

# **Serengeti – webbasierte semi-automatische Annotation semantischer Relationen**

Nils Diewald, Anna Garbar, Daniela  
Goecke, Maik Stührenberg

- Projekt *Sekimo*
- Annotationschema zur Annotation semantischer Relationen
- Annotationswerkzeuge
- Serengeti
- Zusammenfassung und Ausblick

- **Projekt *Sekimo***
- Annotationschema zur Annotation semantischer Relationen
- Annotationswerkzeuge
- Serengeti
- Zusammenfassung und Ausblick

- Sekundäre Informationsstrukturierung und Verknüpfung heterogener texttechnologischer Ressourcen
- Teilprojekt der DFG Forschergruppe 437 „Texttechnologische Informationsmodellierung“
- Projektziele:
  1. Entwicklung einer Architektur zur Integration heterogener Texttechnologischer Ressourcen
  2. Definition eines abstrakten Datenformats für Multi-Ebenen-Annotation
  3. Anwendungsdomäne: Anaphernresolution, speziell die Auflösung von Bridging-Relationen definiter Beschreibungen

## Anwendungsdomäne:

- Anaphorische Relation bestehen zwischen einem phorischen Element und einem oder mehreren Antezedenzien
- Anaphorische Relationen werden zwischen Diskursentitäten annotiert
- Für die Auflösung anaphorischer Relationen, speziell Bridging-Relationen, wird eine Vielzahl von Informationen benötigt
- Linguistische Ressourcen werden mittels Markup-Unifikation miteinander verknüpft
- Jede Diskursentität ist durch ein XML Element und dessen Attribute beschrieben und eindeutig identifizierbar
- XML Element- und Attributinformationen werden als Merkmalsvektoren reinterpretiert

## Anforderungen:

- Die Integration der Ressourcen sowie die Anwendungsressource soll formal evaluierbar sein (Precision/Recall)
- Realisierung der Anaphernauflösung mittels Entscheidungsbäumen (C4.5)  
Die Eignung von Entscheidungsbäumen für die Auflösung anaphorischer Relationen wurde bereits von Xiaofeng et al. (2004); Strube & Müller (2003) gezeigt
- *hier*: Fokus auf Integration und Evaluation verschiedener Ressourcen, d.h. verschiedener Merkmalssets

Herangehensweise:

- Erstellen eines Korpus anaphorischer Relationen für
  - Korpustudien
  - Training und Evaluation

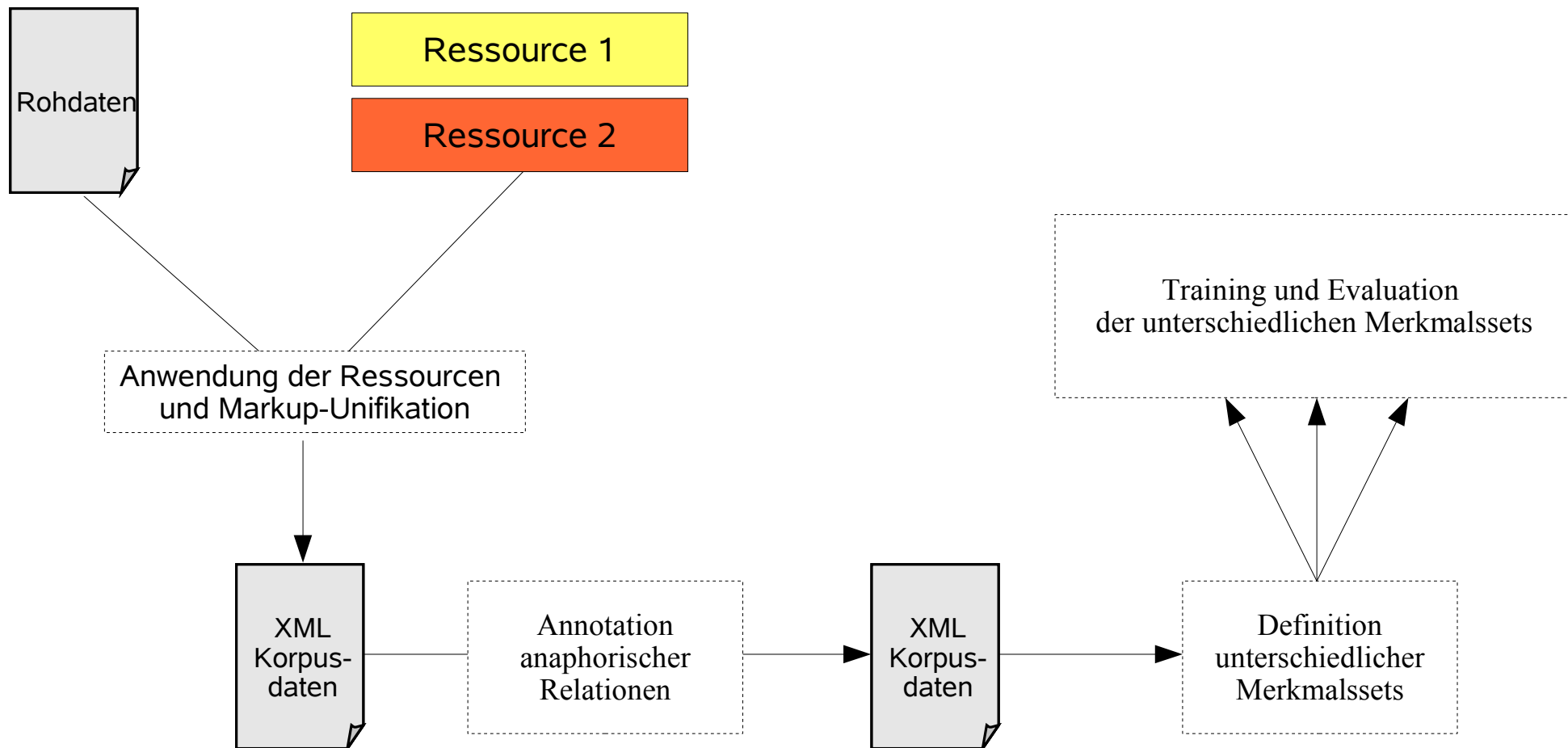
Die Korpuserstellung umfasst:

- die Definition eines Annotationsschemas
- Zusammenstellung des Korpus (C1 Korpus + Zeitungsartikel)
- die Annotation der Daten

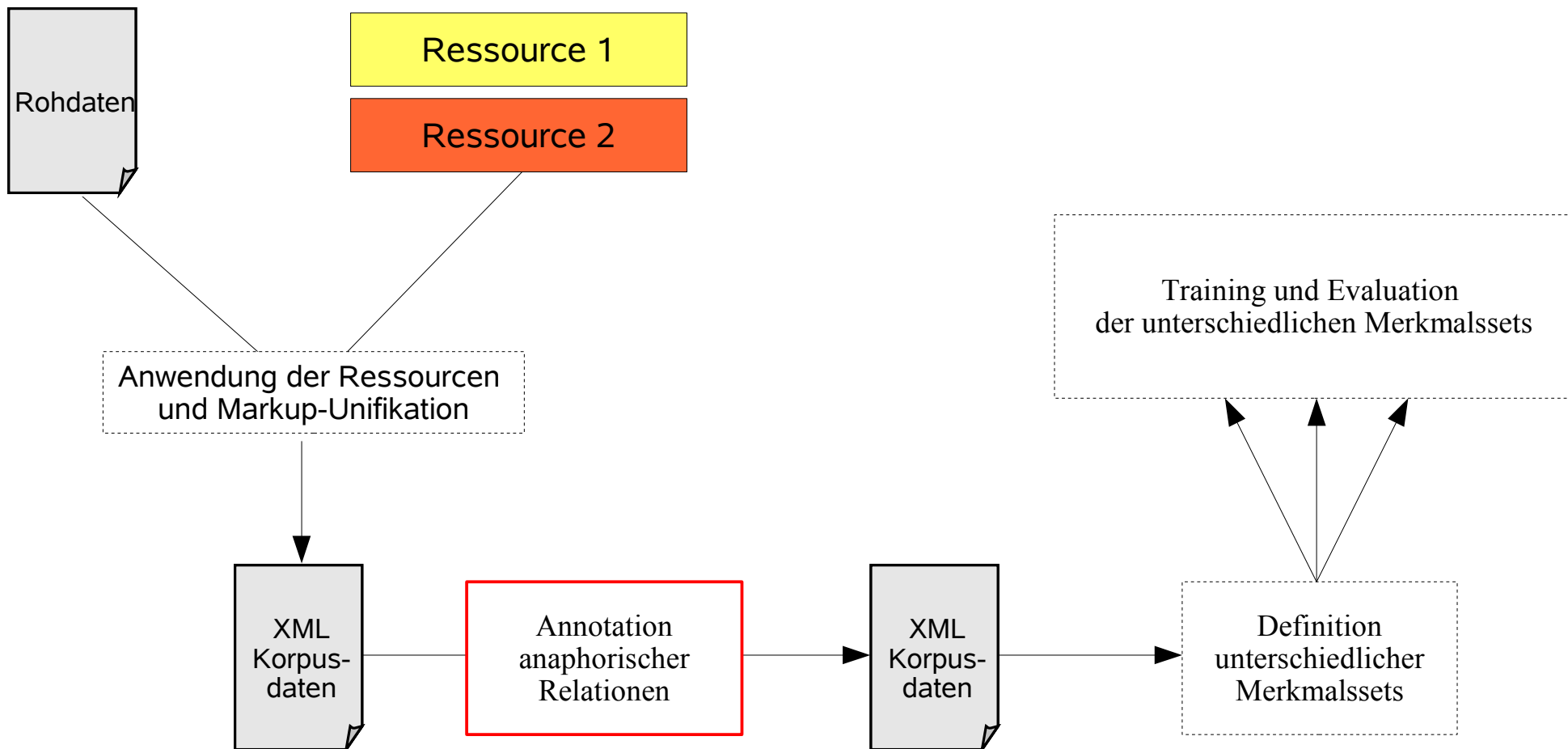
Gewünscht:

- Die Unterstützung der Annotationsarbeit durch ein Annotationswerkzeug, das die folgende Arbeitsschritte ermöglicht:
  - Annotation der Daten
  - Analyse der Annotationen auf Vollständigkeit und Inter-Annotator-Vergleich

## Die Anwendungsdomäne im Überblick



## Die Anwendungsdomäne im Überblick



- Projekt *Sekimo*
- **Annotationschema zur Annotation semantischer Relationen**
- Annotationswerkzeuge
- Serengeti
- Zusammenfassung und Ausblick

Das verwendete Annotationsschema...

- ...ist entwicklungshistorisch gesehen eine Erweiterung des von Holler (2004) vorgestellten Schemas
- ...modelliert anaphorische Relationen als XML Standoff-Annotation

Dem Annotationsschema zugrunde liegt eine Taxonomie hinsichtlich...

- ...des NP-Typs
  - Pronominalanaphern
  - NP-Anaphern (Definite Beschreibungen)
- ...des Typs der anaphorischen Relation
  - Kospezifikation
  - Bridging

- cospecLink
  - ident
  - propName
  - synonym
  - hyperonym
  - hyponym
  - paraphrase
  - addInfo
  - isA
- bridgingLink
  - poss
  - setMember
  - hasMember
  - meronym
  - holonym
  - bridging

- cospecLink
  - ident
  - propName
  - synonym
  - hyperonym
  - hyponym
  - paraphrase
  - addInfo
  - isA
- bridgingLink
  - poss
  - setMember
  - hasMember
  - meronym
  - holonym
  - bridging
- Relationen bestehen zwischen Diskursentitäten
- Diskursentitäten als Markables
- *cospecLink*: Referenzidentität zwischen dem anaphorischen Element und dem Antezedens/den Antezedentien
- *bridgingLink*: keine Referenzidentität, das Antezedens muss aus dem Kontext inferiert werden

- cospecLink
  - ident
  - propName
  - synonym
  - hyperonym
  - hyponym
  - paraphrase
  - addInfo
  - isA
- bridgingLink
  - poss
  - setMember
  - hasMember
  - meronym
  - holonym
  - bridging

ein Mann – er

die Großmutter – die Oma

Lehrwerke – die DAF-Lehrwerke

Jacqueline Schulte – die 20jährige

die Kinder – ihre Lebenschancen

Maik Marie – die Kinder

Kindergeburtstage – Geschenke

- Relationen bestehen zwischen Antezedens/ten
- Diskursentitäten als Markables
- *cospecLink*: Referenzidentität zwischen dem anaphorischen Element und dem Antezedens/den Antezedentien
- *bridgingLink*: keine Referenzidentität, das Antezedens muss aus dem Kontext inferiert werden

Annotiert werden

- Diskursentitäten
  - nominal
  - namedEntity
  - propositional (in Kooperation mit C1)
- Vollständige Annotation für Pronomen und definite Beschreibungen

Maik hat kein eigenes Fahrrad und Marie fährt nicht in den Urlaub. Zwei Kinder, eine Gemeinsamkeit.

```
<de deID="de8" deType="namedEntity" headRef="w36">
  <token ref="w36">Maik</token>
</de>
<token ref="w37">hat</token> <token ref="w38">kein</token>
<token ref="w39">eigenes</token> <token ref="w40">Fahrrad</token>,
<token ref="w42">und</token>
<de deID="de10" deType="namedEntity" headRef="w43">
  <token ref="w43">Marie</token>
</de>
<token ref="w45">fährt</token> <token ref="w46">nicht</token>
<token ref="w47">in</token>
<de deID="de11" deType="nom" headRef="w49">
  <token ref="w48">den</token>
  <token ref="w49">Urlaub</token>
</de>
<de deID="de12" deType="nom" headRef="w53">
  <token ref="w52">Zwei</token>
  <token ref="w53">Kinder</token>
</de>,
<de deID="de13" deType="nom" headRef="w56">
  <token ref="w55">eine</token>
  <token ref="w56">Gemeinsamkeit</token>
</de>.
<token_ref id="w36" dependHead="w37" pos="N" syntax="@NH" dependValue="subj"
  morpho="MSC SG NOM" text="Maik" lemma="maik" />
<semRel>
  <bridgingLink relType="hasMember" antecedentIDRefs="de8 de10" phorIDRef="de12"/>
</semRel>
```

Maik hat kein eigenes Fahrrad und Marie fährt nicht in den Urlaub. Zwei Kinder, eine Gemeinsamkeit.

```
<de deID="de8" deType="namedEntity" headRef="w36">
  <token ref="w36">Maik</token>
</de>
<token ref="w37">hat</token> <token ref="w38">kein</token>
<token ref="w39">eigenes</token> <token ref="w40">Fahrrad</token>,
<token ref="w42">und</token>
<de deID="de10" deType="namedEntity" headRef="w43">
  <token ref="w43">Marie</token>
</de>
<token ref="w45">fährt</token> <token ref="w46">nicht</token>
<token ref="w47">in</token>
<de deID="de11" deType="nom" headRef="w49">
  <token ref="w48">den</token>
  <token ref="w49">Urlaub</token>
</de>
<de deID="de12" deType="nom" headRef="w53">
  <token ref="w52">Zwei</token>
  <token ref="w53">Kinder</token>
</de>,
<de deID="de13" deType="nom" headRef="w56">
  <token ref="w55">eine</token>
  <token ref="w56">Gemeinsamkeit</token>
</de>.
<token_ref id="w36" dependHead="w37" pos="N" syntax="@NH" dependValue="subj"
  morpho="MSC SG NOM" text="Maik" lemma="maik" />
<semRel>
  <bridgingLink relType="hasMember" antecedentIDRefs="de8 de10" phorIDRef="de12"/>
</semRel>
```

Maik hat kein eigenes Fahrrad und Marie fährt nicht in den Urlaub. Zwei Kinder, eine Gemeinsamkeit.

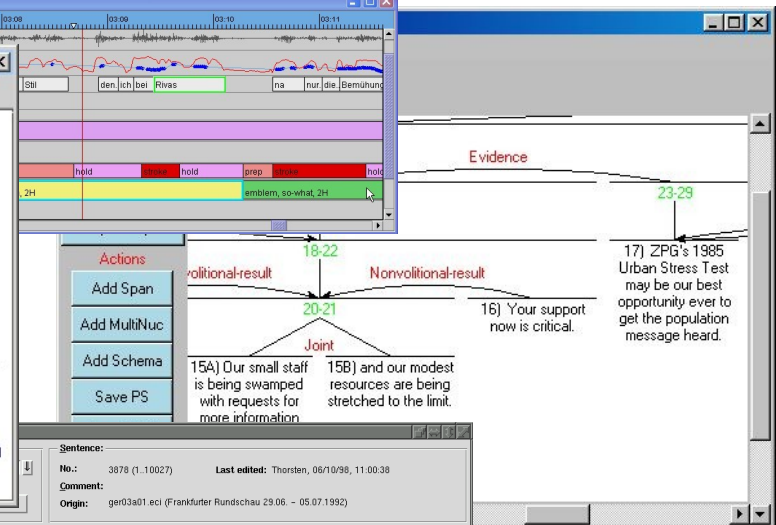
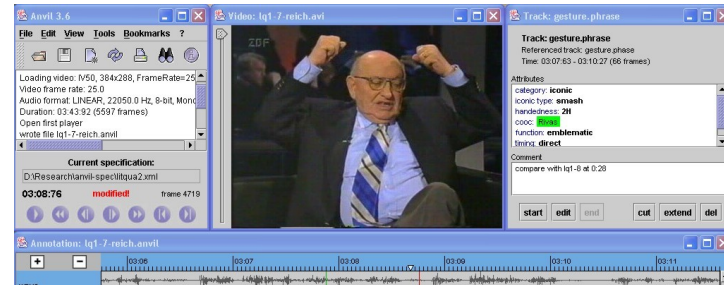
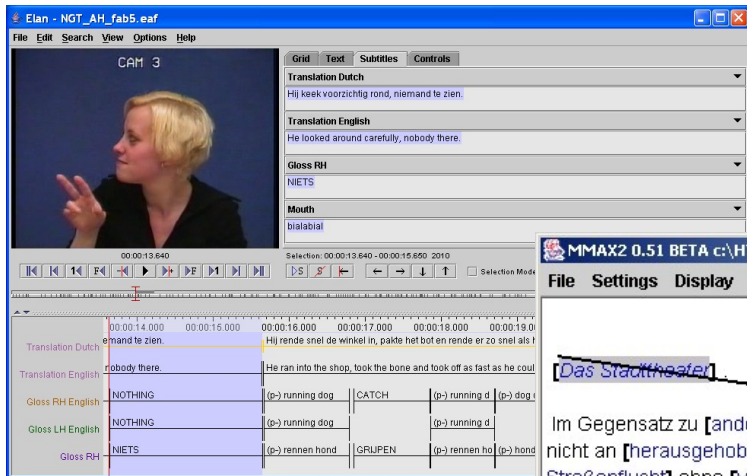
```
<de deID="de8" deType="namedEntity" headRef="w36">
  <token ref="w36">Maik</token>
</de>
<token ref="w37">hat</token> <token ref="w38">kein</token>
<token ref="w39">eigenes</token> <token ref="w40">Fahrrad</token>,
<token ref="w42">und</token>
<de deID="de10" deType="namedEntity" headRef="w43">
  <token ref="w43">Marie</token>
</de>
<token ref="w45">fährt</token> <token ref="w46">nicht</token>
<token ref="w47">in</token>
<de deID="de11" deType="nom" headRef="w49">
  <token ref="w48">den</token>
  <token ref="w49">Urlaub</token>
</de>
<de deID="de12" deType="nom" headRef="w53">
  <token ref="w52">Zwei</token>
  <token ref="w53">Kinder</token>
</de>,
<de deID="de13" deType="nom" headRef="w56">
  <token ref="w55">eine</token>
  <token ref="w56">Gemeinsamkeit</token>
</de>.
<token_ref id="w36" dependHead="w37" pos="N" syntax="@NH" dependValue="subj"
  morpho="MSC SG NOM" text="Maik" lemma="maik" />
<semRel>
  <bridgingLink relType="hasMember" antecedentIDRefs="de8 de10" phorIDRef="de12"/>
</semRel>
```

Maik hat kein eigenes Fahrrad und Marie fährt nicht in den Urlaub. Zwei Kinder, eine Gemeinsamkeit.

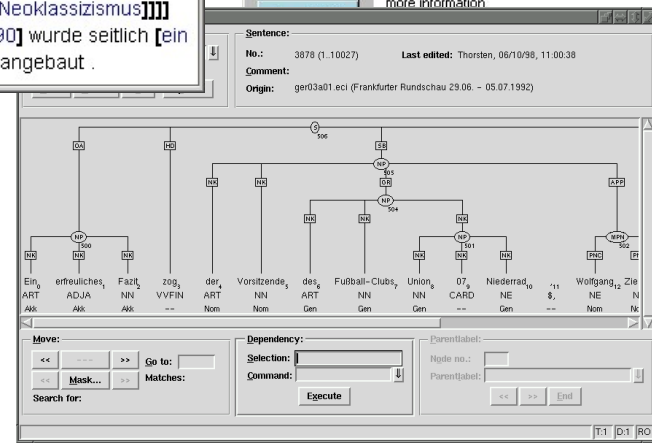
```
<de deID="de8" deType="namedEntity" headRef="w36">
  <token ref="w36">Maik</token>
</de>
<token ref="w37">hat</token> <token ref="w38">kein</token>
<token ref="w39">eigenes</token> <token ref="w40">Fahrrad</token>,
<token ref="w42">und</token>
<de deID="de10" deType="namedEntity" headRef="w43">
  <token ref="w43">Marie</token>
</de>
<token ref="w45">fährt</token> <token ref="w46">nicht</token>
<token ref="w47">in</token>
<de deID="de11" deType="nom" headRef="w49">
  <token ref="w48">den</token>
  <token ref="w49">Urlaub</token>
</de>
<de deID="de12" deType="nom" headRef="w53">
  <token ref="w52">Zwei</token>
  <token ref="w53">Kinder</token>
</de>,
<de deID="de13" deType="nom" headRef="w56">
  <token ref="w55">eine</token>
  <token ref="w56">Gemeinsamkeit</token>
</de>.
<token_ref id="w36" dependHead="w37" pos="N" syntax="@NH" dependValue="subj"
  morpho="MSC SG NOM" text="Maik" lemma="maik" />
<semRel>
  <bridgingLink relType="hasMember" antecedentIDRefs="de8 de10" phorIDRef="de12"/>
</semRel>
```

- Projekt *Sekimo*
- Annotationschema zur Annotation semantischer Relationen
- **Annotationswerkzeuge**
- Serengeti
- Zusammenfassung und Ausblick

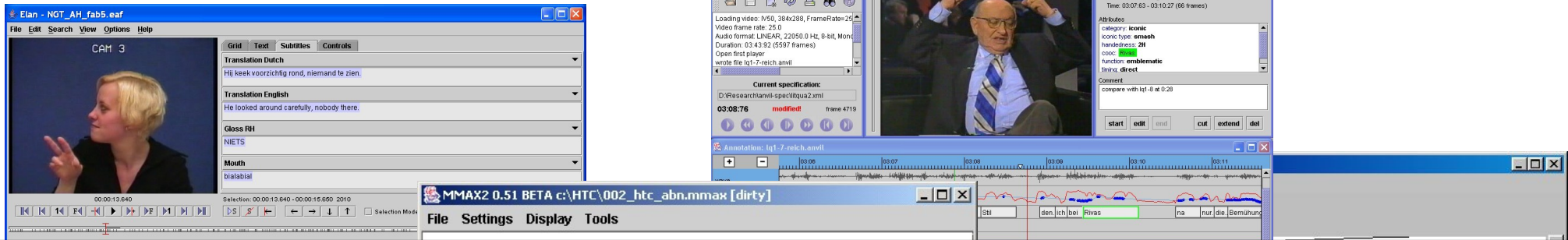
Es gibt bereits eine Reihe von Werkzeugen zur Annotation unimodaler und multimodaler Dokumente:



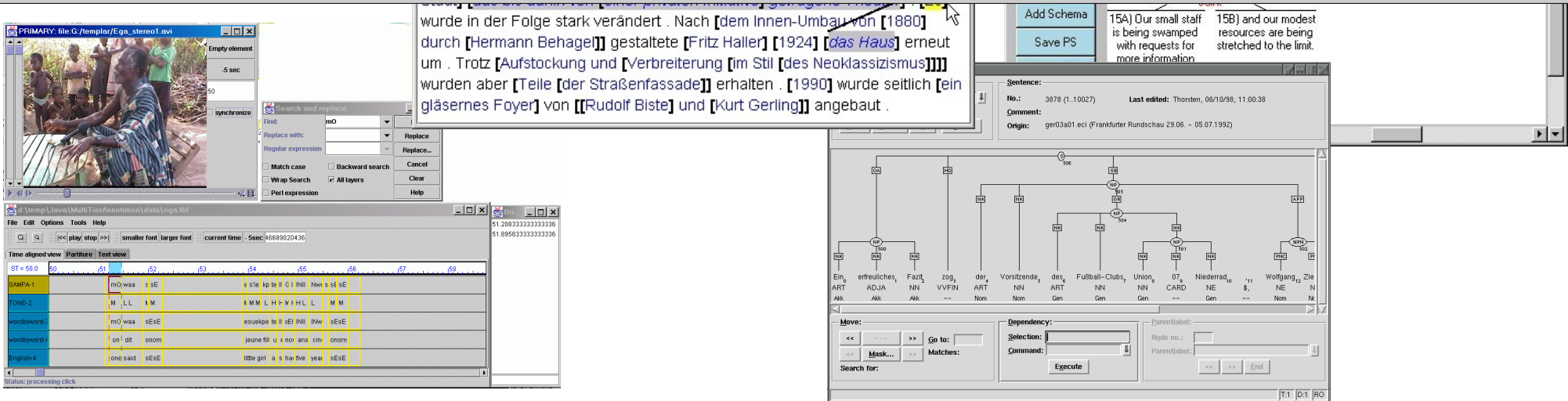
Time	ST	50.0	51	52	53	54	55	56	57	58
SAMPA-1			mCv	vaa	t	sE				
tone-2			M	LL	YM					
word1			mCv	vaa	sE					
word2			on	dt	onom					
English-4			one	said	sE					



Es gibt bereits eine Reihe von Werkzeugen zur Annotation unimodaler und multimodaler Dokumente:



## Warum ein weiteres Annotationswerkzeug?



Nachteil einer Reihe von existierenden Werkzeugen:

- lokale Datenhaltung

Anforderung an ein neues Annotationswerkzeug:

- zentrale Datenhaltung
- Unterstützung einer verteilten Korpuserstellung

- Projekt *Sekimo*
- Annotationschema zur Annotation semantischer Relationen
- Annotationswerkzeuge
- **Serengeti**
- Zusammenfassung und Ausblick

- Webbasierte Client-Server-Applikation (aktuelle Version 0.7 implementiert für Mozilla Firefox > 1.0)
- Ajax-Architektur (Perl, JavaScript, XHTML, CSS)
- Vorteile:
  - Zentrale Verwaltung eines Korpus
  - parallele Annotation eines Dokuments durch mehrere Benutzer
  - Operationen über mehrere Annotationen (Vergleich und Merge)
  - keine Installation notwendig



The screenshot shows a Mozilla Firefox browser window titled "Serengeti Annotator - Mozilla Firefox". The address bar contains the URL "http://coli.lili.uni-bielefeld.de/serengeti/annotator.pl". The search bar contains the text "ndiewald". The main content area features a teal background with the title "Serengeti Annotator V0.7.0" and the subtitle "Semantic Relations Annotation Tool". Below the title, there is a paragraph describing the tool as a web-based client-server application for annotating semantic relations in text documents. It mentions that the tool is developed in the SEKIMD (A2) project, part of the distributed research group "TEXT TECHNOLOGICAL MODELLING OF INFORMATION" founded by the DFG. A note indicates it is optimized for Mozilla Firefox V1.x. On the right side, there are three links: "A BRIEF STARTING GUIDE", "DEMO ACCOUNT", and "CONTACT". At the bottom right, there is a zebra illustration and the word "serengeti" in a stylized font. The footer of the page states "Last Update: 2007-01-19 by Nils Diewald". The browser status bar at the bottom shows the word "Fertig" and two green checkmark icons.

Serengeti Annotator - Mozilla Firefox

http://coli.lili.uni-bielefeld.de/serengeti/annotator.pl

ndiewald

## Serengeti Annotator V0.7.0

### Semantic Relations Annotation Tool

The Serengeti Annotator is a web based client-server-application used for annotating semantic relations in text documents.

It is developed in the SEKIMD (A2) project, part of the distributed research group "TEXT TECHNOLOGICAL MODELLING OF INFORMATION" founded by the DFG.

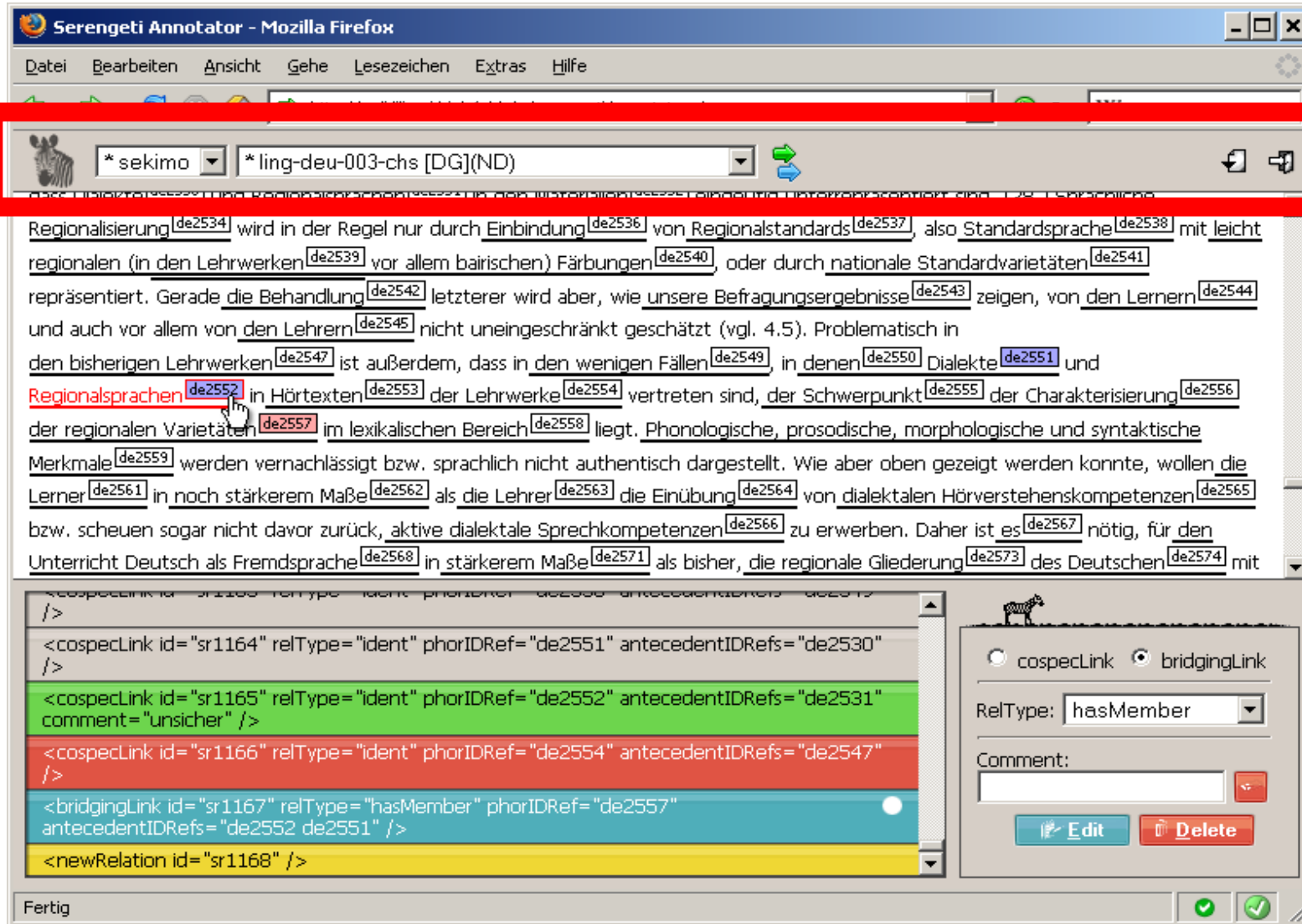
Optimized for Mozilla Firefox V1.x

- A BRIEF STARTING GUIDE
- DEMO ACCOUNT
- CONTACT

serengeti

Last Update: 2007-01-19 by Nils Diewald

Fertig



Serengeti Annotator - Mozilla Firefox

Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe

\* sekimo \* ling-deu-003-chs [DG](ND)

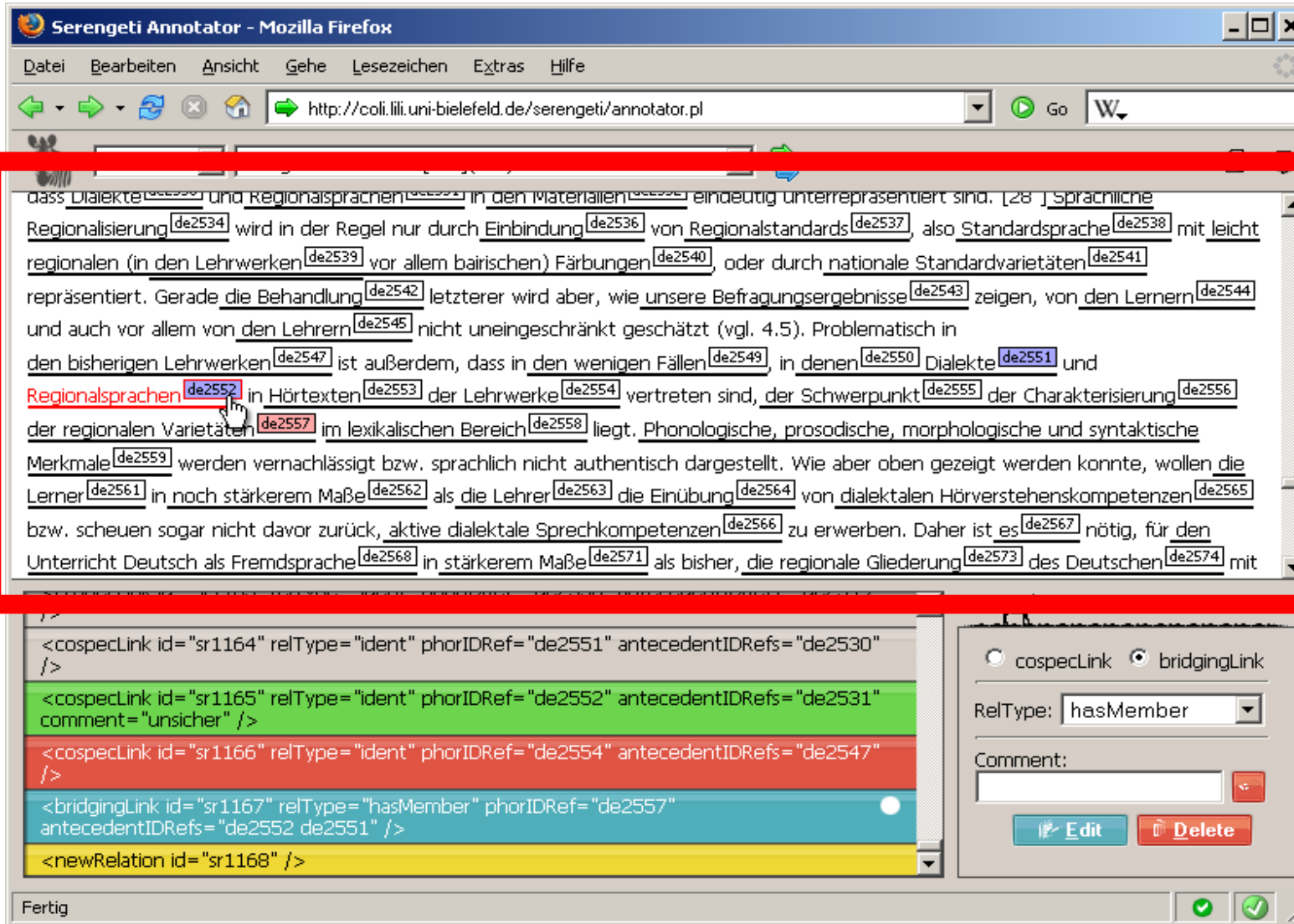
Regionalisierung<sup>[de2534]</sup> wird in der Regel nur durch Einbindung<sup>[de2536]</sup> von Regionalstandards<sup>[de2537]</sup>, also Standardsprache<sup>[de2538]</sup> mit leicht regionalen (in den Lehrwerken<sup>[de2539]</sup> vor allem bairischen) Färbungen<sup>[de2540]</sup>, oder durch nationale Standardvarietäten<sup>[de2541]</sup> repräsentiert. Gerade die Behandlung<sup>[de2542]</sup> letzterer wird aber, wie unsere Befragungsergebnisse<sup>[de2543]</sup> zeigen, von den Lernern<sup>[de2544]</sup> und auch vor allem von den Lehrern<sup>[de2545]</sup> nicht uneingeschränkt geschätzt (vgl. 4.5). Problematisch in den bisherigen Lehrwerken<sup>[de2547]</sup> ist außerdem, dass in den wenigen Fällen<sup>[de2549]</sup> in denen<sup>[de2550]</sup> Dialekte<sup>[de2551]</sup> und Regionalsprachen<sup>[de2552]</sup> in Hörtexten<sup>[de2553]</sup> der Lehrwerke<sup>[de2554]</sup> vertreten sind, der Schwerpunkt<sup>[de2555]</sup> der Charakterisierung<sup>[de2556]</sup> der regionalen Varietäten<sup>[de2557]</sup> im lexikalischen Bereich<sup>[de2558]</sup> liegt. Phonologische, prosodische, morphologische und syntaktische Merkmale<sup>[de2559]</sup> werden vernachlässigt bzw. sprachlich nicht authentisch dargestellt. Wie aber oben gezeigt werden konnte, wollen die Lerner<sup>[de2561]</sup> in noch stärkerem Maße<sup>[de2562]</sup> als die Lehrer<sup>[de2563]</sup> die Einübung<sup>[de2564]</sup> von dialektalen Hörverstehenskompetenzen<sup>[de2565]</sup> bzw. scheuen sogar nicht davor zurück, aktive dialektale Sprechkompetenzen<sup>[de2566]</sup> zu erwerben. Daher ist es<sup>[de2567]</sup> nötig, für den Unterricht Deutsch als Fremdsprache<sup>[de2568]</sup> in stärkerem Maße<sup>[de2571]</sup> als bisher, die regionale Gliederung<sup>[de2573]</sup> des Deutschen<sup>[de2574]</sup> mit

```
<cospecLink id="sr1164" relType="ident" phorIDRef="de2551" antecedentIDRefs="de2530" />
<cospecLink id="sr1165" relType="ident" phorIDRef="de2552" antecedentIDRefs="de2531" comment="unsicher" />
<cospecLink id="sr1166" relType="ident" phorIDRef="de2554" antecedentIDRefs="de2547" />
<bridgingLink id="sr1167" relType="hasMember" phorIDRef="de2557" antecedentIDRefs="de2552 de2551" />
<newRelation id="sr1168" />
```

Fertig

- Auswahl der Gruppe
- Auswahl des Dokuments
- Funktionen für
  - Speicherung
  - Export
  - Commit
  - Print
  - Logout

# Serengeti – Semantic Relations Annotation Tool



Serengeti Annotator - Mozilla Firefox

http://coli.lili.uni-bielefeld.de/serengeti/annotator.pl

dass Dialekte und Regionalsprachen in den Materialien eindeutig unterrepräsentiert sind. [28] Sprachliche Regionalisierung wird in der Regel nur durch Einbindung von Regionalstandards, also Standardsprache mit leicht regionalen (in den Lehrwerken vor allem bairischen) Färbungen, oder durch nationale Standardvarietäten repräsentiert. Gerade die Behandlung letzterer wird aber, wie unsere Befragungsergebnisse zeigen, von den Lernern und auch vor allem von den Lehrern nicht uneingeschränkt geschätzt (vgl. 4.5). Problematisch in den bisherigen Lehrwerken ist außerdem, dass in den wenigen Fällen in denen Dialekte und Regionalsprachen in Hörtexten der Lehrwerke vertreten sind, der Schwerpunkt der Charakterisierung der regionalen Varietäten im lexikalischen Bereich liegt. Phonologische, prosodische, morphologische und syntaktische Merkmale werden vernachlässigt bzw. sprachlich nicht authentisch dargestellt. Wie aber oben gezeigt werden konnte, wollen die Lerner in noch stärkerem Maße als die Lehrer die Einübung von dialektalen Hörverstehenskompetenzen bzw. scheuen sogar nicht davor zurück, aktive dialektale Sprechkompetenzen zu erwerben. Daher ist es nötig, für den Unterricht Deutsch als Fremdsprache in stärkerem Maße als bisher, die regionale Gliederung des Deutschen mit

```
<cospecLink id="sr1164" relType="ident" phorIDRef="de2551" antecedentIDRefs="de2530" />
<cospecLink id="sr1165" relType="ident" phorIDRef="de2552" antecedentIDRefs="de2531" comment="unsicher" />
<cospecLink id="sr1166" relType="ident" phorIDRef="de2554" antecedentIDRefs="de2547" />
<bridgingLink id="sr1167" relType="hasMember" phorIDRef="de2557" antecedentIDRefs="de2552 de2551" />
<newRelation id="sr1168" />
```

Fertig

cospecLink bridgingLink

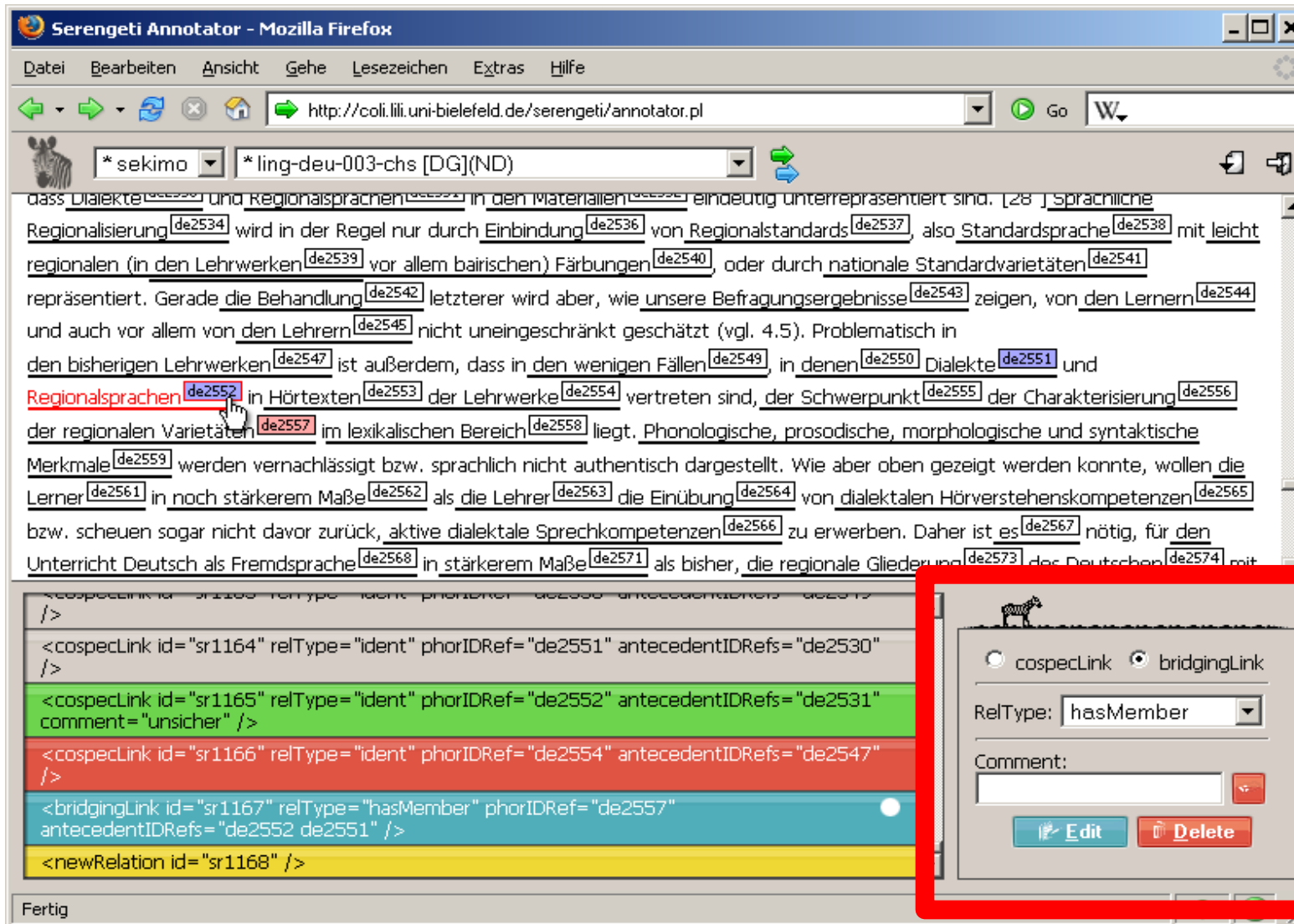
RelType: hasMember

Comment:

Edit Delete

- Anzeige des Dokuments
- Layout
- Auswahl der Diskursentitäten

# Serengeti – Semantic Relations Annotation Tool



Serengeti Annotator - Mozilla Firefox

http://coli.lili.uni-bielefeld.de/serengeti/annotator.pl

\* sekimo \* ling-deu-003-chs [DG](ND)

... dass Dialekte und Regionalsprachen in den Materialien eindeutig unterrepräsentiert sind. [28] Sprachliche Regionalisierung wird in der Regel nur durch Einbindung von Regionalstandards, also Standardsprache mit leicht regionalen (in den Lehrwerken vor allem bairischen) Färbungen, oder durch nationale Standardvarietäten repräsentiert. Gerade die Behandlung letzterer wird aber, wie unsere Befragungsergebnisse zeigen, von den Lernern und auch vor allem von den Lehrern nicht uneingeschränkt geschätzt (vgl. 4.5). Problematisch in den bisherigen Lehrwerken ist außerdem, dass in den wenigen Fällen in denen Dialekte und Regionalsprachen in Hörtexten der Lehrwerke vertreten sind, der Schwerpunkt der Charakterisierung der regionalen Varietäten im lexikalischen Bereich liegt. Phonologische, prosodische, morphologische und syntaktische Merkmale werden vernachlässigt bzw. sprachlich nicht authentisch dargestellt. Wie aber oben gezeigt werden konnte, wollen die Lerner in noch stärkerem Maße als die Lehrer die Einübung von dialektalen Hörverstehenskompetenzen bzw. scheuen sogar nicht davor zurück, aktive dialektale Sprechkompetenzen zu erwerben. Daher ist es nötig, für den Unterricht Deutsch als Fremdsprache in stärkerem Maße als bisher, die regionale Gliederung des Deutschen mit

```
<cospecLink id="sr1164" relType="ident" phorIDRef="de2551" antecedentIDRefs="de2530" />
<cospecLink id="sr1165" relType="ident" phorIDRef="de2552" antecedentIDRefs="de2531" comment="unsicher" />
<cospecLink id="sr1166" relType="ident" phorIDRef="de2554" antecedentIDRefs="de2547" />
<bridgingLink id="sr1167" relType="hasMember" phorIDRef="de2557" antecedentIDRefs="de2552 de2551" />
<newRelation id="sr1168" />
```

Fertig

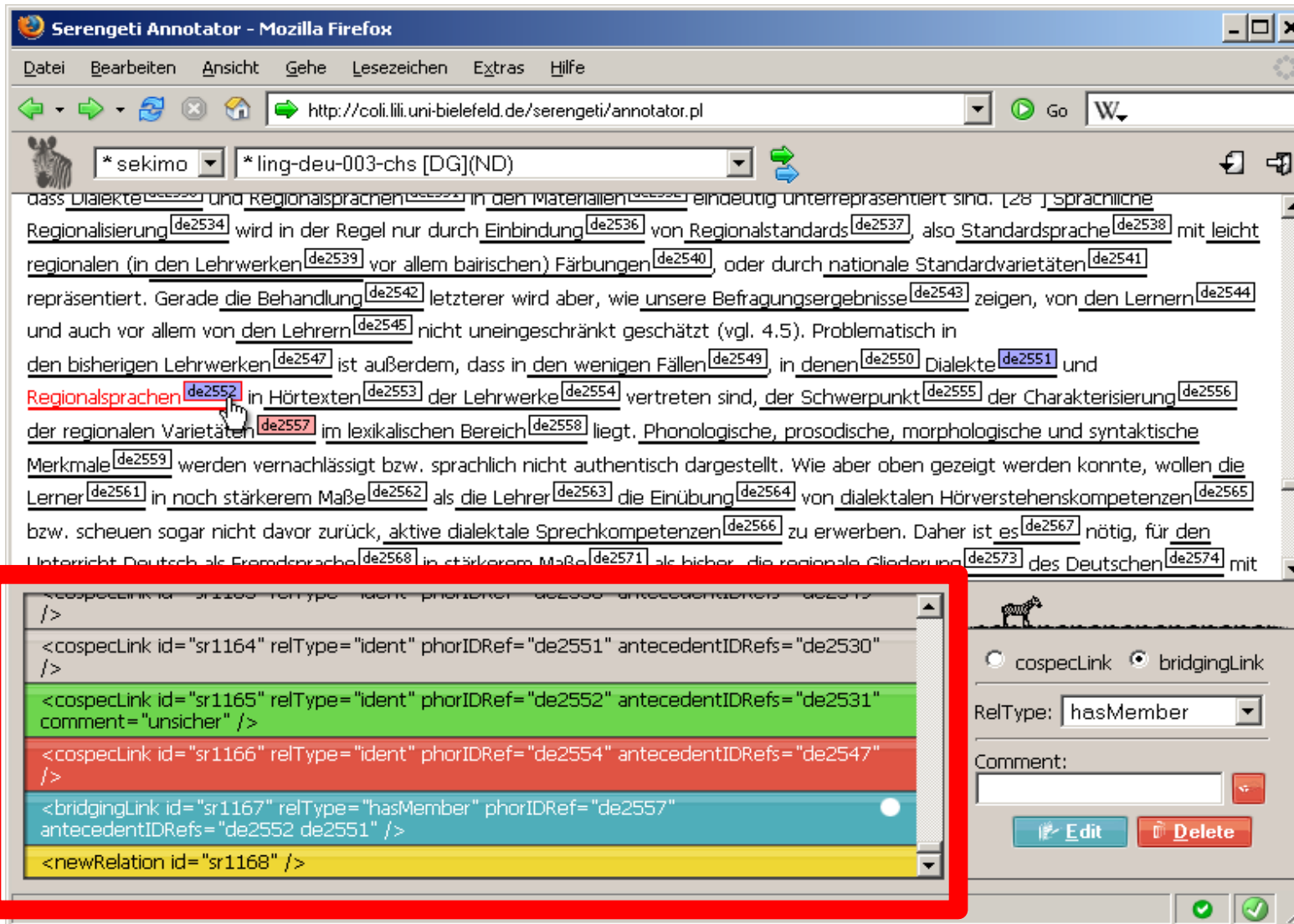
RelType: hasMember

Comment:

Edit Delete

- Spezifizierung des Relationstyps
- Editieren und Löschen
- Status-Anzeige

# Serengeti – Semantic Relations Annotation Tool



Serengeti Annotator - Mozilla Firefox

http://coli.lili.uni-bielefeld.de/serengeti/annotator.pl

\* sekimo \* ling-deu-003-chs [DG](ND)

class Dialekte und Regionalsprachen in den Materialien eindeutig unterrepräsentiert sind. [28] Sprachliche Regionalisierung wird in der Regel nur durch Einbindung von Regionalstandards, also Standardsprache mit leicht regionalen (in den Lehrwerken vor allem bairischen) Färbungen, oder durch nationale Standardvarietäten repräsentiert. Gerade die Behandlung letzterer wird aber, wie unsere Befragungsergebnisse zeigen, von den Lernern und auch vor allem von den Lehrern nicht uneingeschränkt geschätzt (vgl. 4.5). Problematisch in den bisherigen Lehrwerken ist außerdem, dass in den wenigen Fällen in denen Dialekte und Regionalsprachen in Hörtexten der Lehrwerke vertreten sind, der Schwerpunkt der Charakterisierung der regionalen Varietäten im lexikalischen Bereich liegt. Phonologische, prosodische, morphologische und syntaktische Merkmale werden vernachlässigt bzw. sprachlich nicht authentisch dargestellt. Wie aber oben gezeigt werden konnte, wollen die Lerner in noch stärkerem Maße als die Lehrer die Einübung von dialektalen Hörverstehenskompetenzen bzw. scheuen sogar nicht davor zurück, aktive dialektale Sprechkompetenzen zu erwerben. Daher ist es nötig, für den Unterricht Deutsch als Fremdsprache in stärkerem Maße als bisher, die regionale Gliederung des Deutschen mit

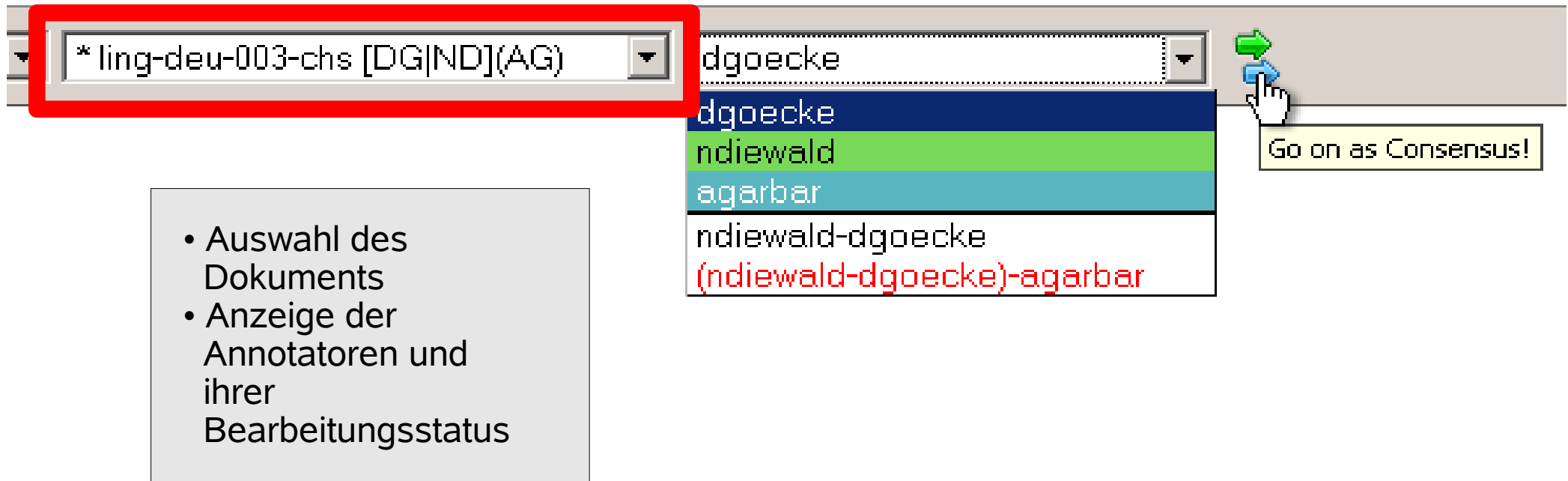
```
<cospecLink id="sr1164" relType="ident" phorIDRef="de2551" antecedentIDRefs="de2530" />
<cospecLink id="sr1165" relType="ident" phorIDRef="de2552" antecedentIDRefs="de2531" comment="unsicher" />
<cospecLink id="sr1166" relType="ident" phorIDRef="de2554" antecedentIDRefs="de2547" />
<bridgingLink id="sr1167" relType="hasMember" phorIDRef="de2557" antecedentIDRefs="de2552 de2551" />
<newRelation id="sr1168" />
```

RelType: hasMember

Comment:

Edit Delete

- Anzeige und Auswahl erstellter Relationen
- Farbliche Hervorhebungen



\* ling-deu-003-chs [DG|ND](AG)

dgoecke

dgoecke

ndiewald

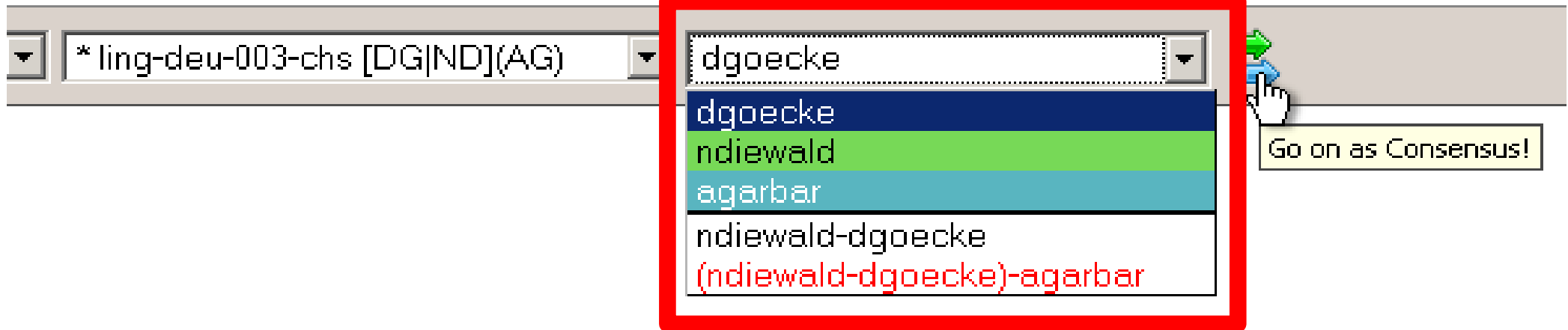
agarbar

ndiewald-dgoecke

(ndiewald-dgoecke)-agarbar

Go on as Consensus!

- Auswahl des Dokuments
- Anzeige der Annotatoren und ihrer Bearbeitungsstatus



\* ling-deu-003-chs [DG|ND](AG)

dgoecke

dgoecke

ndiewald

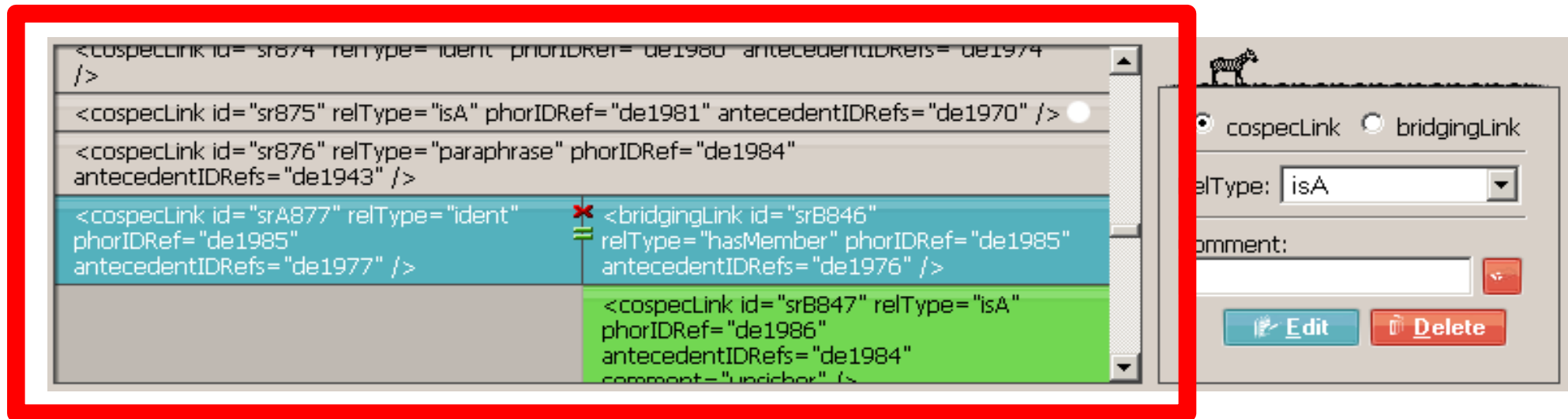
agarbar

ndiewald-dgoecke

(ndiewald-dgoecke)-agarbar

Go on as Consensus!

- Auswahl der Annotationen
- Anzeige der Annotatoren und ihrer Bearbeitungsstatus



```
<cospecLink id="sr874" relType="ident" phorIDRef="de1980" antecedentIDRefs="de1974" />
<cospecLink id="sr875" relType="isA" phorIDRef="de1981" antecedentIDRefs="de1970" />
<cospecLink id="sr876" relType="paraphrase" phorIDRef="de1984" antecedentIDRefs="de1943" />
<cospecLink id="srA877" relType="ident" phorIDRef="de1985" antecedentIDRefs="de1977" />
<bridgingLink id="srB846" relType="hasMember" phorIDRef="de1985" antecedentIDRefs="de1976" />
<cospecLink id="srB847" relType="isA" phorIDRef="de1986" antecedentIDRefs="de1984" comment="unsicher" />
```

- Anzeige der Relationen beider Annotationen
- Konfliktrelationen werden gegenübergestellt
- Konfliktart wird dargestellt

- Projekt *Sekimo*
- Annotationschema zur Annotation semantischer Relationen
- Annotationswerkzeuge
- Serengeti
- **Zusammenfassung und Ausblick**

- Serengeti bietet die Möglichkeit große Datenmengen in hoher Qualität zu annotieren
- zentrale Verwaltung eines Korpus
- parallele Annotation eines Dokuments durch mehrere Benutzer
- Operationen über mehrere Annotationen
- keine Installation notwendig
- modular konzipiertes System

- Generalisierung des Annotationswerkzeuges für die Annotation weiterer Diskursrelationen, z.B. lexikalischer Ketten
- Externe Definition von *Markables* und *Relationen*, aktuelle Anwendungsfälle:
  - Markable=DE; Relation=cospecLink/bridgingLink
  - Markable=Lemma; Relation=lexicalChain
- Anwendung in der DFG-Forschergruppe:
  - Korpusstudien
  - Annotationsschichten als Ressource für die Diskursanalyse
- Erweiterung um Berechnung der Inter-Annotator-Reliabilität (z. B.  $\kappa$ ,  $AC_1$ )
- Abfragetool zur Suche in Annotationen

Vielen Dank für Ihre  
Aufmerksamkeit!

serengeti

