

# Einführung in die Benutzung von ANNIS 1.1 – DRAFT –

SFB 632, D1  
annis@ling.uni-potsdam.de

27. August 2004

## **Zusammenfassung**

Dies ist eine praxisorientierte Einführung in die Benutzung von ANNIS. ANNIS bietet einen einfachen Zugang zu sprachlichen Daten und ihren Annotationen über das Internet. Es ermöglicht sowohl die gleichzeitige Betrachtung von sehr verschiedenartiger Annotation auf mehreren Ebenen als auch die Suche in diesen Daten.

Neben der Benutzeroberfläche von ANNIS behandelt diese Einführung vor allem die Navigation in den Korpusdaten, die Visualisierung der Annotation und die Suche.

# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>3</b>
<b>2</b>	<b>Benutzeroberfläche</b>	<b>3</b>
2.1	Die Menüleiste . . . . .	3
2.2	Die Arbeitsoberfläche . . . . .	4
2.3	Konventionen . . . . .	4
<b>3</b>	<b>Navigation in den Korpusdaten</b>	<b>5</b>
3.1	Die Annotationsebenen der Korpusdaten . . . . .	5
3.2	Die Annotationsdokumente einer Annotationsebene . . . . .	5
<b>4</b>	<b>Zur Darstellung der annotierten Daten</b>	<b>7</b>
4.1	Detailansicht . . . . .	7
4.2	Diskursansicht . . . . .	9
4.3	Navigationsleiste . . . . .	10
4.4	Zeigerrelationen . . . . .	10
4.5	Zur schichtenbasierten Darstellung von Baumstrukturen . . . . .	11
<b>5</b>	<b>Die Suche</b>	<b>12</b>
5.1	Einfache Suche . . . . .	13
5.1.1	Suche nach Text . . . . .	13
5.1.2	Suche nach Annotationen . . . . .	15
5.2	Kombinierte Suche . . . . .	17
5.3	Speichern und Laden ausgewählter Suchergebnisse . . . . .	19
<b>6</b>	<b>Datenimport und -export</b>	<b>20</b>
6.1	Datenimport . . . . .	20
6.2	Datenexport . . . . .	20
<b>7</b>	<b>Verschiedenes</b>	<b>21</b>
7.1	Benutzereinstellungen . . . . .	21
7.2	Benutzerrechte . . . . .	21
<b>8</b>	<b>Zusammenfassung und Ausblick</b>	<b>21</b>

# 1 Einführung

ANNIS steht für **A**nnotierte **I**nformations**S**truktur und ist der Arbeitstitel für das Programm, das den Zugriff auf die linguistischen Daten im SFB erlaubt. Dies beinhaltet neben der Visualisierung der annotierten Daten auch eine Suche.

Diese Einführung soll die Nutzer in die Lage versetzen, die wichtigsten Funktionen von ANNIS anzuwenden. Neben der Benutzeroberfläche (Abschnitt 2) wird die Navigation in den Korpusdaten (Abschnitt 3), die Visualisierung der Annotation (Abschnitt 4) und die Suche (Abschnitt 5) behandelt. In Abschnitt 6 geht es um den Import und Export von Daten, in Abschnitt 7 um die Benutzereinstellungen und die Benutzerrechte.

Am Ende vieler Abschnitte werden kleine praktische Aufgaben angeboten, die auf den vorherigen Ausführungen aufbauen. Diese beziehen sich auf ein kleines Beispielkorpus von zehn aus dem Potsdamer Kommentarkorpus entnommenen Texten – dem *PCC10 (Potsdam Commentary Corpus 10)* – die auf mehreren linguistischen Ebenen annotiert wurden.

## 2 Benutzeroberfläche

Die Benutzeroberfläche besteht aus der dauerhaft sichtbaren Menüleiste auf der linken und der Arbeitsfläche auf der rechten Seite.<sup>1</sup>

### 2.1 Die Menüleiste

Die Menüleiste besteht, von oben beginnend, aus den folgenden Komponenten:

- dem Logo mit Links zum Sonderforschungsbereich und den beiden beteiligten Universitäten.
- einem Icon mit einer Landesflagge für das Umschalten zwischen verschiedenen Sprachen, derzeit Englisch und Deutsch.
- einem Feld *'Daten'*, das den Zugriff auf die Korpusdaten und Datenformate erlaubt.
- einem Feld *'Suche'* mit einem Suchfenster für die Eingabe von Suchanfragen und dem Trefferspeicher.
- einem Feld *'Benutzer'* mit Benutzeroptionen wie An- und Abmeldung und den Benutzereinstellungen.

---

<sup>1</sup>Da ANNIS aus einem Webbrowser heraus aufgerufen wird, stehen dessen Funktionalitäten ebenfalls zur Verfügung. Dies beinhaltet vor allem die Navigationsfunktionen wie Vorwärts- oder Zurückblättern, Links, Druckfunktionen, usw. ...

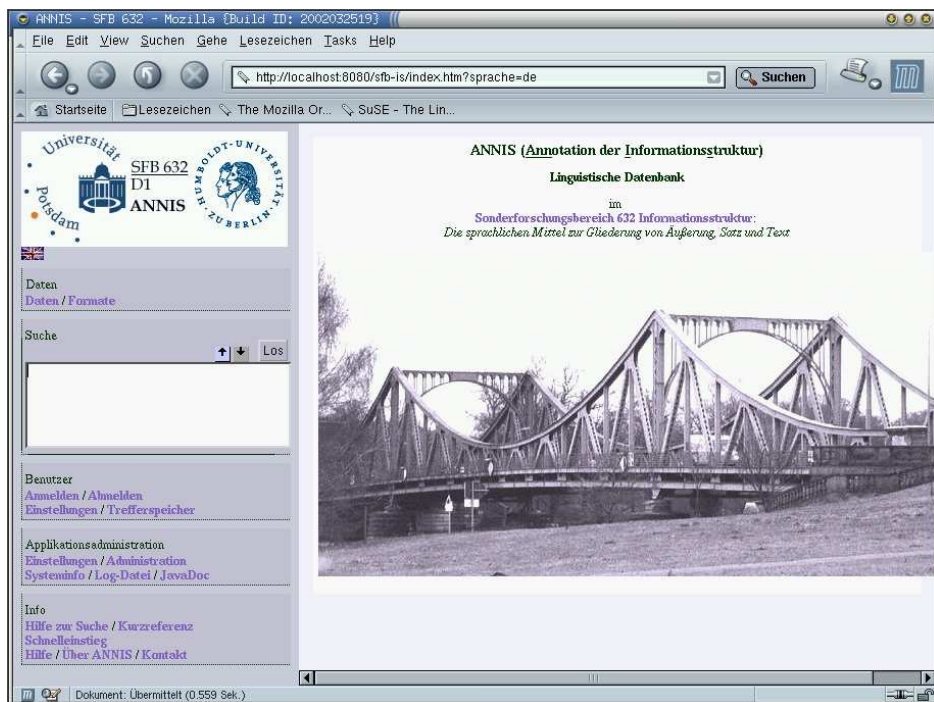


Abbildung 1: Die Benutzeroberfläche

- einem Feld 'Info' mit diesem Tutorial, einem Glossar, Hilfen zur Suche und Hintergrundinformationen zu ANNIS.

## Aufgaben

- Testen Sie die Links in der Menüleiste!

## 2.2 Die Arbeitsoberfläche

Die Arbeitsoberfläche ist der dynamische Teil der Benutzeroberfläche. Sie dient u.a. zur Navigation in den Korpusdaten, dem Betrachten der annotierten Daten und dem Betrachten von Suchergebnissen.

## 2.3 Konventionen

Um die Benutzerführung in ANNIS zu erleichtern wurde versucht, einige einfache Regeln zu befolgen:

- Weiterführende Links werden **violett** (in diesem Dokument: **fettgedruckt**) hervorgehoben.

- Dreiecksschalter ermöglichen das Ein- und Ausblenden zusätzlicher Informationen und stehen rechts von ihren Bezugsbegriffen.
- Ein längeres Verweilen des Mauszeigers über einem Ausdruck lässt häufig eine Erläuterung zu diesem erscheinen.

### 3 Navigation in den Korpusdaten

In diesem Abschnitt geht es um den Umgang mit den Korpusdaten in ANNIS.

#### 3.1 Die Annotationsebenen der Korpusdaten

Ein Klick auf den Link **Daten** im Abschnitt '*Daten*' in der Menüleiste führt zur Übersicht über die Annotationsebenen in ANNIS. Diese werden wie in Abbildung 2 aufgelistet: für jede Annotationsebene wird die Anzahl der annotierten Dokumente angegeben. So sind z.B. 7 Dokumente<sup>2</sup> auf der Annotationsebene *pcc10.co-reference* annotiert.

Zusätzlich können Informationen zum Annotationsschema einer Ebene – derzeit eine Aufzählung der Tags des Schemas – mit Hilfe der Dreiecksschalter ein- und ausgeblendet werden. Im Beispiel sind die Tags der Annotationsebene *pcc10.syntax-tiger* eingeblendet (d.h. 'cat', 'char', 'rel', usw.).

Ein Klick auf die Angabe der Zahl der Dokumente führt zur Ansicht der Dokumente dieser Annotationsebene.

#### Aufgaben

- Klicken Sie auf den Dreiecksschalter hinter *pcc10.cognitive-status*, um die Tags dieser Annotationsebene einzublenden. Ein nochmaliger Klick blendet sie wieder aus.
- Klicken Sie auf den Link **10 Dokumente** der Annotationsebene *pcc10.is.cognitive-status*, um zur Ansicht der Dokumente dieser Annotationsebene zu gelangen.

#### 3.2 Die Annotationsdokumente einer Annotationsebene

Die Annotationsdokumente einer Annotationsebene werden zusammen mit einem Ausschnitt der Primärdaten aufgelistet. Im Beispiel in Abbildung 3 sehen wir die Dokumente der Annotationsebene *pcc10.is.cognitive-status*, d. h. die Dokumente, in denen Kognitiver Status annotiert wurde. Eines der

---

<sup>2</sup>Die Anzahl und die Namen der Dokumente in dieser Einführung kann vom jeweils aktuellen Datensatz in ANNIS abweichen, da dieser noch vervollständigt wird.

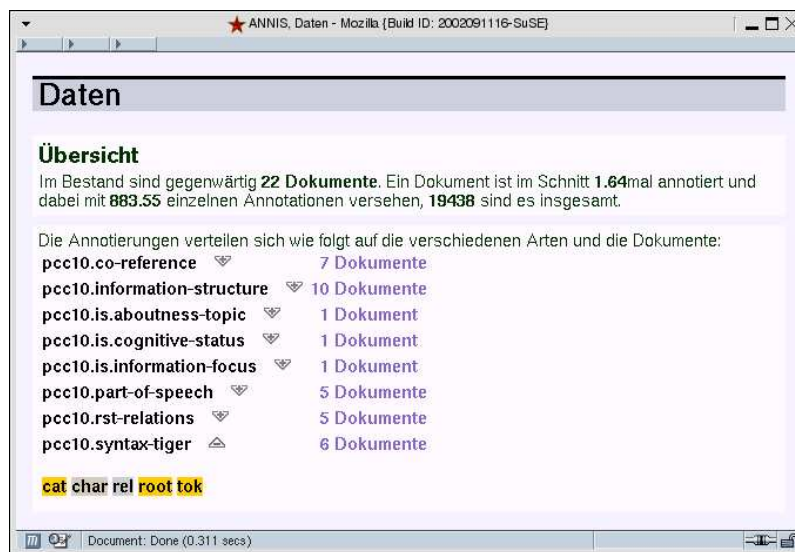


Abbildung 2: Übersicht über die Korpusdaten in ANNIS

Dokumente hat den Namen *maz-alles-16153* (bzw. *maz16153\_anno*) und beginnt mit den Worten “Alles offen In Berlin”.

Zu jedem Dokument können mit Hilfe eines Dreiecksschalters weitere Informationen eingeblendet werden. So wurden zum Annotationsdokument **maz-keine-10205** Informationen darüber eingeblendet, auf welchen Annotationsebenen das Primärdatum annotiert wurde – in unserem Beispiel zusätzlich mit *pcc10.is.aboutness-topic* und *pcc10.is.information-focus*. Ausserdem ist für jedes Annotationsdokument eine Liste mit Exportoptionen eingeblendet.<sup>3</sup>

Ein Klick auf den Dokumentnamen führt zu der Annotationsansicht eines Dokuments.

### Aufgaben

- Klicken Sie auf den Dreiecksschalter hinter dem Dokumentnamen **maz-5715\_anno**, um weitere Informationen zu den Annotationen des Primärdatums einzublenden.
- Klicken sie auf den Dokumentnamen **maz5715**. Dies führt zu dessen Annotationsansicht, die im nächsten Abschnitt näher erläutert wird.

<sup>3</sup>Siehe dazu Abschnitt 6.

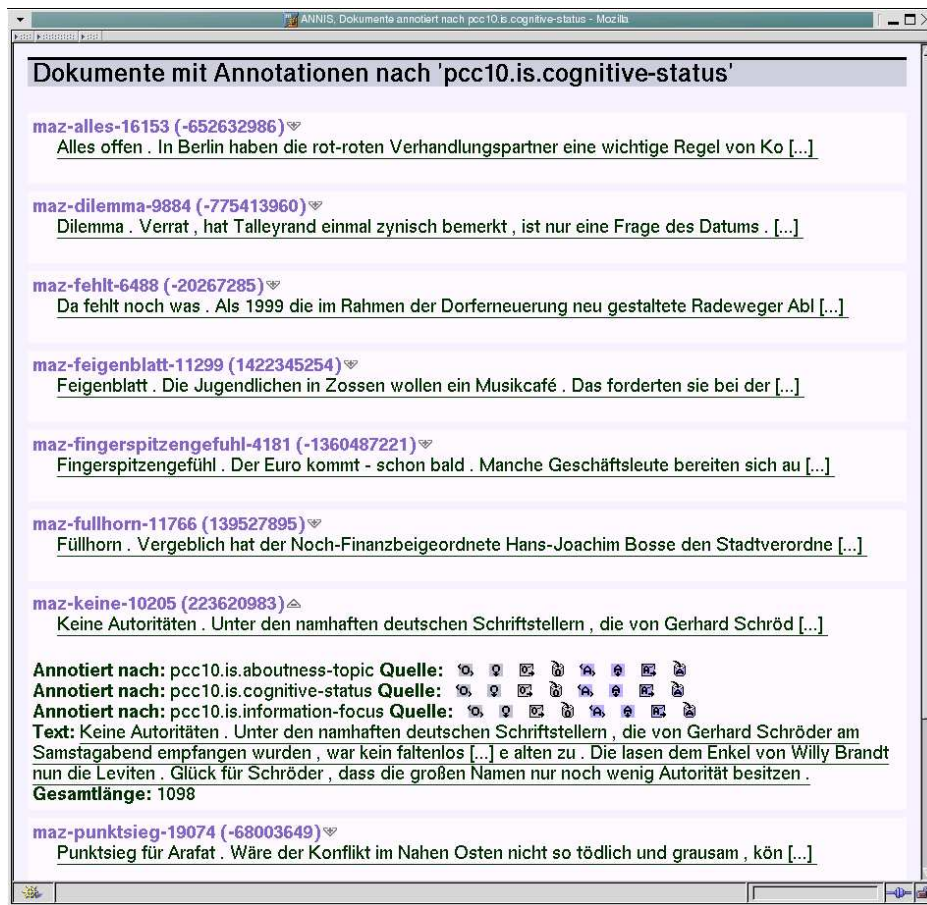


Abbildung 3: Übersicht über die Dokumente einer Annotationsebene

## 4 Zur Darstellung der annotierten Daten

Die Annotationsansicht besteht aus drei Komponenten, der *Detailansicht*, der *Diskursansicht* und der *Navigationsleiste*.

### 4.1 Detailansicht

Die Detailansicht besteht aus i) der Primärdatenzeile mit einem Ausschnitt des Primärdatums und ii) den Annotationen auf den einzelnen Annotationsebenen darüber.

Die Primärdatenzeile, die im Beispiel in Abbildung 4 am unteren Rand zu sehen ist, enthält eine tokenisierte Repräsentation des gegenwärtig ausgewählten Ausschnitts der Primärdaten. Die Annotationen darüber sind entsprechend ihrer Zugehörigkeit zu den einzelnen Annotationsebenen angeordnet und können mit Hilfe der Dreiecksschalter ein- und ausgeblendet wer-

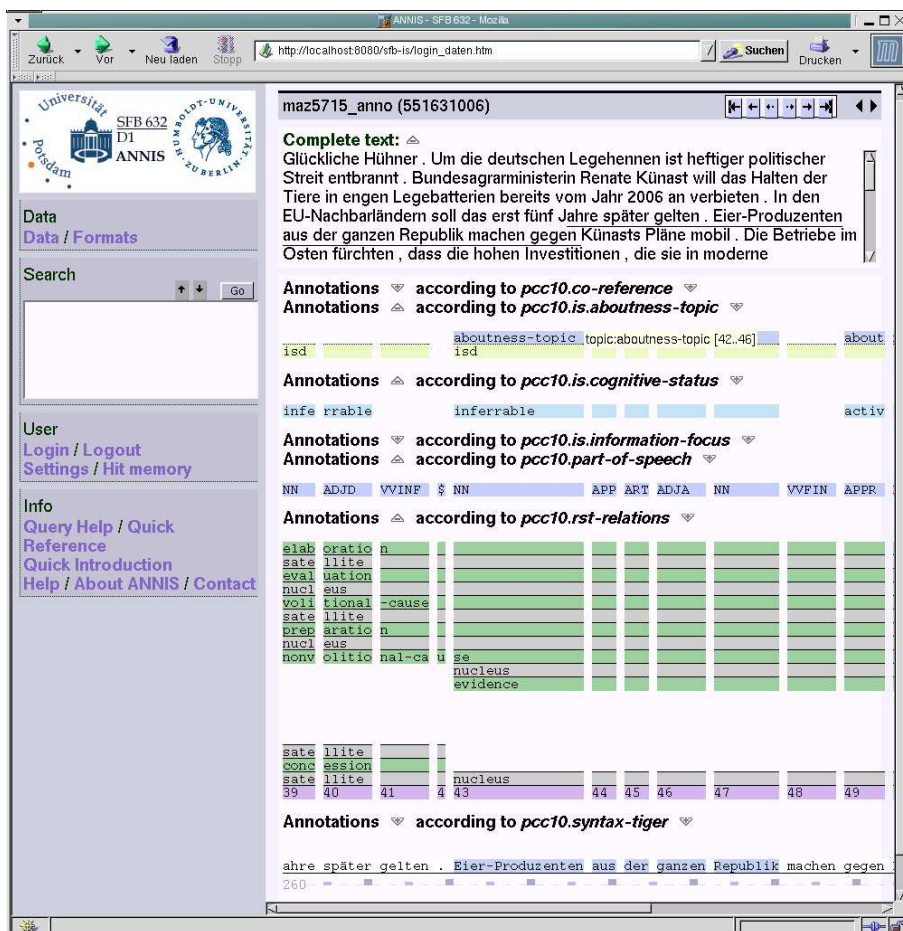


Abbildung 4: Annotationsansicht

den. So sind in Abbildung 4 die Annotationen der Ebene *pcc10.is.aboutness-topic*, *pcc10.is.cognitive-status*, *pcc10.part-of-speech* und *pcc10.rst-relations* ein-, und die der Annotationsebenen *pcc10.co-reference*, *pcc10.information-focus* und *pcc10.syntax-tiger* ausgeblendet.

**Annotation und Primärdaten** Annotationen auf den einzelnen Annotationsebenen beziehen sich in der Regel direkt auf das Segment der Primärdaten auf der Primärdatenzeile, welches sie überspannen. Die mit der Annotation verbundenen Primärdaten werden hervorgehoben, wenn sich der Mauszeiger über einer Annotation befindet. Im Beispiel in Abbildung 4 befindet sich der Mauszeiger über einer Annotation auf der Annotationsebene *pcc10.is.aboutness-topic*, über einem *aboutness-topic* (im oberen mittleren Teil des Bildschirms). Das zugehörige Segment *Eier-Produzenten aus der ganzen Republik* in der Primärdatenzeile wird dabei farblich markiert. Wenn



der Mauszeiger für längere Zeit über einer Annotation verweilt, erscheinen zusätzliche Informationen, in Abbildung 4 über das Attribut-Wert-Paar und die überspannten Token.

**Markierung von Annotationen** Ein Klick auf den Dreiecksschalter hinter dem Namen der Annotationsebene blendet Informationen über das Annotationsschema, derzeit eine Auflistung der Tags, ein. Fahren wir mit der Maus über eines der Tags in dieser Liste, werden auf der Primärdatenzeile alle mit diesem Tag annotierten Ausdrücke hervorgehoben. Mit einem Klick auf ein Tag kann die farbliche Hervorhebung der zugehörigen Annotationsschicht aus- und eingeschaltet werden.

## Aufgaben

- Blenden Sie die Annotationen der Annotationsebene *pcc10.is.aboutness-topic* ein, indem Sie auf den Dreiecksschalter hinter dem Ausdruck “Annotationen” klicken.
- Bewegen Sie den Mauszeiger über die Annotationen auf verschiedenen Schichten und beobachten Sie, welche Segmente der Primärdaten farblich hervorgehoben, d.h. mit Annotation versehen sind. Verweilen Sie mit dem Mauszeiger über einer Annotation und interpretieren Sie die angezeigte Information.
- Blenden Sie mit einem Klick auf den Dreiecksschalter hinter dem Namen der Annotationsebene *pcc10.syntax-tiger* die Tags dieser Annotationsebene ein. Fahren Sie über einzelne Tags und beobachten Sie die Primärdatenzeile. Klicken Sie auf das Tag *rel*, das sich auf die Kanten zwischen den Knoten bezieht.

## 4.2 Diskursansicht

Die Diskursansicht soll dabei helfen, die Daten aus der Detailansicht in einen grösseren Kontext einzuordnen. Der Ausschnitt der Daten, der gegenwärtig in der Detailansicht betrachtet wird, wird dabei unterstrichen dargestellt.

Die Diskursansicht kann zudem dazu verwendet werden, ein anderes Segment in den Primärdaten zu fokussieren. Hierzu kann mit der Maus auf einen Ausdruck in der Diskursansicht geklickt werden, der dann den neuen Mittelpunkt der Detailansicht bildet.

Mit Hilfe eines Dreiecksschalters kann die Diskursansicht ein- und ausgeblendet werden.

## Aufgaben

- Blenden Sie die Diskursansicht mit einem Klick auf den Dreiecksschalter hinter “Diskursansicht” ein.
- Springen Sie zu einem anderem Diskurssegment, indem sie auf einen Ausdruck in der Diskursansicht klicken.

### 4.3 Navigationsleiste

Die Navigationsleiste enthält neben dem Namen des Primärdatums eine Leiste mit Navigationspfeilen, die das Scrollen in den Daten erlauben.

Derzeit kann i) an den Anfang und das Ende des Primärdatums gesprungen, ii) der Diskursausschnitt um eine halbe Ausschnittsbreite und iii) um ein Zeichen verschoben werden.

Falls Suchergebnisse betrachtet werden, erscheinen zusätzliche Buttons zum Abspeichern des Treffers in der Trefferliste (Pfeil nach unten) und zum Blättern zum vorherigen bzw. nächsten Treffer.

## Aufgaben

- Probieren Sie die Navigationspfeile aus!



Abbildung 5: Die Navigationsleiste

### 4.4 Zeigerrelationen

Typische Beispiele für die Verwendung von Zeigerrelationen sind neben anaphorischen sogenannte ‘Bridging’-Relationen: ein Ausdruck verweist auf einen anderen, mit dem er in einer bestimmten Beziehung steht.

Im aktuellen Prototyp von ANNIS werden diese Beziehungen zwischen Ausdrücken nur indirekt repräsentiert: Eine Annotation der Form “-> *markable ID*” zeigt auf einen Ausdruck mit der Kennung *ID*.

So befindet sich in Abbildung 6 über der Annotation “*part-whole*” (auf der Annotationsebene *pcc10.co-reference*) eine Annotation “-> *markable 17*”, die auf das Antezedens mit der ID “17” verweist. Damit wird angezeigt, dass der Ausdruck “Die Betriebe im Osten” in einer Bridging-Relation vom Typ *part-whole* zum Ausdruck “Eier-Produzenten aus der ganzen Republik” im vorherigen Satz steht (Der Ausdruck “Eier-Produzenten aus der ganzen Republik” wird von der Annotation mit der ID “17” überspannt).



NP				
	MNR			
	PP			
NK	AC	NK	NK	NK
t	t	t	t	t

Eier-Produzenten aus der ganzen Republik

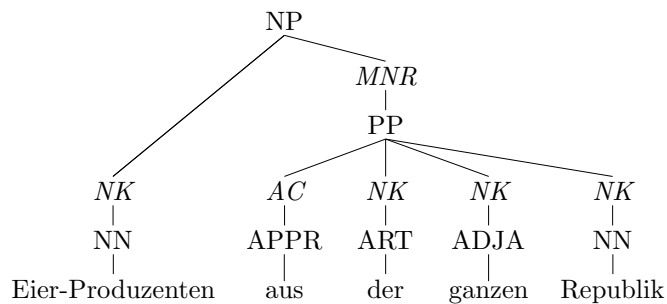


Abbildung 7: Schichten- und baumbasierte Darstellung der Syntaxstruktur einer NP

Syntaxstruktur gegenüber. Von der obersten Schicht beginnend wechseln Knoten repräsentierende Schichten – mit 'cat'-Tags wie NP und PP – mit Kanten darstellenden Schichten – mit Funktions-Tags wie MNR und NK – ab.

## 5 Die Suche

Thema dieses Abschnittes ist

- das Formulieren von Suchanfragen in ANNIS
- die Darstellung der Suchergebnisse
- sowie das Speichern und Laden ausgewählter Suchergebnisse.

Die Beispiel-Suchanfragen in diesem Abschnitt sind so gewählt, dass der Benutzer sie selbst eingeben und aktiv nachvollziehen kann (und das auch tun sollte!). Die Ausdrücke, die der Benutzer einzugeben hat, sind jeweils fett gedruckt.

Auf der linken Seite des ANNIS Bildschirms (in der Menüleiste) befindet sich das Suchfenster. In diesem Fenster werden die Anfragen formuliert; durch Drücken des "Los"-Buttons (oberhalb der Suchmaske) wird der Suchprozess gestartet. Die Suchergebnisse werden im Hauptfenster (rechts) angezeigt.

Hinweis: Sie können mit den Oben-/Unten-Pfeilen (links vom “Los”-Button) in Ihren alten Anfragen blättern und eine Anfrage erneut abschicken.

## 5.1 Einfache Suche

Die allgemeine Form einer einfachen Suchanfrage lautet: **“Text”** oder **Attribut=Wert**. Diese Form wird im Folgenden genauer erklärt.

### 5.1.1 Suche nach Text

Wir beginnen mit der Suche nach einzelnen Wörtern, Wortfragmenten oder ganzen Textpassagen. Dazu wird der gesuchte Ausdruck, umgeben von doppelten Anführungszeichen “...”, in das Suchfenster eingegeben und anschließend der Suchprozess durch Drücken des “Los”-Buttons gestartet.

**Beispiel 1** Wir suchen nach Texten, die den Namen “Johannes” enthalten. Geben Sie das gesuchte Wort in die Suchmaske ein: **“Johannes”**, und drücken Sie auf den “Los”-Button. Nach kurzer Zeit erscheinen im Hauptfenster die Suchergebnisse, die sogenannten ‘Treffer’.

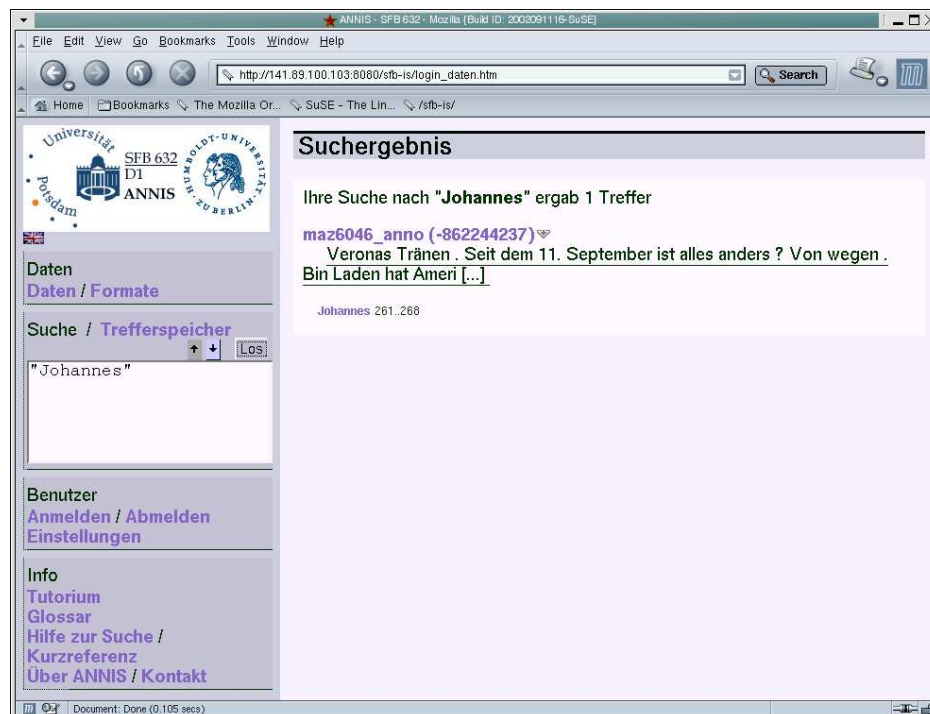


Abbildung 8: Eine Suchanfrage

In der ersten Zeile wird die Anzahl der Treffer angegeben. In unserem

Beispiel sollten Sie 1 Treffer erhalten. Das Dokument, das den Namen “Johannes” enthält, heißt “maz6046\_anno”. Direkt unter dem Dokumentnamen wird der Textanfang dieses Dokuments gezeigt: “Veronas Tränen ...”. Darunter kommt dann die Auflistung der Treffer, in unserem Fall nur einer: das Wort “Johannes”, violett hinterlegt, gefolgt von einer Positionsangabe (die Buchstaben des Wortes “Johannes” nehmen die Positionen 261-268 im Dokument ein).

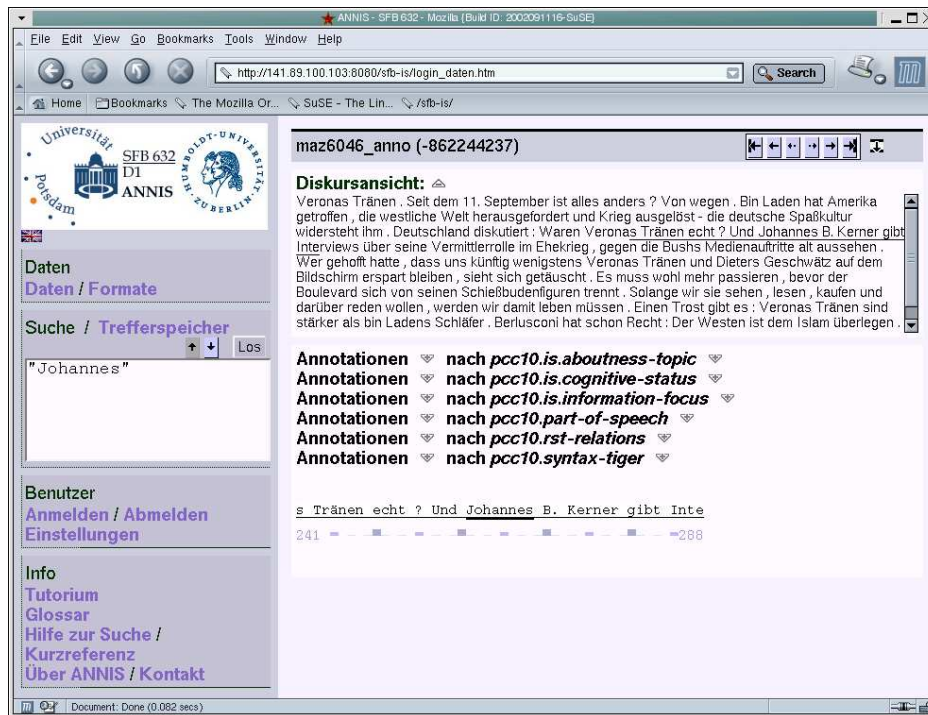


Abbildung 9: “Johannes” im Kontext

Die Treffer fungieren als Links, d.h. durch Klicken auf den Link können Sie sich das gesuchte Wort im Kontext anzeigen lassen. Klicken Sie nun auf “Johannes”, dann erhalten Sie den Kontext “...Und Johannes B. Kerner ...”. Im Text ist das Wort “Johannes” fett unterstrichen: die Unterstreichung markiert den Ausdruck, nach dem gesucht wurde.

Sie können nun entweder direkt eine neue Suchanfrage in das Suchfenster eingeben oder aber Sie kehren mit dem “zurück”-Knopf Ihres Browsers zur Trefferauflistung zurück.

Hinweis: Groß- und Kleinschreibung wird momentan noch nicht unterschieden. Sie erhalten also dieselben Treffer, wenn Sie “johannes” eingeben.

**Beispiel 2** Wir suchen nun nach dem Wort “erst”. Geben Sie dazu den entsprechenden Suchausdruck ein: **“erst”**. Sie sollten eine Reihe von Treffern erhalten, z.T. sogar mehrere innerhalb eines Dokuments. Z.B. finden sich im Dokument “maz-feigenblatt-11299” (bzw.: “maz11299\_anno”) 2 Treffer. Wählen Sie den ersten davon aus (“erst 91..94”) und lassen Sie sich den Kontext dazu anzeigen (durch Klicken auf “erst”). Der dazugehörige Kontext ist “bei der ersten Zossener Runde” — ein wohl unerwarteter Treffer. Es werden also nicht nur Instanzen des Wortes “erst” gefunden, sondern auch Ausdrücke, in denen “erst” enthalten ist, z.B. “ersten” oder **“Verständnis”** oder **“widersteht”**.

Um die anderen Treffer anzuschauen, haben Sie zwei Möglichkeiten: Entweder Sie machen (wie oben erwähnt) Gebrauch vom “zurück”-Knopf Ihres Browsers und kehren damit zur Trefferauflistung zurück. Oder Sie benutzen spezielle Pfeile, die zum Blättern zwischen den einzelnen Treffern gedacht sind: Oben rechts im Hauptfenster (rechts neben den normalen scroll-Pfeilen) erscheinen, sobald es mehr als einen Treffer gibt, Pfeile nach rechts und links. Klicken Sie auf den rechten Pfeil, um zum nächsten Treffer zu gelangen (“der Stadt dafür Verständnis haben”), oder auf den linken Pfeil, um zum vorhergehenden Treffer zu kommen. (Der Pfeil nach unten dient zum Speichern ausgewählter Treffer. Mehr dazu s. unten.)

Wie kann man nun die Anfrage so modifizieren, dass tatsächlich nur das Wort “erst” gefunden wird? Das gewünschte Ergebnis können Sie erreichen, indem Sie vor und hinter dem gesuchten Begriff ein Leerzeichen einfügen: **“ erst ”**. Jetzt sollten Sie deutlich weniger Treffer erhalten.

Man kann sich die besondere Eigenschaft der Textsuche auch zu Nutzen machen: z.B. können auf diese Weise alle Zusammensetzungen mit “Fußball” gefunden werden. Geben Sie den Suchausdruck **“Fußball”** ein. Sie sollten dann neben dem Wort “Fußball” auch “Fußball-Weltmacht” und “Fußball-Kleinmacht” finden (im Dokument “maz-furchtet-10374”).

### 5.1.2 Suche nach Annotationen

Die Texte in ANNIS sind mit verschiedener linguistischer Information annotiert, wie z.B. Informationsstruktur, Syntax etc. In den meisten Fällen besteht die annotierte Information aus einem Attribut mit einem Wert, wie z.B. dem Attribut “case” mit dem Wert “nominative”.

Neben reinem Text können Sie auch in diesen Annotationen suchen. Hier lautet der allgemeine Suchaufbau **Attribut=Wert**, also z.B. **case=nominative**.

**Beispiel 3** Wir suchen nach Topikausdrücken. In unseren Beispieldokumenten heißt das relevante Attribut “Topic”, ein möglicher Wert dieses Attributs ist “aboutness-topic”. Daher lautet die Suchanfrage **topic=aboutness-topic**. (Beachten Sie, dass der Suchausdruck hier ohne Anführungszeichen

eingetragen wird — dies unterscheidet die Suche nach Annotationen von der Suche nach Text.)

Sie sollten ca. 180 Treffer erhalten. Klicken Sie nun z.B. auf den ersten Treffer im Dokument “maz-alles-16153”: “In Berlin”. Sie gelangen damit, wie immer, zur Kontextansicht. Wenn Sie nun auf den Dreiecksschalter hinter dem Wort “Annotationen” der Annotationsebene *pcc10.is.aboutness-topic* klicken, so werden die zum Text gehörenden Annotationen mit angezeigt: “In Berlin” ist annotiert mit der Information “medium” (was die Konfidenz der Annotation angibt) und “aboutness-topic” (= dem Wert für das Attribut “topic”). Der Wert “aboutness-topic” ist dabei fett unterstrichen, da nach dieser Annotation gesucht wurde.

Fahren Sie nun mit dem Mauszeiger auf die Annotation “aboutness-topic” und warten Sie, bis der dazugehörige Informationstext erscheint, der die vollständige Annotation (inklusive des Attributs) anzeigt: “topic=aboutness-topic [4..5]”. Wir sehen also, dass der Informationstext der Suchanfrage entspricht. (“4..5” bedeutet: die Wörter “In Berlin”, die als Topik annotiert sind und daher den Treffertext darstellen, sind die Wörter 4-5 im Dokument. Der linke Kontext, “Alles offen .”, sind die Wörter 1-3.)



Abbildung 10: Suche nach Annotationen



## 5.2 Kombinierte Suche

Bisher haben wir in unseren Suchanfragen immer nur genau eine Suchbedingung angegeben: ein Text(fragment) oder eine Annotation.

Will man mehrere Bedingungen innerhalb einer Suche kombinieren, so werden zuerst die einzelnen Bedingungen aufgezählt, getrennt durch “&”: **Bedingung1 & Bedingung2** .... Anschließend können die Konditionen mit Hilfe verschiedener Operatoren kombiniert werden. Dabei wird auf die Bedingungen der Reihe nach mit #1 (= entspricht Bedingung1), #2 etc. Bezug genommen. Die allgemeine Form der Suchanfrage lautet damit: **Bedingung1 & Bedingung2 & ... & #1 Operator #2 & ....**

**Beispiel 4** In dem folgenden Beispiel suchen wir nach mehreren Wörtern, z.B. nach “nur noch”. Sie können entweder, wie oben gezeigt, direkt nach diesem Textstring suchen: **“nur noch”**. Ihre Anfrage sollte 2 Treffer erzielen.

Alternativ können Sie auch mit Hilfe des Präzedenz-Operators “.” fordern, dass der Ausdruck “nur” unmittelbar vor dem Ausdruck “noch” stehen soll. Die Suchanfrage lautet dann: **“nur” & “noch” & #1 . #2**. Wir suchen also nach dem Wort “nur” und dem Wort “noch”, wobei #1 “nur” entspricht und #2 “noch”. Der “.”-Operator fordert, dass das Wort “nur” dem Wort “noch” unmittelbar vorangeht.

Um die Zuordnung von #1 und #2 zu verdeutlichen, können wir die Bedingungen in der Suchanfrage auch umkehren: Geben Sie dazu ein: **“noch” & “nur” & #2 . #1**. D.h. der zweite Ausdruck soll nun vor dem ersten stehen. Jetzt sollten Sie das gleiche Ergebnis wie vorher im Hauptfenster angezeigt bekommen (2 Treffer).

Hinweis: Der “.”-Operator fordert unmittelbare Präzedenz. Sie können aber auch nach indirekter, mittelbarer Präzedenz suchen. Wenn Sie z.B. nach Sätzen suchen, die eine “je ... desto”-Konstruktion enthalten, können Sie das folgendermaßen tun: **“je” & “desto” & #1 .\* #2**. “.\*” erlaubt einen beliebigen Abstand zwischen den beiden Ausdrücken — Sie können ja auch nicht vorher wissen, wie groß der Abstand zwischen “je” und “desto” in den Beispielen tatsächlich ist.

**Beispiel 5** Analog zur kombinierten Suche nach mehreren Wörtern können Sie auch nach mehreren Annotationen suchen. Dazu können Sie die Form der obigen Suchanfrage einfach übertragen. Beispielsweise können Sie nach einem kognitiv inaktiven Element suchen, dem unmittelbar ein Topik folgt. In unseren Beispieldokumenten heißen die entsprechenden Attribute und Werte: “cognitive\_status=inactive” und “topic=aboutness-topic”. Die Anfrage würde also wie folgt lauten: **cognitive\_status=inactive & topic=aboutness-topic & #1 . #2**. Sie sollten 10 Treffer erhalten. Klicken Sie z.B. auf

den Treffer “von Gerhard Schröder am Samstag abend”, um zur Kontextansicht zu gelangen. Um die Annotationen zum kognitiven Status und Topik anzuschauen, müssen Sie wieder auf die entsprechenden Dreiecksschalter hinter “Annotations” klicken, in unserem Fall die von *pcc10.is.cognitive-status* und *pcc10.is.aboutness-topic*. Nun sollten Sie sehen, dass “von Gerhard Schröder” als “inactive” annotiert ist (hervorgehoben durch die fette Unterstreichung) und “am Samstagabend” als “topik”.

**Beispiel 6** Alternativ können Sie auch nach mehreren Annotationen eines einzigen Ausdrucks suchen, z.B. nach einem Ausdruck, der gleichzeitig als inaktiv wie auch als Topik annotiert ist. Dafür machen wir Gebrauch vom Abdeckungsgleichheit-Operator “\_=\_”, der verlangt, dass die genannten Konditionen für genau denselben Textstring gelten müssen. Die Anfrage lautet dann: **cognitive\_status=inactive & topic=aboutness-topic & #1 \_=\_ #2**. Hier erhalten Sie weit mehr Treffer als oben: 85.

**Beispiel 7** Es ist auch möglich, die Suche nach Text mit Suche nach Annotationen zu kombinieren. Nehmen wir an, wir suchen nach dem Wort “die”, das als Subjekt (“rel=sb”) fungiert. Dafür machen wir wieder Gebrauch vom Abdeckungsgleichheit-Operator “\_=\_”: **“die” & rel=sb & #1 \_=\_ #2**. Die Anfrage verlangt, dass das Wort “die” und die Annotation als Subjekt an der gleichen Position im Korpus stehen. Für diese Anfrage sollten Sie 6 Treffer erhalten.

Hinweis für besonders Interessierte: Man kann die Anfrage auch unterspezifiziert lassen und nur eine Abdeckungsüberschneidung (statt Abdeckungsgleichheit) fordern. Dazu lässt man die 3. Kondition einfach weg: **“die” & rel=sb**. Mit dieser Anfrage erhalten Sie dann z.B. auch Treffer wie “die ewige Besserwisserei der Alten” (= 2. Treffer im Dokument “maz10205\_anno”). Hier ist “die” nur (echter) Bestandteil eines Subjekts.

**Beispiel 8** Für die Suche in syntaktischen Annotationen gibt es einen weiteren Operator, den Dominanz-Operator “>”. Im folgenden Beispiel suchen wir einen Satz (“cat=S”), der einen anderen Satz unmittelbar dominiert. Wieder werden zuerst die Einzelbedingungen genannt, die anschließend durch den Operator “>” miteinander verknüpft werden: **cat=S & cat=S & #1 > #2**. Sie sollten 22 Treffer erzielen.

Der Dominanz-Operator erlaubt auch eine Spezifizierung der Funktion: “>func”. Sucht man z.B. nach eingebetteten Sätzen, die als Objekt (“rel=oc”) fungieren, lautet die Anfrage: **cat=S & cat=S & #1 >oc #2**.

Hinweis für besonders Interessierte: Die Operatoren zur Suche in Bäumen sind von TIGERSearch entlehnt, s. <http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERSearch/>.

**Beispiel 9** Wenn Sie wie oben nach Text oder Annotationen suchen, so durchsucht ANNIS stets alle Texte in der Datenbank. Sie können die Suche aber auch auf ein oder mehrere Dokumente beschränken.

Als Beispiel nehmen wir die Anfrage aus Beispiel 7: **“die” & rel=sb & #1 \_=#2**. (Zur Erinnerung: Sie können mit den Oben-/Unten-Pfeilen links des “Los”-Buttons in Ihren alten Anfragen blättern und eine Anfrage erneut abschicken.) Wenn Sie die Anfrage erneut starten, erhalten Sie 6 Treffer aus 2 verschiedenen Dokumenten. Jetzt können Sie die Suche weiter einschränken, indem sie den Suchraum auf ein Dokument beschränken, z.B. auf das Dokument maz10205\_anno. Der Dokumentname kann als eine Art von Annotation betrachtet werden; das Attribut heißt “doc”, der Name des Dokuments bestimmt den Wert des Attributs: “doc=maz10205\_anno”.

Die Suchanfrage für dieses Beispiel lautet dann: **“die” & rel=sb & #1 \_=#2 & doc=maz10205\_anno**. Jetzt werden nur noch Treffer innerhalb dieses Dokuments gefunden (3 Treffer).

The screenshot shows a web browser window with the ANNIS search interface. The search query entered is "die" & rel=sb & #1 \_=#2 & doc=maz10205\_anno. The results table, titled "Trefferspeicher", contains the following data:

Nr.	Dok.-ID	Dok.name	Zeit	Sucheingabe
5	223620983	maz10205_anno	25.08.2004 16:12	"die" & rel=sb & #1 _=#2 & doc=maz10205_anno
4	-407053837	maz5010_anno	25.08.2004 16:01	cat=S & cat=S & #1 >oc #2
3	-953197389	maz10374_anno	25.08.2004 16:01	cat=S & cat=S & #1 > #2
2	-862244237	maz6046_anno	25.08.2004 15:04	cognitive_status=inactive & topic=aboutness-topic & #1 _=#2
1	223620983	maz10205_anno	25.08.2004 15:03	"die" & rel=sb & #1 _=#2

Abbildung 11: Trefferliste

### 5.3 Speichern und Laden ausgewählter Suchergebnisse

Oft möchte man bestimmte Treffer abspeichern und zu einem späteren Zeitpunkt wieder ansehen. Das ist möglich in ANNIS: Als Beispiel nehmen wir wieder die Suchanfrage aus Beispiel 7: **“die” & rel=sb & #1 \_=#2**.

Wir erhalten im Hauptfenster die Liste der Treffer, welche violett hinterlegt sind. Klicken Sie den ersten der Links im Dokument “maz10205\_anno” an, um zur Kontextansicht zu gelangen: “die 69..71”.

Wie schon erwähnt, gibt es oben rechts im Hauptfenster einen Button (ein Pfeil nach rechts), um in den verschiedenen Treffern zu blättern. Klicken Sie auf den Pfeil, um zum zweiten Treffer zu gelangen. Jetzt können Sie entweder zum dritten Treffer weitergehen (Pfeil nach rechts) oder zum ersten zurückkehren (Pfeil nach links).

Links von den Pfeilbuttons gibt es einen Button mit einem Pfeil nach unten, der zum Speichern des aktuellen Treffers dient. Klicken Sie den Pfeil an, dann wird Ihnen die ID genannt, unter der der Treffer abgespeichert wird. (Um weitere Treffer abzuspeichern, kehren Sie mit dem “zurück”-Knopf Ihres Browsers zur Detailansicht zurück und blättern Sie zu weiteren Treffern.)



Abbildung 12: Pfeilbuttons: zum Navigieren, zum Abspeichern der Treffer und zum Blättern in den Treffern

Wenn Sie zu einem späteren Zeitpunkt auf die abgespeicherten Treffer zugreifen wollen, klicken Sie auf den Link “Trefferpeicher” (oberhalb des Suchfensters). Die bisher gespeicherten Treffer werden nun im Hauptfenster aufgelistet. In der ersten Spalte steht die ID des Treffers, in der dritten der Name des Dokuments, das den Treffer enthält. Spalte 4 gibt den Zeitpunkt des Abspeicherns an, Spalte 5 den Suchausdruck, mit der der entsprechende Treffer gefunden wurde.

Durch Klicken auf die ID gelangen Sie zur Kontextansicht des Treffers. Ein Klick auf das rote Kreuz rechts neben dem Treffer löscht den Eintrag.

## 6 Datenimport und -export

### 6.1 Datenimport

Der Datenimport in ANNIS 1.1 wird allein durch die Administratoren vorgenommen. Zu importierende Daten können an [annis@ling.uni-potsdam.de](mailto:annis@ling.uni-potsdam.de) geschickt werden.

### 6.2 Datenexport

Annotationsdokumente können sowohl im SFB Encoding Standard als auch im Originalformat, d.h. im Format, in dem die Daten annotiert wurden, exportiert werden.

Dazu dient die Exportleiste, die in der Übersicht über die Annotationsdokumente einer Annotationsebene (Abschnitt 3) und in der Annotationsansicht (Abschnitt 4) zur Verfügung steht. Neben der Betrachtung des Quellcodes der einzelnen Annotationsdokumente können sie heruntergeladen als auch per e-mail verschickt werden.

## 7 Verschiedenes

### 7.1 Benutzereinstellungen

Die Option 'Einstellungen' im Menüabschnitt '*Benutzer*' stellt eine Reihe von Möglichkeiten zur Anpassung der Oberfläche an die Nutzerbedürfnisse zur Verfügung.

### 7.2 Benutzerrechte

Benutzerrechte werden durch die Administratoren vergeben. Derzeit sind das Betrachten der Daten, der Export und das Administrieren von den individuellen Nutzerrechten abhängig.

## 8 Zusammenfassung und Ausblick

ANNIS versucht, einer Reihe von sehr unterschiedlichen Anforderungen gerecht zu werden, die hier kurz skizziert werden.

- *Heterogene Sprachdaten*: Die Primärdaten sind sehr heterogen. Sie umfassen Monologe und Dialoge, gesprochene und geschriebene Sprache, Sprachen mit verschiedenen Schriften, usw. ...
- *Heterogene Annotation*: Annotationen der Daten nutzen verschiedene Datentypen, von einfachen Attribut-Wert-Paaren, über Zeigerrelationen, Bäume bis hin zu Graphen. Zudem ist die Annotation sowohl Wort- (Part-of-Speech), Satz- (Syntaxstruktur, grammatische Funktionen, ...) als auch Diskursbezogen (Diskursstruktur, Koreferenz).
- *Visualisierung und Suche auf mehreren Ebenen*: Annotationen auf mehreren Ebenen sollen gleichzeitig betrachtet werden können, um z.B. das Zusammenspiel verschiedener linguistischer Ebenen (Prosodie, Wortstellung, Grammatische Funktion) bei der Realisierung von Informationsstruktur zu erforschen. Aus dem gleichen Grund sollen sich Suchanfragen auf Annotationen verschiedener Annotationsebenen beziehen können.
- *Benutzerfreundlichkeit* Die Nutzer von ANNIS sollen in der Lage sein, ohne umfangreiche Installationen und aufwendiges Training mit der

Datenbank umzugehen. So soll ANNIS mit einfachen Webbrowsern aufgerufen werden können.

Der aktuelle Prototyp erfüllt einige dieser Anforderungen in Ansätzen, lässt andere jedoch noch unberücksichtigt. Dies betrifft insbesondere die folgenden Punkte:

- *Visualisierung*: Das derzeit verwendete schichten-basierte Modell erlaubt die gleichzeitige Betrachtung von Annotationen verschiedener Art (d.h. vor allem verschiedener Datentypen wie einfache Attribut-Wert-Paare, Zeiger und Baumstrukturen) auf mehreren Annotations-ebenen. In zukünftigen Versionen soll dieses Modell durch weitere interaktive Komponenten ergänzt bzw. ersetzt werden, insbesondere für eine geeignete Visualisierung von Baumstrukturen und Zeigern.
- *Suche*: Die Suche erlaubt ebenen-übergreifende Anfragen auf beliebigen Daten<sup>4</sup>, die ohne weitere Vorverarbeitungs- und Optimierungsschritte (wie z.B. Indizierung) in ANNIS importiert werden. Dies spiegelt sich jedoch in der geringen Performanz der Suche bei komplexen Suchanfragen. Das soll verbessert werden.
- *Organisation der Daten*: In zukünftigen Versionen von ANNIS werden weitere Ordnungsprinzipien wie Sprache der Primärdaten, Annotierer, Zeitpunkt der Annotation etc. unterstützt.

Die Verwendung von ANNIS durch die Teilprojekte im SFB und die Integration von im SFB erhobenen Daten werden die weitere Entwicklung von ANNIS bestimmen.

Rückmeldungen und Kritik sind überaus willkommen: [annis@ling.uni-potsdam.de](mailto:annis@ling.uni-potsdam.de).

---

<sup>4</sup>D.h. auf Daten, die in einem der Importformate – derzeit MMAX, Exmaralda, RSTTool-RS3, TIGER – vorliegen.