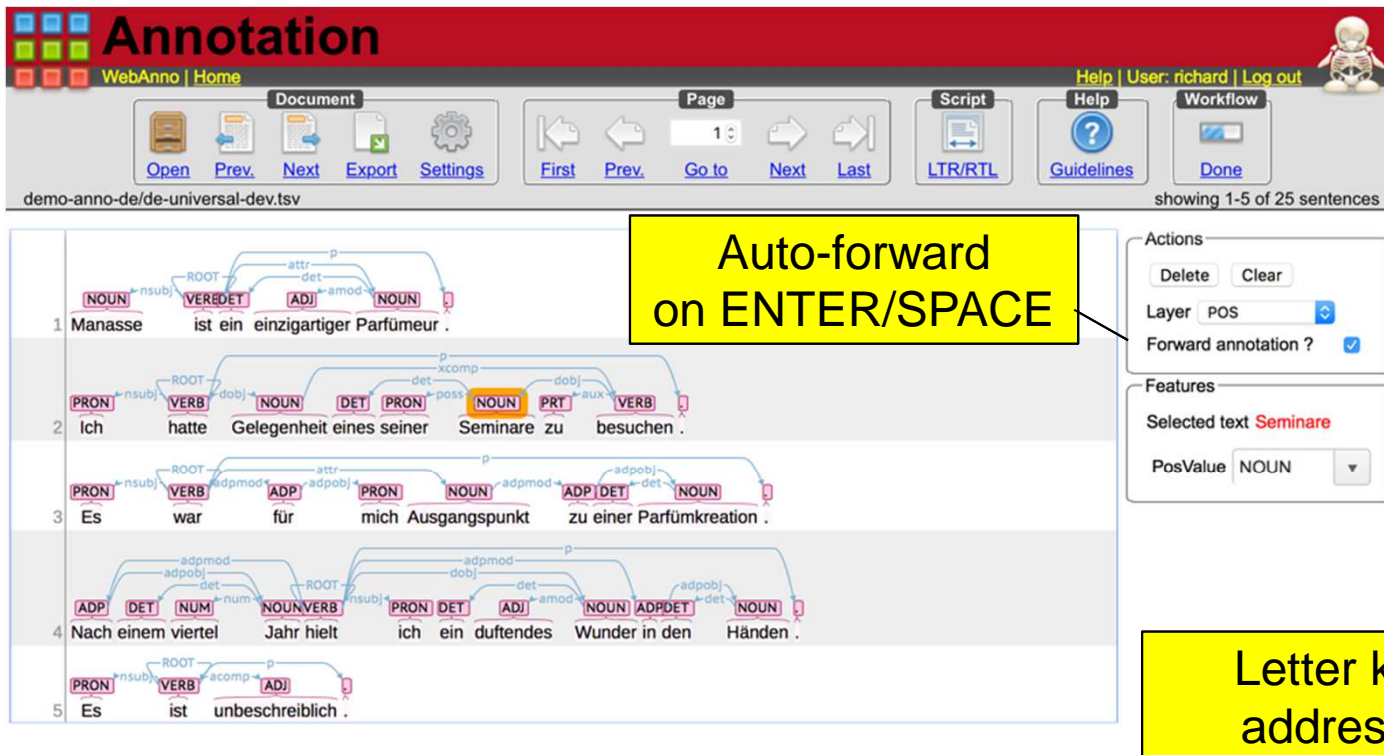


Coreference



Keyboard-based annotation

- Works for token-locked span annotations with single feature



Annotation
WebAnno | Home

demo-anno-de/de-universal-dev.tsv

Page 1 of 1

showing 1-5 of 25 sentences

1 Manasse ist ein einzigartiger Parfümeur .

2 Ich hatte Gelegenheit eines seiner Seminare zu besuchen .

3 Es war für mich Ausgangspunkt zu einer Parfümcreation .

4 Nach einem viertel Jahr hielt ich ein duftendes Wunder in den Händen .

5 Es ist unbeschreiblich .

Auto-forward on ENTER/SPACE

Letter key to address tag

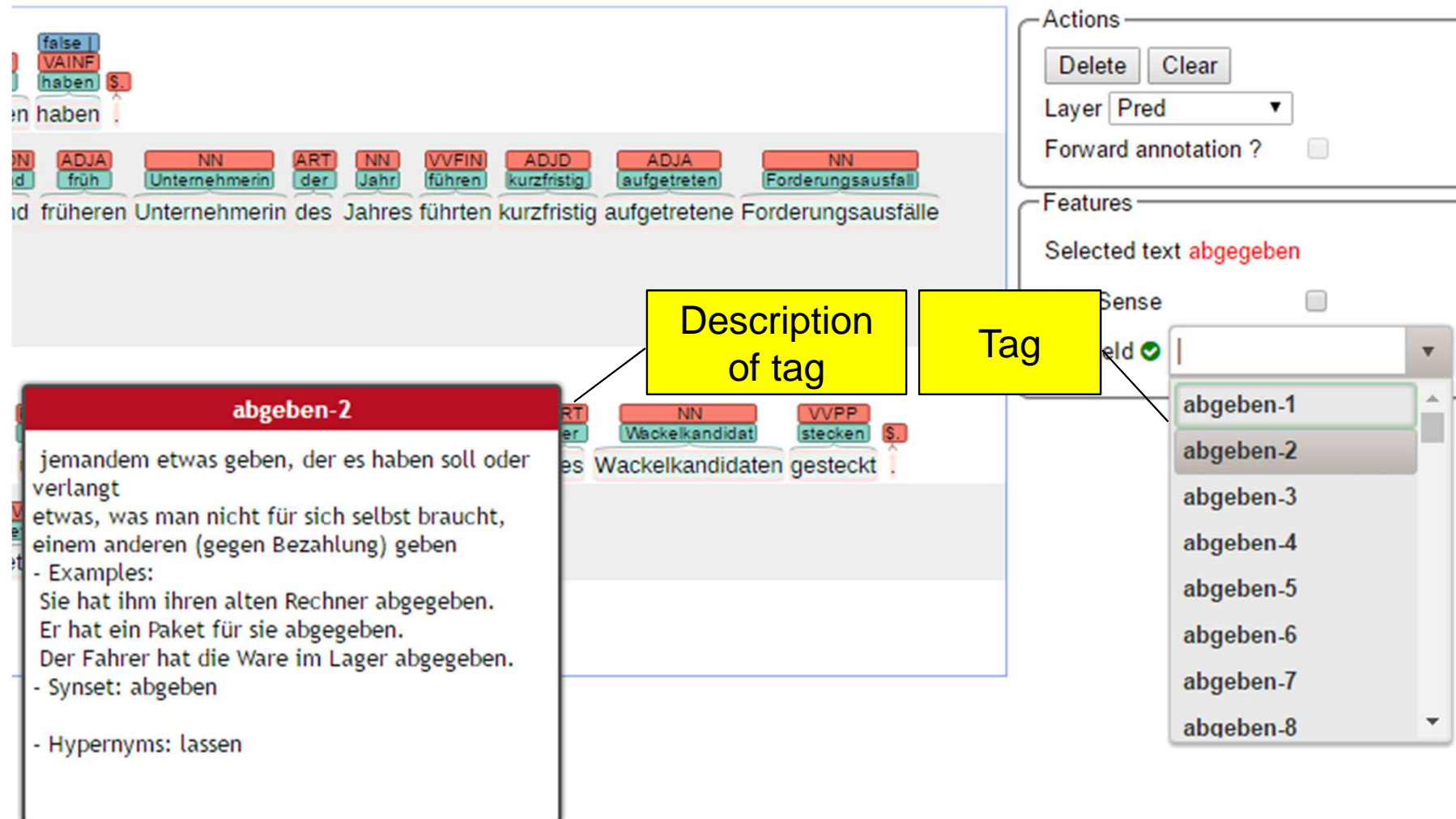
Actions: Delete, Clear

Layer: POS

Forward annotation?

Features: Selected text **Seminare**, PosValue: NOUN

Tag descriptions show as tooltips



The screenshot shows a morphological analysis interface. At the top, there's a search bar with 'false | VAINF haben S.'. Below it, a list of words is shown with their corresponding tags: 'früh' (ADJA), 'Unternehmerin' (NN), 'der' (ART), 'Jahr' (NN), 'führten' (VVFIN), 'kurzfristig' (ADJD), 'aufgetretene' (ADJA), and 'Forderungsausfälle' (NN). A tooltip for the tag 'abgeben-2' is displayed, containing the following text:

abgeben-2
jemandem etwas geben, der es haben soll oder verlangt
etwas, was man nicht für sich selbst braucht, einem anderen (gegen Bezahlung) geben
- Examples:
Sie hat ihm ihren alten Rechner abgegeben.
Er hat ein Paket für sie abgegeben.
Der Fahrer hat die Ware im Lager abgegeben.
- Synset: abgeben
- Hypernyms: lassen

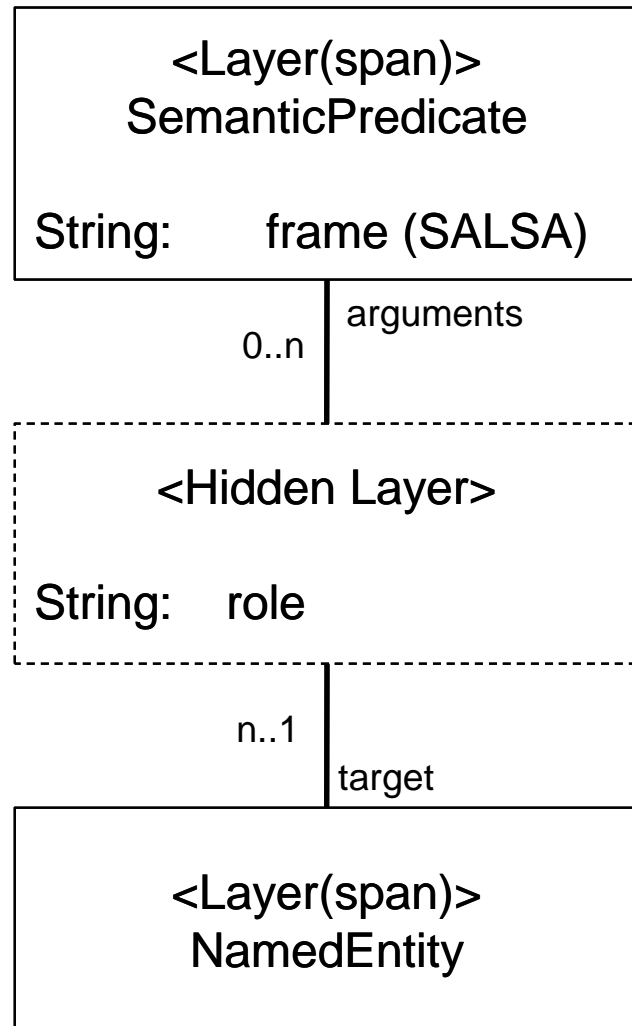
On the right side, there's a control panel with 'Actions' (Delete, Clear), 'Layer' (Pred), and 'Forward annotation?' (checkbox). Below that, 'Features' shows 'Selected text abgegeben'. A dropdown menu is open, showing a list of tags: abgeben-1, abgeben-2, abgeben-3, abgeben-4, abgeben-5, abgeben-6, abgeben-7, and abgeben-8. Two yellow boxes labeled 'Description of tag' and 'Tag' point to the tooltip and the dropdown menu respectively.

Semantic annotation

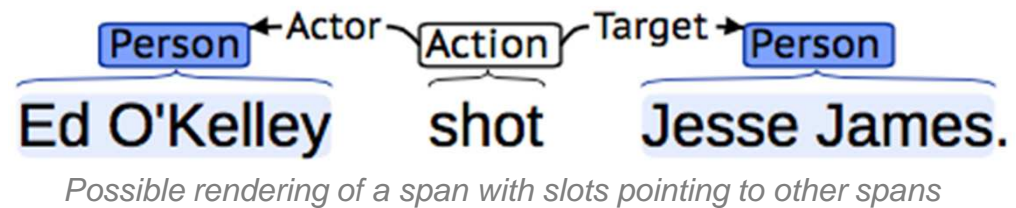
-
- Semantic role labelling
 - Frame annotation
 - Event annotation

 - What does it look like?
 - Trigger / frame-evoking element permitting certain argument slots to be filled
 - Arguments that act as slot fillers

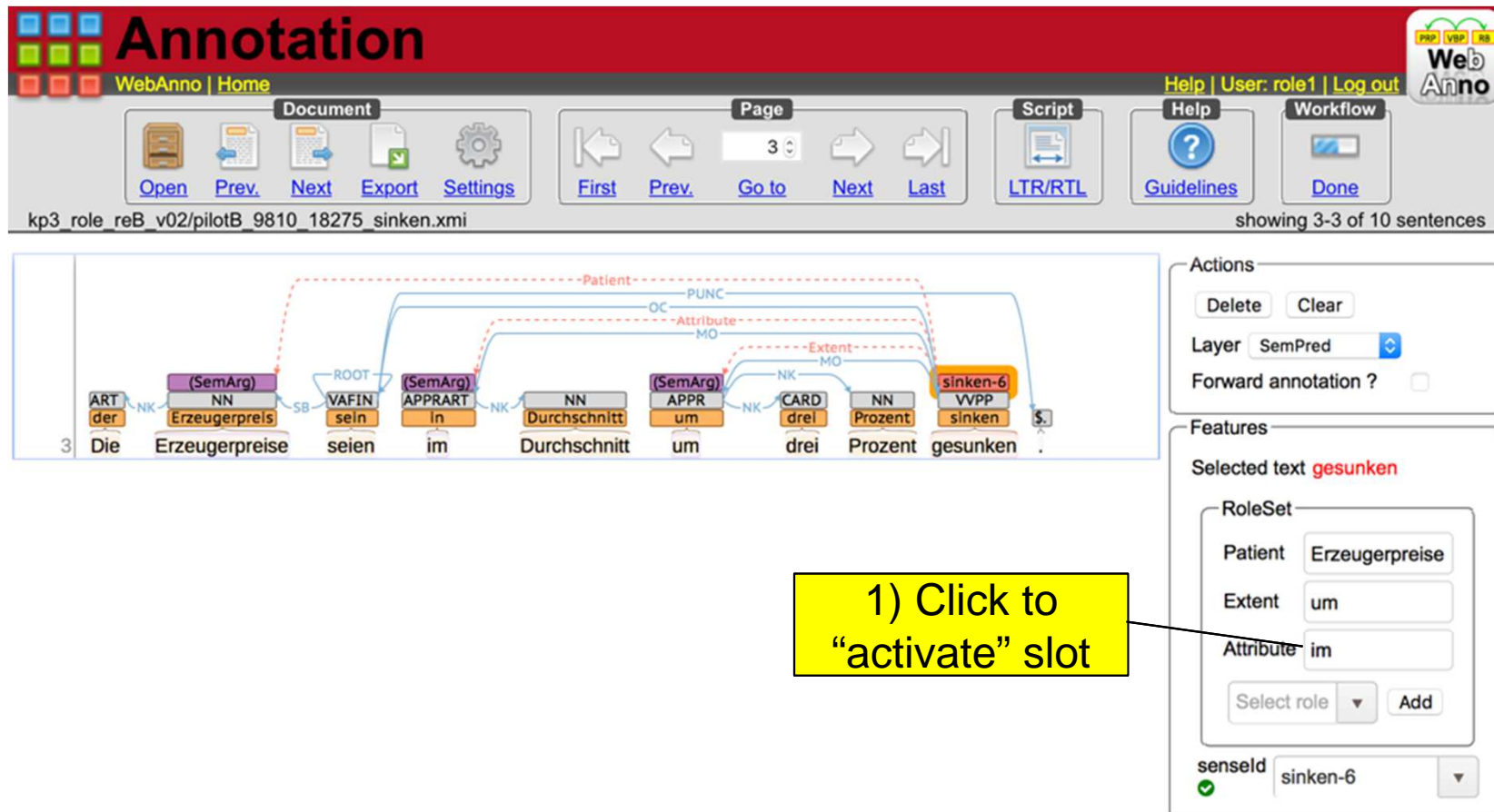
New “Slot” feature type



- **Slot** represents link to **target spans**
- Each target is associated with a **role**
- Slot features only allowed on **span layers**
- Slot can **point to many target** annotations
- Annotation can be **pointed to by many slots**

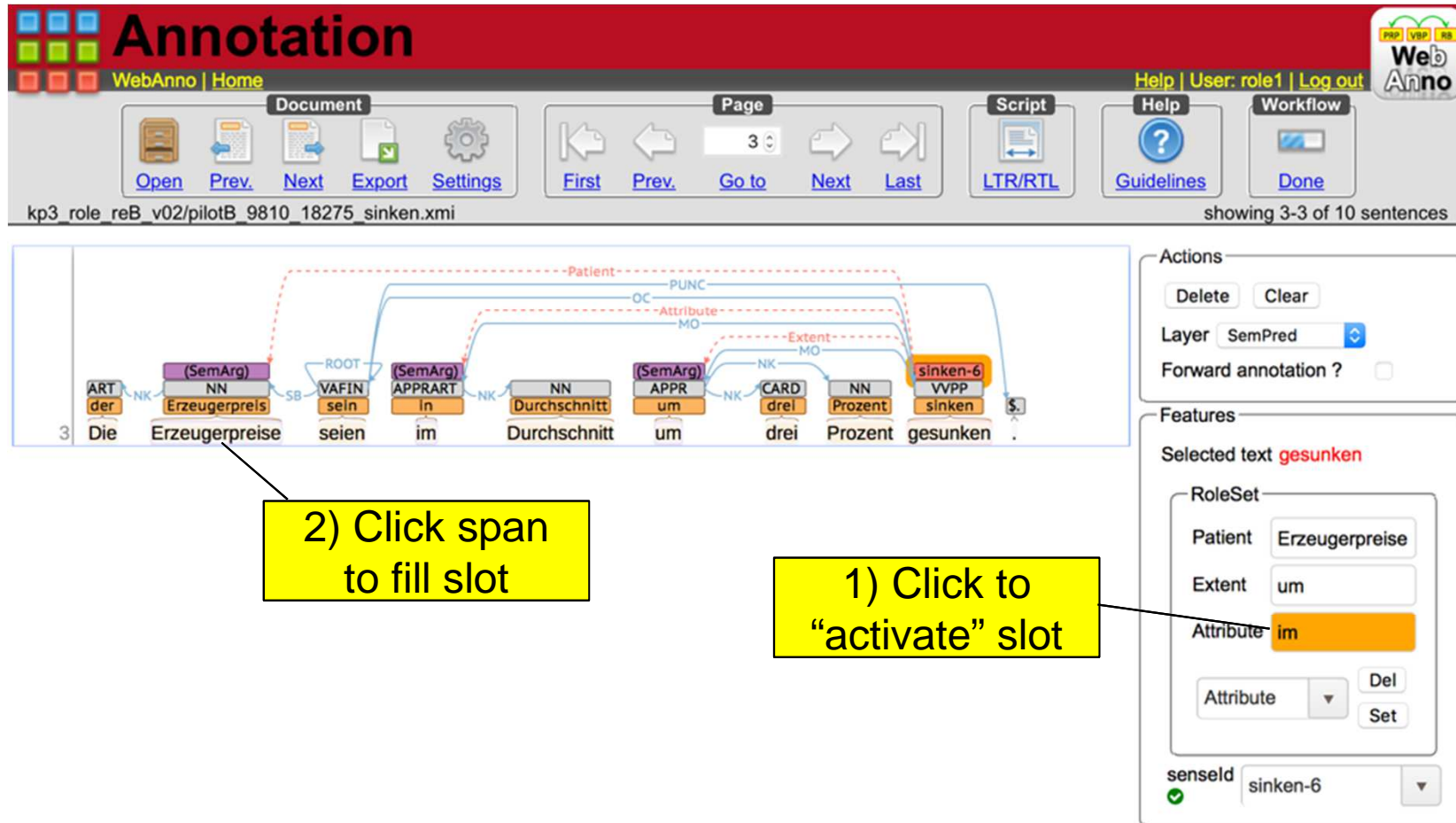


Slot features in the annotation editor



The screenshot shows the WebAnno annotation editor interface. At the top, there is a navigation bar with buttons for Document (Open, Prev., Next, Export, Settings), Page (First, Prev., Go to, Next, Last), Script (LTR/RTL), Help (Guidelines), and Workflow (Done). The current document is identified as 'kp3_role_reB_v02/pilotB_9810_18275_sinken.xml' and it is showing '3-3 of 10 sentences'. The main area displays a sentence: 'Die Erzeugerpreise seien im Durchschnitt um drei Prozent gesunken.' with various semantic annotations. A yellow box highlights the word 'gesunken' with the instruction '1) Click to "activate" slot'. On the right, the 'Features' panel shows the 'Selected text' as 'gesunken' and the 'RoleSet' with 'Patient' (Erzeugerpreise), 'Extent' (um), and 'Attribute' (im). The 'sensed' field is set to 'sinken-6' with a green checkmark.

Slot features in the annotation editor



The screenshot shows the WebAnno annotation editor interface. At the top, there's a navigation bar with "Annotation" in large letters, "WebAnno | Home", and user information "Help | User: role1 | Log out". Below this are several toolbars: "Document" (Open, Prev, Next, Export, Settings), "Page" (First, Prev, Go to, Next, Last), "Script" (LTR/RTL), "Help" (Guidelines), and "Workflow" (Done). The main area displays a sentence: "Die Erzeugerpreise seien im Durchschnitt um drei Prozent gesunken." with various semantic annotations like (SemArg), (SemPred), and (SemArg) connected by lines. A right-hand panel contains "Actions" (Delete, Clear), "Layer" (SemPred), "Forward annotation?", and "Features" (Selected text: gesunken, RoleSet: Patient: Erzeugerpreise, Extent: um, Attribute: im, senseld: sinken-6). Two yellow callout boxes provide instructions: "2) Click span to fill slot" pointing to the "im" slot, and "1) Click to 'activate' slot" pointing to the "im" slot in the RoleSet.

Constraints

- Typically certain tags are only applicable in certain contexts

“Assistance” semantic frame label only applicable to certain lemmata

Between given POS tags, only certain dependency relations possible.

- Large tagsets are unwieldy
 - Tagsets represent a controlled vocabulary – not context sensitive
- **New concept required to model context sensitivity**
 - Applied on top of tagsets, not replacing them
 - Allows annotator to focus on labels *valid* in the given context
 - Constraints are *soft* – annotator can chose to override them

Example: dependency relations

```
import de.tudarmstadt.ukp.dkpro.core.api.syntax.type.dependency.Dependency as  
    Dependency;
```

```
Dependency {  
    governor.pos.value = "NN" &  
    dependent.pos.value = "DET" ->  
        dependencyType = "DET";  
}
```

Automation – integrated machine-learning

- Annotating data from scratch is more work than correcting
- WebAnno learns from pre-annotated data and makes suggestions
- Accept suggestions with a single click
- Correct suggestions to improve training data

Annotation pane

Annotation

1 A blue moon is an extra full moon that appears in a subdivision of a year, either the third of four full moons in a season or, recently, a second full moon in a month of the common calendar.

Suggestion pane

User: Suggestion

1 A blue moon is an extra full moon that appears in a subdivision of a year, either the third of four full moons in a season or, recently, a second full moon in a month of the common calendar.

POS tagging automation for Amharic

- Initially manually annotated 21 sentences
- Iterative automatic annotation suggestion until 300 sentences are annotated.
- We obtained an F-score of 0.89 with the final model.

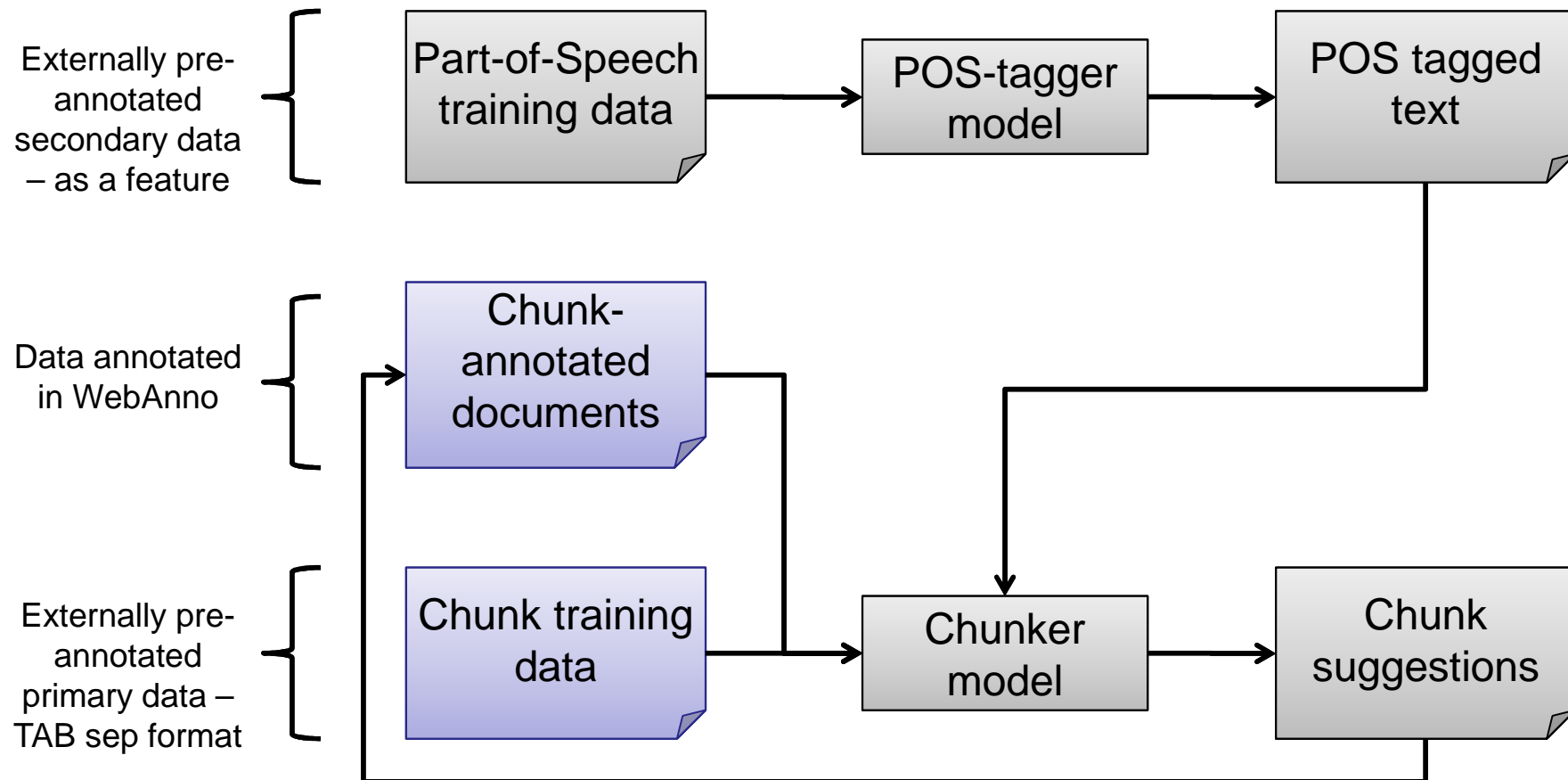
Annotation

6	PUNC	PUNC	NOUN	NOUN	NOUN	NOUN	ADV	VERB	PUNC		
	<	<	ግብፅ	የህዳሴው	ግድብ	ግንባታን	ፈፀሞ	አትፈቅድም	::		
7	PRON	VERB	NOUN	PRON	NOUN	NOUN	NOUN	NOUN	VERB	NOUN	VERB
	ይህንን	ለማሳካትም	ኢትዮጵያ	ምንም	ዓይነት	የውጭ	ዕርዳታ	ለግድቡ	አንዳታገኝ	ግፊት	አናደርጋለን

Suggestion

6	PUNC	PUNC	NOUN	NOUN	NOUN	NOUN	NOUN	VERB	PUNC		
	<	<	ግብፅ	የህዳሴው	ግድብ	ግንባታን	ፈፀሞ	አትፈቅድም	::		
7	PRON	NOUN	NOUN	PRON	NOUN	NOUN	NOUN	NOUN	VERB	NOUN	VERB
	ይህንን	ለማሳካትም	ኢትዮጵያ	ምንም	ዓይነት	የውጭ	ዕርዳታ	ለግድቡ	አንዳታገኝ	ግፊት	አናደርጋለን

Automation Example: Chunking



Correction – correct pre-annotated document

Annotation panel

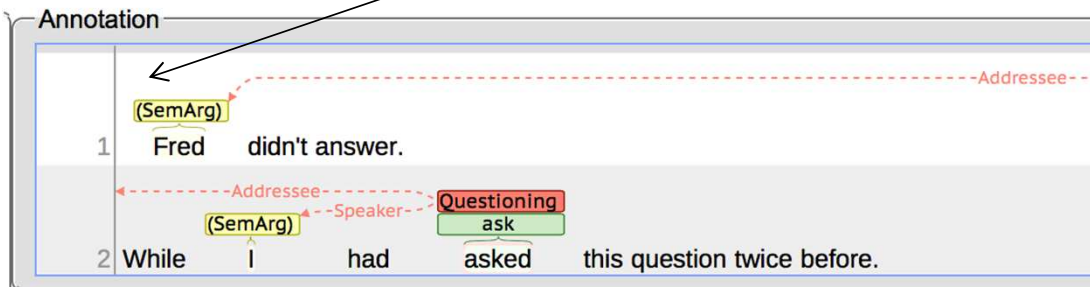
- 1 Täglich muß ich die verschwundene Gottheit wieder rufen .
- 2 Wenn ich an große Männer denke , in großen Zeiten , wie sie , ein heilig Feuer , um sich die mit ihnen aufflog zum Himmel , und dann an mich , wie ich oft , ein glimmend Lämpc die Nacht hindurch zu scheinen - siehe !
- 3 da geht ein wunderbarer Schauer mir durch alle Glieder , und leise ruf ich mir das Schre

Correction panel

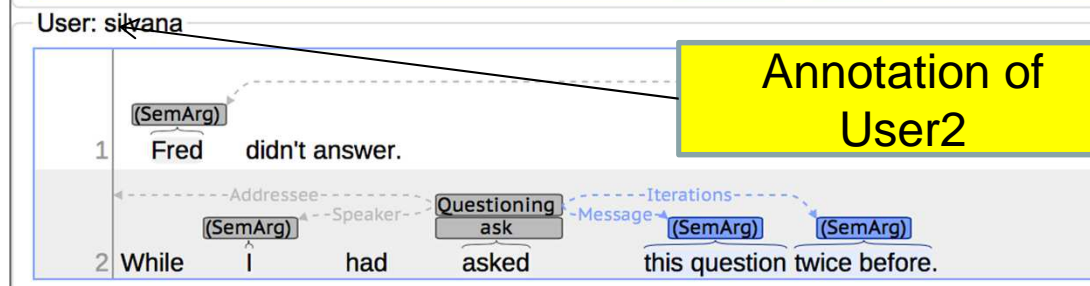
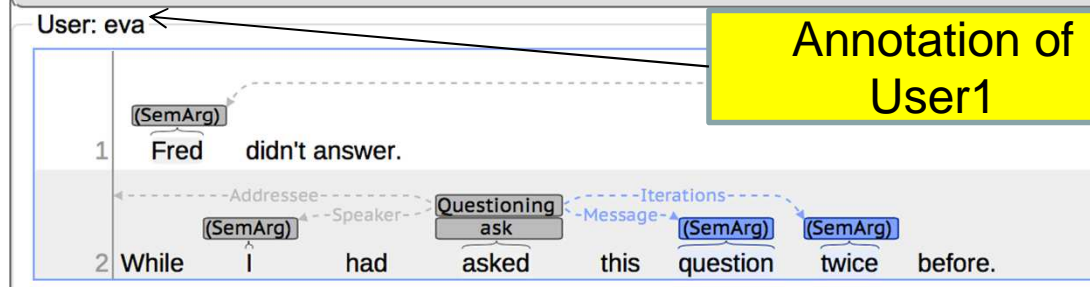
- 1 Täglich muß ich die verschwundene Gottheit wieder rufen .
- 2 Wenn ich an große Männer denke , in großen Zeiten , wie sie , ein heilig Feuer , um sich die mit ihnen aufflog zum Himmel , und dann an mich , wie ich oft , ein glimmend Lämpc die Nacht hindurch zu scheinen - siehe !
- 3 da geht ein wunderbarer Schauer mir durch alle Glieder , und leise ruf ich mir das Schre

Curation

Curated annotation



Invisible annotations in other pages



- Curation interface can handle several sentences at the same time
- indication if cross-sentence relations are not shown because they fall out of the sentence window
- improved sidebar indication for disagreements



Agreement calculation

-
- Support for multiple agreement measures
 - Cohen's Kappa
 - Fleiss' Kappa
 - Krippendorff's Alpha (nominal)
 - Improved level of detail
 - Export of raw agreement data

Agreement calculation



Agreement: [\(Help\)](#)

Feature:

Pred : bestSense
 Pred : lemmaMapped
 Pred : senseld
 Pred : senseMapped
 SemArg : head
 SemPred : RoleSet
 SemPred : aFrame
 SemPred : senseld

Measure: Krippendorff's Alpha (nominal) ▾

Exclude incomplete:

Link treatment: Use link role as label ▾

Export format: CSV ▾

Legend:

- Lower triangle: Positions included in agreement / total number of positions
- Upper triangle: Agreement score - click on cell to export agreement report

	anette	eva	role1	role2	role3	silvana
anette	-	no positions	no labels from anette	no labels from anette	no positions	no positions
eva	0/0	-	no labels from eva	no labels from eva	no positions	no positions
role1	406/407	406/407	-	0.55	no labels from role3	no labels from silvana
role2	396/405	396/405	398/408	-	no labels from role3	no labels from silvana
role3	0/0	0/0	406/407	396/405	-	no positions
silvana	0/0	0/0	406/407	396/405	0/0	-

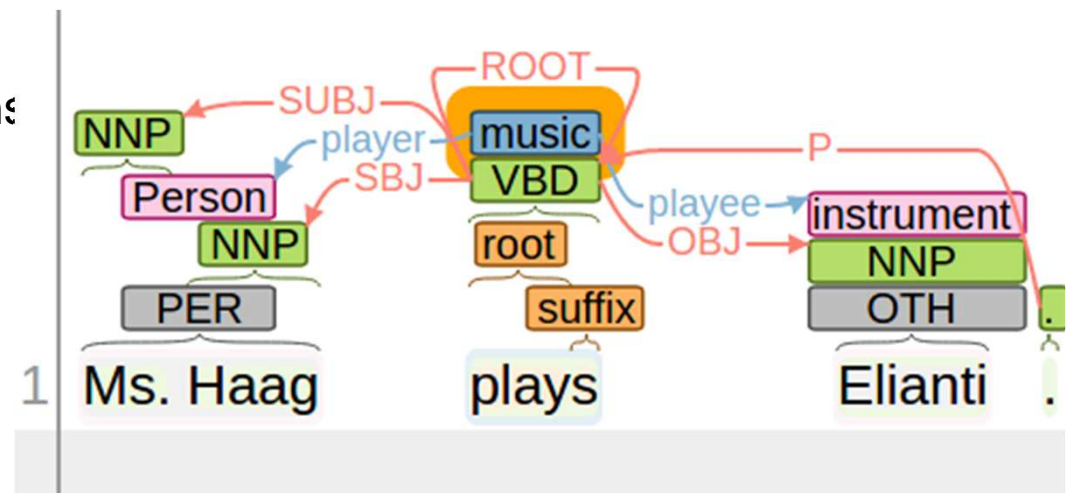
Agreement export



Type	Collection	Document	Layer	Feature	Position	role1	role2
SpanPosition	kp3_role_reb_v02	pilot8_9810_18275_ablehnen.xml	webanno.custom.SemPred	RoleSet	243-252 [ablehnten] -> [206-215 [Aktionäre]	Agent	Agent
SpanPosition	kp3	Feature				role1	role2
SpanPosition	kp3	Position					Topic
SpanPosition	kp3	RoleSet				Agent	Agent
SpanPosition	kp3	243-252 [ablehnten] -> [227-242 [Kapitalerhöhung]				Beneficiary	Topic
SpanPosition	kp3	RoleSet				Beneficiary	Topic
SpanPosition	kp3	360-371 [abgeschafft] -> [335-346 [Wehrpflicht]				Patient	Patient
SpanPosition	kp3	RoleSet				Patient	Patient
SpanPosition	kp3	628-640 [abzuschaffen] -> [582-594 [Beschränkung]				Patient	Theme
SpanPosition	kp3	RoleSet				Patient	Theme
SpanPosition	kp3	394-404 [anzubieten] -> [350-352 [er]				Agent	Agent
SpanPosition	kp3	RoleSet				Agent	Agent
SpanPosition	kp3	394-404 [anzubieten] -> [380-386 [Gästen]				Recipient	Recipient
SpanPosition	kp3	RoleSet				Recipient	Recipient
SpanPosition	kp3	394-404 [anzubieten] -> [387-393 [Kaffee]				Theme	Theme
SpanPosition	kp3	RoleSet				Theme	Theme
SpanPosition	kp3	1230-1236 [bietet] -> [1215-1229 [Andrianopoulos]				Agent	Agent
SpanPosition	kp3	RoleSet				Agent	Agent
SpanPosition	kp3	1230-1236 [bietet] -> [1245-1253 [Hunderterte]				Theme	Extent
SpanPosition	kp3	RoleSet				Theme	Extent
SpanPosition	kp3	1230-1236 [bietet] -> [1284-1287 [zum]				Goal	Theme
SpanPosition	kp3	RoleSet				Goal	Theme
SpanPosition	kp3	1905-1908 [bot] -> [1913-1924 [Sowjetunion]				Agent	Agent
SpanPosition	kp3	RoleSet				Agent	Agent
SpanPosition	kp3	1905-1908 [bot] -> [2027-2036 [abzugeben]				Theme	null
SpanPosition	kp3	RoleSet				Theme	null
SpanPosition	kp3	265-271 [führen] -> [253-256 [Als]				Manner	Attribute
SpanPosition	kp3_role_reb_v02	pilot8_9810_18275_arbeiten.xml	webanno.custom.SemPred	RoleSet	317-325 [arbeiten] -> [345-353 [Menschen]	Agent	Agent

All-inclusive TSV export/import

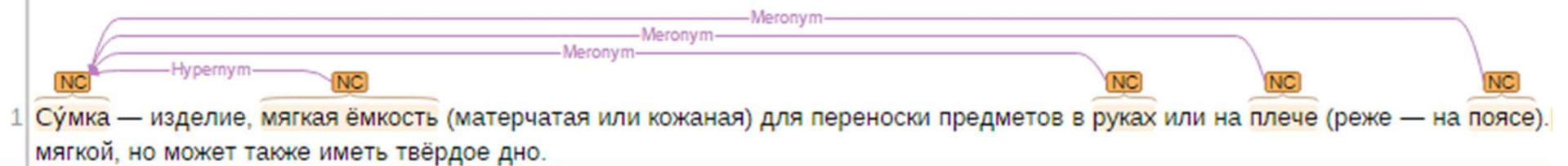
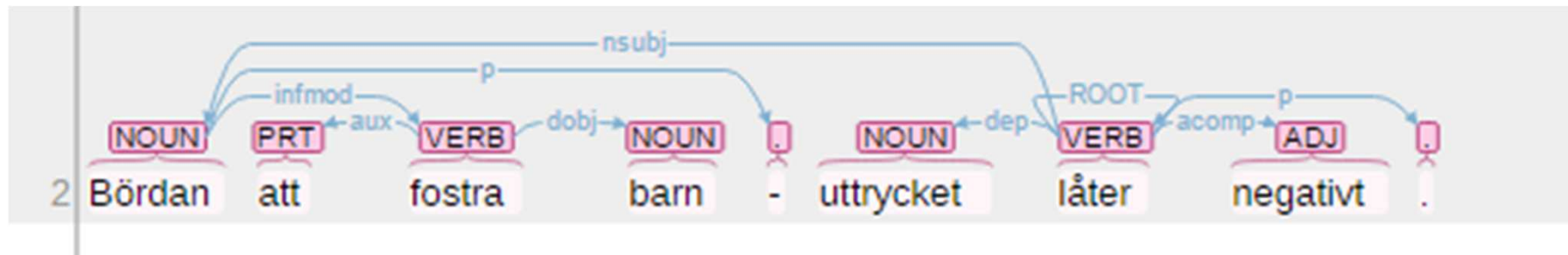
- Off course, WebAnno is the beginning
 - Export data in human readable format
 - Re-import for further annotation
 - Support almost all kind of
 - Sub-token annotation
 - Chain and relation annotations
 - Slot features
 - Multiple and stacked annotations



Other import/export formats

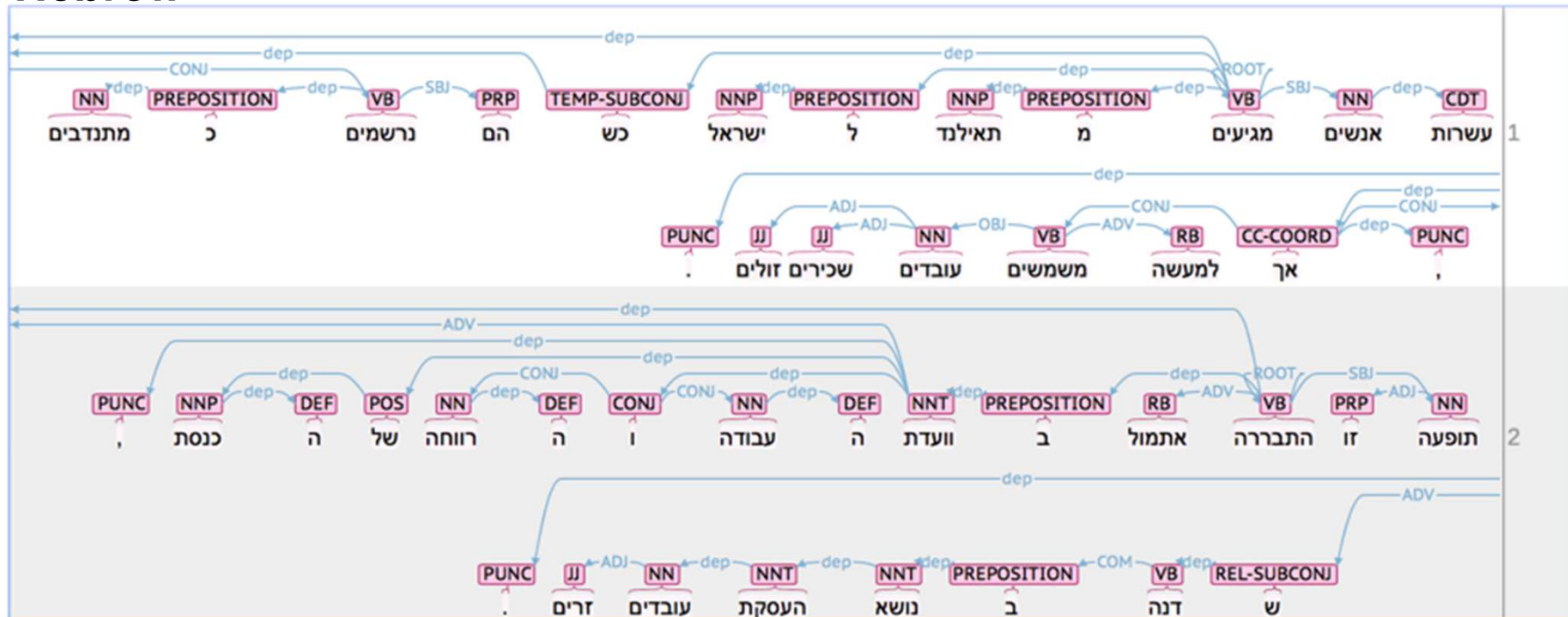
- WebLicht Text Corpus Format TCF
- Different CoNNL formats (CoNNL 2000, 2002, 2006, 2009, U)
- UIMA XMI
- UIMA CAS
- WebAnno CPH TEI (only imports)
- Plain text

WebAnno for Swedish and Russian



Right-to-left support

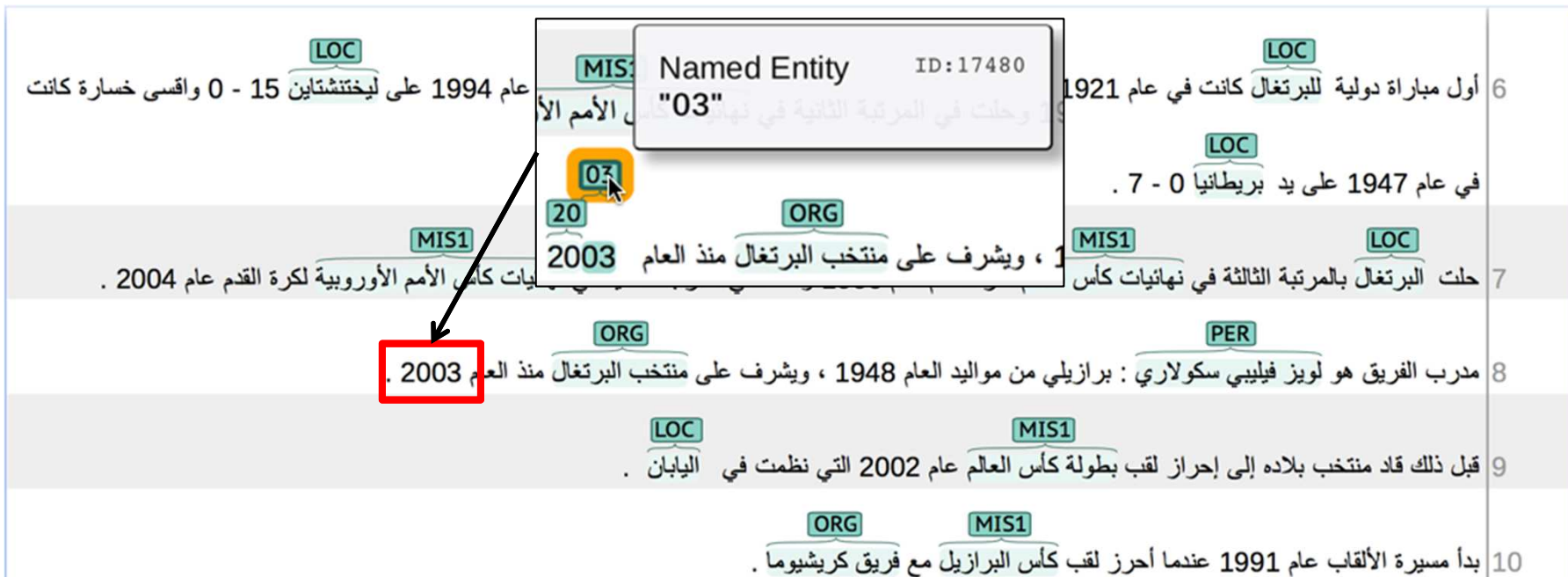
Hebrew



- Subtoken annotation in LTR tokens within RTL texts (e.g. numbers)
- Cross-browser compatibility

Right-to-left support

Arabic



6	أول مباراة دولية للبرتغال كانت في عام 1921	LOC
7	في عام 1947 على يد بريطانيا 0 - 7 .	LOC
7	حلت البرتغال بالمرتبة الثالثة في نهائيات كأس	MIS1 LOC
8	مدرّب الفريق هو لويز فيليب سكولاري : برازيلي من مواليد العام 1948 ، ويشرف على منتخب البرتغال منذ العام 2003 .	ORG PER
9	قبل ذلك قاد منتخب بلاده إلى إحراز لقب بطولة كأس العالم عام 2002 التي نظمت في اليابان .	LOC MIS1
10	بدأ مسيرة الألقاب عام 1991 عندما أحرز لقب كأس البرازيل مع فريق كريشيوما .	ORG MIS1

- Subtoken annotation in LTR tokens within RTL texts (e.g. numbers)
- Cross-browser compatibility

Amharic

PUNC PUNC NOUN NOUN NOUN NOUN ADV VERB PUNC
< < ግብፅ የህዳሴው ግድብ ግንባታን ፈፀሞ አትፈቅድም ::

PRON VERB NOUN PRON NOUN NOUN NOUN NOUN VERB NOUN VERB
ይህንን ለማሳካትም ኢትዮጵያ ምንም ዓይነት የውጭ ዕርዳታ ለግድቡ አንዳታገኝ ግፊት አናደርጋለን

Future work for historical data

- Importing OCR/PDF documents
- Enable editing source documents (OCR/PDF)
- Curation for the editing of documents
- Agreement for the editing of OCR data
- Curation interface for document translations

Thank you