

Guidelines

zur Annotation semantischer Rollen im Deutschen mit VerbNet-basiert definierten Rollen

verwendet in der Annotationsstudie des CLARIN-D-Kurationsprojekts
"Semantische Annotation für Digital Humanities"

(Éva Mújdricza-Maydt, Silvana Hartmann, Anette Frank)

(Finale Version 20170809)

Inhaltsverzeichnis

1 Einführung	2
1.1 Annotationsvarianten	2
2 Lesartendisambiguierung der Prädikate (WSD)	
2.1 Was wird annotiert?	4
2.1.1 Grundlagen zur Prädikatannotation	4
2.1.2 Grundlagen zur Annotation von Lesarten	5
2.2 Wie wird annotiert?	5
2.2.1 Annotationsvarianten	6
2.3 Korrektur der Annotation	6
2.4 Schwierige Fälle	6
2.4.1 Mehrteilige Prädikate: Pred- und PredPart-Annotation	6
2.4.2 Kommentar-Annotation: Note-Layer	7
3 Annotation semantischer Rollen (SRL)	
3.1 Was wird annotiert?	9
3.2 Wie wird annotiert?	11
3.3 Schwierige Fälle	13
3.3.1 Mehrfachannotation bei mehrmaliger Referenz	13
3.3.1.1 Überlappungen in Prädikat-Argument-Strukturen	13
3.3.1.2 Elidiertes Prädikat	13
3.3.1.3 Multiple Lesarten	14
3.3.2 Nicht lokal realisierte Argumente	14
3.3.3 Annotationsvarianten für Argumente	14
3.3.3.1 Annotation ganzer Phrasen	14
3.3.3.2 Annotation des Kopfelements	15
3.3.4 Mehrteilige semantische Prädikate und Argumente	15
3.3.5 Kommentar-Annotation: Note-Layer	16
4 Verwendung von WebAnno	
4.1 Allgemeine Hinweise	17
4.2 Einstellungen des Annotationsprozesses (Settings)	17
4.3 Annotation verschiedener Annotationstypen in WebAnno	18
4.3.1 „span“-Annotation	18
4.3.2 „relation“-Annotation	18
4.3.3 „slot“-Annotation für semantische Rollen	19
5 Anhang	
5.1 Übersicht zu GermaNet-Lesarten	20
5.2 Übersicht der semantischen Rollen	20
6 Literaturverzeichnis	

1 Einführung

Die Annotationsstudie “*Semantische Annotation für Digital Humanities*”¹ wurde im Rahmen des gleichnamigen CLARIN-D-Kurationsprojekts durchgeführt, und hatte die Zielsetzung, die Annotation von **Lesarten von Prädikaten** (*Word Sense Disambiguation, WSD*) sowie die Annotation **semantischer Rollen** (*Semantic Role Labeling, SRL*) für die Argumente der Prädikate auf **deutschen** Korpora zu erproben.

In der Studie wurden zusammenhängende Textteile aus dem **SALSA2.0-Korpus**² sowie aus dem **Dortmunder Chat-Korpus**³ als Grundlage für die Annotation verwendet. Lesarten wurden für ausgewählte Prädikate anhand der **GermaNet**⁴-Lesarten (V9.0) annotiert. Bei der Annotation der semantischen Rollen für die Argumente der Prädikate wurde ein angepasstes Rolleninventar hauptsächlich auf Grundlagen der semantischen Rollen aus **VerbNet**⁵ ausgearbeitet. Die Beschreibung der Rollen ist als Begleitmaterial zu diesem Dokument online herunterladbar⁶.

Im Verlauf der Annotationsstudie wurden mehrere **Varianten des Annotationsvorgangs** durchgeführt, um eine Vergleichsgrundlage zwischen relevanten Kriterien zu ermöglichen und bestmögliche Annotationsgrundlagen zu erproben.

Im Folgenden werden die Varianten zusammenfassend dargestellt, anschließend werden die Teilaufgaben WSD und SRL im Detail beschrieben. Das abschließende Kapitel beschreibt die Annotationsschritte im verwendeten Annotationstool **WebAnno**⁷.

1.1 Annotationsvarianten

Generell wurden folgende Annotationsvarianten festgelegt und in Kombination erprobt:

1. Prädikat-Auswahl
 - a) Markierte Prädikate im Textfluss annotieren: Der ursprüngliche Text mit seinem Textzusammenhang bleibt erhalten, die Annotatoren blättern zu dem jeweils nächsten vorannotierten Prädikat im Text.
 - b) Markierte Prädikate gebündelt nach Lemma annotieren: Die Prädikate werden anhand vorliegender Lemma-Annotation aus dem Textfluss extrahiert. Die Annotatoren können alle Prädikate zu einem Lemma gebündelt annotieren. Um den Textzusammenhang aufrechtzuerhalten, werden die Sätze mit dem aktuellen Zielprädikat in einem Kontextfenster von 3-5 Sätzen vor und nach dem Satz mit dem aktuellen, vorannotierten Prädikat jeweils mit extrahiert.
2. Prädikat-Annotation
 - a) Prädikate nur mit Lemma annotieren (keine weitere Lesartenannotation)
 - b) Prädikate mit Lesart annotieren
 - i. Annotation einer besten Lesart

1 Webseite des Projekts: <http://www.cl.uni-heidelberg.de/projects/clarin-dcp/>

2 SALSA-Korpus: <http://www.coli.uni-saarland.de/projects/salsa/>

3 Dortmunder Chat-Korpus: <http://www.chatkorpus.tu-dortmund.de/>

4 GermaNet: <http://www.sfs.uni-tuebingen.de/GermaNet/>

5 VerbNet: <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>, Rollenreferenz: <http://verbs.colorado.edu/verb-index/vn/reference.php>

6 SemAnno-Rolleninventar: http://www.cl.uni-heidelberg.de/projects/GNVN_semanno/material/SemAnno_RoleInventory.pdf

7 WebAnno-Webseite: <https://webanno.github.io/webanno/>

ii. Annotation aller passenden Lesarten mit zusätzlicher Auswahl einer besten Lesart

3. Rollen-Annotation

- a) Annotation der Rollen auf der gesamten Wortsequenz des Arguments: Je nach Argument kann dies ein Wort, eine Nominal- oder Präpositionalphrase, oder längere Sequenzen bis zu einem Satz umfassen. Diskontinuierliche Sequenzen müssen erkannt und mit annotiert werden.
- b) Annotation der Rollen auf dem Kopfwort des Arguments: Immer ein Wort wird annotiert, welches im Idealfall anhand einer unterliegenden Dependenzstruktur-Annotation für die Repräsentation des Arguments zur Annotation gewählt wird.

Für die endgültige Annotation wurden folgende Annotationsrichtlinien als Grundlage verwendet: **Prädikate** wurden **lemmaweise** (Option 1.b) **mit allen passenden GermaNet-9.0-Lesarten** annotiert, wobei jeweils eine **beste Lesart** ausgewählt wurde (2.b.ii.). Zu jeder annotierten Prädikatlesart wurden die **Argumente am Kopfwort** durch eine semantische Rolle ausgezeichnet (3.b). In wenigen Fällen wurden auch mehrere Rollen für eine Lesart annotiert.

Die Annotationen sind für weitere Verwendung in der Forschung online zugänglich⁸.

8 http://www.cl.uni-heidelberg.de/projects/GNVN_semanno/data.shtml

2 Lesartendisambiguierung der Prädikate (WSD)

Annotationsaufgabe: Gegebene Prädikate werden mit ihrer Lesart, z.B. **GermaNet⁹-Lesart-Id** (V9.0) markiert. Eine Beschreibung der Lesarten liegt den Annotatoren vor (siehe Kapitel 5.1).

2.1 Was wird annotiert?

2.1.1 Grundlagen zur Prädikatannotation

Im Rahmen dieser Studie werden in erster Linie **Vollverben** als **Prädikate** annotiert.

Vollverben sind Verben, die **ohne Hilfe eines anderen** Verbs das Prädikat, die zentrale strukturgebende Kategorie des Satzes, bilden können. Sie besitzen eine **eigene Bedeutung**. Sie drücken Ereignisse (Bsp.: *feiern*), Veränderungen (*abbrechen*), oder auch Zustände (*besitzen*) aus. Gemäß ihrer Bedeutung verlangen Vollverben eine Anzahl an Mitspielern, die als Subjekt, Objekte oder freie Ergänzungen die im Verb beschriebene Situation vervollständigen.

Im Rahmen dieser Annotationsaufgabe werden auch **Funktionsverbgefüge** annotiert. Diese sind Verbkonstruktionen, die aus einem Verb und einer Nominal- oder Präpositionalphrase bestehen wie

- Bsp. *jemandem **Nachricht geben*** 'jemanden benachrichtigen'; *einer **Prüfung unterziehen*** 'prüfen'; *in **Kraft treten***.

Funktionsverben haben eine geschwächte oder keine eigene Semantik, sondern sie bilden mit der nominalen Phrase zusammen eine Bedeutungseinheit. Man kann sie oft auch durch einfache Verben (wie oben in Beispielen) ersetzen. Aus diesen Gründen werden die zugehörigen Nominal- oder Präpositionalphrasen als Teil des Prädikats behandelt.

Idiomatische Verbalkonstruktionen verhalten sich wie Funktionsverbgefüge, z.B. *ins **Gras beißen*** für 'sterben' wird als ein Prädikatkomplex annotiert.

Mit diesen Konventionen orientieren wir uns unter Anderem an den im SALSA-Projekt entwickelten Annotationsrichtlinien¹⁰.

Nicht annotiert werden Hilfsverben, Modalverben und Kopulakonstruktionen.

- Die Hilfsverben *haben*, *sein* und *werden* dienen dazu, Formen der Vergangenheit, der Zukunft, bzw. Passiv und Konjunktiv zu bilden. Beispiele:
 - *Er **ist** spät nach Hause gekommen.*; *Er **wird** wieder spät ankommen.*; *Wir **wurden** für diese Aufgabe ausgewählt.*; *Das Fenster **ist** geöffnet.*; ***Wäre** er nur früher angekommen!*
- Modalverben modifizieren die Bedeutung des Hauptverbs, sie bezeichnen eine Möglichkeit, Notwendigkeit, Erlaubnis oder Fähigkeit. Dazu gehören: *können*, *dürfen*, *müssen*, *sollen*, *wollen*, *mögen* und ihre Formen.
 - *Er **kann** warten.*; *Morgen **dürfte** Hans wieder da sein.*
- Kopulakonstruktionen sind eine Verbindung der Verben *sein*, *werden*, *bleiben* mit einem Nomen oder Adjektiv, die als Prädikative bezeichnet werden. Die Bedeutung des Prädikats trägt das Prädikativ. Man kann diese Konstruktion von Funktionsverbgefügen

9 GermaNet: <http://www.sfs.uni-tuebingen.de/GermaNet/>

10 Vgl. [Burchardt et al. 2009].

unterscheiden, denn Prädikative kommen hauptsächlich mit den oben genannten Verben vor.

- *Daniela ist wieder **gesund**.; Meine kleine Katze wird mal ein **guter Jäger**.; Sie **bleiben** die ganze Nacht **online**.*

2.1.2 Grundlagen zur Annotation von Lesarten

Für die Lesartenannotation für die zugrundeliegenden Vollverben bzw. Verbkomplexe verwenden wir im Rahmen dieser Annotationsstudie das Bedeutungsinventar von GermaNet.

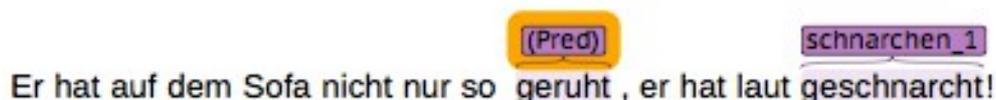
Eine **Lesart** ist eine semantisch definierbare Bedeutung für ein Wort -- das Zielverb bzw. den Verbalkomplex --, wobei ein Wort auch mehrere Lesarten aufweisen kann. In diesem Fall gilt das Wort als mehrdeutig, ambig. Die Lesarten eines Wortes können in der Regel (jedoch nicht immer) durch die Verwendung im textuellen Kontext unterschieden werden. Die Bestimmung der korrekten Lesart für ein Wort nennt man Disambiguierung.

Beispielsweise hat das Verb *ruhen* in GermaNet 9.0 vier **Lesarten**:

1. (Allgemein): {pausieren, rasten, ruhen}¹¹ – Bsp.: *Die Produktion **ruht** wegen des Streiks.*
2. (Körperfunktion): {ruhen} – Bsp.: *Mein Opa **ruht** auf dem Sofa.*
3. (Körperfunktion): {nächtigen, ruhen} – Bsp.: *Er **ruhte** bis zur Mittagszeit.*
4. (Lokation): {ruhen} – Bsp.: *Das Dach **ruht** auf Säulen*

Für diese Annotationsaufgabe werden **vorannotierte Verben** betrachtet. Die Vorannotation wird als eine **generische** Markierung gesetzt: Die annotierte Verbinstantz enthält in der Annotation das unspezifische Label „(Pred)“¹². Diese Annotation soll durch die passende GermaNet-Lesart-Id ergänzt werden.

Im folgenden Beispiel¹³ ist das erste Prädikat generisch vorannotiert, das zweite bereits mit der passenden Lesart annotiert.



Er hat auf dem Sofa nicht nur so geruht, er hat laut geschnarcht!

Abbildung 1: Generische Vorannotation "(Pred)" für das Wort "geruht", sowie annotierte Verbinstantz "geschnarcht" annotiert mit WebAnno.

2.2 Wie wird annotiert?

Zur Annotation der vorhandenen generischen Annotationsinstanz führen Sie einen **Doppelklick auf der „(Pred)“-Annotation** aus, sodass die annotierte Instanz gelb umrandet erscheint wie in Abbildung 1 gezeigt wird. Auf der rechten Seite der Annotationsoberfläche werden die Eigenschaften (Features) dieser Instanz aktiviert bzw. eingeblendet.

¹¹ Die Bedeutung der Wörter wird in den nächsten Beispielen durch synonyme, gleichbedeutende Wörter veranschaulicht, wobei die synonymen Wörter in geschweiften Klammern zusammengefasst werden.

¹² Die Annotationsebenen *Pred* und *SemPred* sind Namensvarianten für den gleichen Inhalt, und werden in diesen Guidelines als gleichwertig und gleichbedeutend behandelt.

¹³ Folgende Beispiele zeigen Bildschirmfotos vom Annotationstool WebAnno, in der Version 2.

Sie müssen ausschließlich das Feature „**senseld**“ mit Inhalt füllen: Wählen Sie eine der angebotenen Lesarten aus. Halten Sie den Mauszeiger über eine der möglichen Lesarten in der Drop-box-Liste für senseld. Es erscheint eine **Beschreibung** für die Lesart, anhand der Sie sich für oder gegen der Lesart entscheiden können. Bei der Entscheidung für eine **passende Lesart** ist der **Kontext** des Prädikats einzubeziehen. Nach Auswahl einer Lesart wird die Änderung der Annotation automatisch übernommen, und statt „(Pred)“ erscheint die gewählte Lesart über dem Token.

Es besteht die Möglichkeit, dass keine der angebotenen Lesarten zu dem Prädikat passt. Die richtige **Lesart ist** in diesem Fall **nicht im aktuellen Lexikon** enthalten. Wählen Sie in diesem Fall die Lesart „**XXX**“.

Es werden auch die Verben nur generisch vorannotiert, für die **nur eine Lesart** in GermaNet vorliegt. Der Annotator muss auch hier **prüfen**, ob die angebotene Lesart im aktuellen Kontext angemessen ist. Ist dies nicht der Fall, ist die Default-Lesart „**XXX**“ zu wählen.¹⁴

2.2.1 Annotationsvarianten

Es ist möglich, dass **mehrere Lesarten** passend scheinen. In diesem Fall können **alle** passenden Lesarten annotiert werden. Erstellen Sie zu diesem Zweck eine neue „Pred“-Annotation für jede Lesart und versehen Sie diese mit der passenden Lesart-Id. Zusätzlich soll eine Lesart, die am **besten** zum ausgewählten Prädikat passt, ausgewählt werden. Dies erfolgt durch das Anklicken des Attributs "bestSense".

2.3 Korrektur der Annotation

Eine annotierte Instanz kann man jederzeit ändern oder auch löschen. Mit einem Doppelklick auf das Instanz-Label können die bisher annotierten Eigenschaften zum Prädikat aktiviert werden. Die Änderung einer Eigenschaft oder das Löschen der Lesarten-Instanz (durch Anklicken des Knopfes „Delete“) wird automatisch sofort durchgeführt und angezeigt.

Aktuell besteht keine Möglichkeit, eine Änderung rückgängig zu machen.

2.4 Schwierige Fälle

2.4.1 Mehrteilige Prädikate: Pred- und PredPart-Annotation

Besteht das Prädikat aus mehr als einem Token wie im folgenden Satz

Er _{Hauptteil} **kam** *pünktlich* _{Verbzusatz} **an**,

annotiert man nur die Pred-Instanz, die man als Hauptbestandteil des Prädikats erkennt. Dieses Wort wird allen anderen Teilen übergeordnet.



Abbildung 2: Annotation mehrteiliger Prädikate mit PredPart-Relation zum Hauptverb am Beispiel "lehnen ... ab".

Im Beispiel der Abbildung 2 erscheint der Verbzusatz *ab* des Prädikats *ablehnen* am Ende des Satzes. **Beide Teile** werden als **Pred**-Instanzen vorannotiert. Ihre **Zusammengehörigkeit** wird

¹⁴ In diesem Fall kann der Annotator einen Kommentar mit der Beschreibung der Verbbedeutung hinterlassen (siehe Kapitel 2.4.2).

durch eine generische **PredPart**-Kante angezeigt, wobei diese Kante vom Hauptverb zum Verbzusatz zeigt. An der Richtung der Kante erkennt man, dass hier das Verb *lehnen* als übergeordnete Instanz gilt, und somit muss man nur diese Instanz disambiguieren.

Wird bei einem mehrteiligen Prädikat ein Teil nicht vorannotiert, ist diese Annotation manuell zu ergänzen:

1. Nachannotieren der fehlenden *Pred*-Instanz: Wählen Sie das Layer *Pred* zum Annotieren aus. Durch Doppelklick auf das Wort mit der fehlenden Annotation legt man die Instanz an.
2. Verbinden der *Pred*-Instanzen durch eine *PredPart*-Kante: Klicken Sie die übergeordnete *Pred*-Instanz (den Verbstamm) an. Ziehen Sie mit gedrückter Maustaste eine Kante zur untergeordneten *Pred*-Instanz (z.B. Verbzusatz oder nominale Teile eines mehrteiligen Prädikats), so dass diese Instanz auch hervorgehoben wird. Lassen Sie die Maustaste los, die Kante erscheint zwischen den beiden Teilen.
3. Alle weiteren Prädikat-Teile sollten vom einen Verbstamm abhängig nachannotiert werden.

Relevante Fälle: Die semantisch untergeordneten Prädikatteile sind in den nächsten Beispielen mit eckigen Klammern markiert, die Hauptbestandteile mit geschweiften Klammern:

- vom Hauptverb trennbare **Verbzusätze** – Bsp.: {lehnte} [ab]; {nahm} [teil]; [kennen] {lernen}
- Reflexivpronomen bei **echt reflexiven Verben** – Bsp.: er {freute} [sich] sehr; Der Mann {bildet} [sich] etwas [ein].
 - Bei echt reflexiven Verben ist das Reflexivpronomen **nicht weglassbar** und **stimmt** in Person und Numerus **mit dem Subjekt überein**:
 - ich {freue} [mich]; du {freust} [dich]; er/sie/es {freut} [sich]; wir {freuen} [uns] ...
 - Grammatisch nicht korrekt: *ich freue dich; *er freut mich
- Vorsicht: Auch nicht echt reflexive Verben können **reflexiv verwendet** werden. Bei diesen Verben ist das Reflexivpronomen ein Argument und muss als solches behandelt werden (**nicht** Teil des Prädikats). Diese Reflexivpronomen könnten auch durch andere Argumente ersetzt werden. – Bsp.: Er {rasiert} sich/den Kunden. Sie {achten} gut auf sich/auf ihn.
- **Platzhalter- oder unpersönliches „es“** – In diesen Fällen referiert das Personalpronomen es nicht auf eine Entität, sondern trägt eine rein grammatische Funktion.
 - Bsp.: [es] {donnerte}; [es] {gibt} mehr zu verstehen; [Es] {besteht} die Möglichkeit, dass ...
 - Aber: es als Korrelat wird nicht als Teil des Prädikats gezählt, sondern als Argument, denn Korrelat-es weist eine Referenz auf einen anderen Satzteil auf. Bsp.: Ich {finde} es toll, mit euch zu fahren.
- **idiomatische** Verbalkonstruktionen, Funktionsverbgefüge: Diese Teile idiomatisierter Verben werden als eine Bedeutungseinheit empfunden und verstanden. Beispiele:
 - etwas [in Frage] {stellen}; jemanden [auf die Palme] {bringen}; etwas [zur Verfügung] {stellen}, etwas [in Betracht] {ziehen}
- **Nicht** Teil des Prädikats sind die „zu“-Partikeln bei Infinitiv-Konstruktionen wie – Bsp.: Ich habe dir {versprochen} dich nach Hause zu {fahren}. – Die Partikel „zu“ wird nicht annotiert.

2.4.2 Kommentar-Annotation: Note-Layer

Fällt die Lesartendisambiguierung besonders schwer, oder ist der Fall im gegebenen Kontext nicht entscheidbar, wählt man die Lesart, die am meisten passt oder alternativ die unspezifische Lesart "XXX".

Zusätzlich wird empfohlen eine Annotation *Note* mit Anmerkungen, Kommentaren oder Fragen zu erstellen, die auf den problematischen Fall hinweist (siehe z.B. Abbildung 3). Auch auf Fehler in der vorhandenen Annotation kann man mit einem Kommentar hinweisen.



Abbildung 3: Unsichere Annotation wird mit Layer "Note" beschrieben -- hier gelb umrandet.

Eine **Note-Annotation** erstellt man, indem man das Layer Note zum Annotieren in der rechten Annotationsoberfläche auswählt, und anschließend das fragliche Token anklickt. Über dem Token erscheint die neue Annotation zunächst generisch. Auf der rechten Seite der Annotationsoberfläche können Sie unter dem Feature „Beschreibung“ eine (kurze) Anmerkung zum problematischen Fall frei formulieren. Diese Formulierung wird als Beschriftung der Note-Instanz gespeichert.

3 Annotation semantischer Rollen (SRL)

Prädikate (Verben) und ihre **Argumente** werden in dieser Annotationsaufgabe miteinander **verbunden**. Die annotierten Annotationsebenen sind **SemPred** für die Prädikate und **SemArg** für die Argumente¹⁵. Die Prädikate, und (je nach gewähltem Annotationsschema) auch die Argumente liegen bereits annotiert vor. Durch das Feature „**RoleSet**“ des Layers SemPred (siehe Kapitel 4.3.3) werden für die SRL-Annotation SemArg-Instanzen mit der Zusatzinformation über semantische Rollen ans Prädikat gebunden. Abbildung 4 zeigt eine vollständige Annotation, wobei die semantischen Rollen als Labels der Kanten zu den Argumenten gekennzeichnet sind.



Abbildung 4: Fertige Annotation mit semantischen Rollen zu Prädikaten. Bei mehrfacher Abbindung eines Arguments (wie "Fritz") wird die SemArg-Instanz mehrfach erzeugt, sodass jede SemArg-Instanz eindeutig an eine SemPred-Instanz angebinden wird.

3.1 Was wird annotiert?

Die Satzglieder, die vom gegebenen Verb grammatisch und semantisch verlangt werden, werden als **Argumente** oder Komplemente bezeichnet. Ohne diese Argumente ist der Satz entweder grammatisch inkorrekt (das gegebene Satzglied fehlt), oder semantisch nicht vollständig.

Für diese Annotationsaufgabe werden **obligatorische und fakultative Argumente** (Komplemente) zu einem Prädikat annotiert, nicht aber freie Ergänzungen (Adjunkte). Letztere kann man ohne eine grammatische oder semantische Beeinträchtigung des Satzes weglassen. Sie ergänzen das beschriebene Ereignis mit Informationen, die nicht von der Verbbedeutung her verlangt werden. In den folgenden Beispielen werden Argumente mit eckigen, freie Ergänzungen mit runden Klammern gekennzeichnet.

- Bsp.: **[Fritz] liest (im Garten) [ein Buch]**. – *Fritz* ist ein obligatorisches Argument, *ein Buch* ein optionales. Beide werden annotiert. Die Präpositionalphrase *im Garten* ist eine freie Ergänzung und bleibt unannotiert.
- Subjekte und Objekte sind in der Regel Argumente. Adverbiale (Präpositionalphrasen, Adverbien oder ganze Sätze) können sowohl als Adjunkt, als auch als Argument auftreten. Ein Beispiel für präpositionale Argumente: **[Der Affe] sitzt [auf dem Baum]**.; **[Er] benimmt sich [wie ein Idiot]**.;
- Dativobjekte sind auch nicht eindeutig: **[Er] widmet [ihr] [das Buch]**. Aber: **[Er] tanzt ihr zu schlecht**.

Eine klare Unterscheidung zwischen fakultativen Argumenten und freien Ergänzungen ist nicht immer möglich. Als Entscheidungshilfe können Sie die zusammengefassten Eigenschaften dieser Elemente in der folgenden Tabelle¹⁶ verwenden.

¹⁵ Die Annotationsebenen Pred und PredPart stellen die gleichen Inhalte wie die hier beschriebenen Typen SemPred und SemPredPart.

¹⁶Die Tabelle wurde nach [Pittner und Berman 2006], S. 49 erstellt.

Argument (Komplement)		freie Ergänzung (Adjunkt)
obligatorisch	fakultativ	
verbspezifisch nicht weglassbar ohne dass der Satz ungrammatisch wird	unter bestimmten Bedingungen weglassbar <ul style="list-style-type: none"> • mitverstanden • unbestimmt • kontextuell gegeben 	weglassbar
(a) Weglasstest, (b) Ausklammerungstest und (c) Nachtragstest erzeugen einen <u>ungrammatischen</u> Satz (a) *Hans beantwortet. (b) *Er hat den Fernseher geworfen, und das geschieht aus dem Fenster . (c) *Hans wohnt, und zwar unter der Brücke .	(a) <u>Weglasstest</u> : Der Satz bleibt auch ohne das gegebene Satzglied grammatisch korrekt. <i>Anna isst.</i> (c) <u>Nachtragstest</u> : Nachtrag mit <i>und zwar</i> möglich <i>Anna isst, und zwar Bananen.</i>	(b) <u>Ausklammerungstest</u> : Umformung in einen extra Satz möglich <i>Die Kinder spielen, und das geschieht im Garten.</i> (c) <u>Nachtragstest</u> : Er schlägt die Nagel in die Wand, und zwar mit einem Hammer .

Tabelle 1: Differenzierung zwischen Argumenten und Adjunkten nach [Pittner und Berman 2006], S. 49.

Die **verwendeten semantischen Rollen** sind mit einer ausführlichen Beschreibung und mit Beispielen sowie einer Entscheidungshilfe in der anliegenden Datei *SemAnnoRoleInventory.pdf* aufgelistet¹⁷.

Beispiele für oft vorkommende semantische Rollen mit einer gekürzten Erläuterung -- Die Argumente werden durch eckige Klammern gekennzeichnet und mit der passenden Rolle markiert:

- **Agent**: Der Teilnehmer, der eine Handlung absichtlich und bewusst ausführt.
 - [_{Agent} **Der Torwart**] hat [_{Theme} *den Ball*] {*gefangen*}.
- **Patient**: Ein Betroffener im Ereignis, der auf eine Weise manipuliert, beeinträchtigt wird.
 - [_{Agent} *Die Hexe*] hat [_{Patient} **ihn**] [_{Goal} *in einen Frosch*] {*verwandelt*}.
- **Beneficiary**: Der Betroffene, der potentiell einen Vor- oder Nachteil aus dem Ereignis zieht.
 - [_{Agent} *Daniel*] {*half*} [_{Beneficiary} **mir**] [_{Theme} *bei der Hausaufgabe*].
- **Theme**: worum es sich in der Situation handelt.
 - [_{Agent} *Emma*] hat [_{Theme} **den Ofen**] {*angemacht*}.

Semantische Rollen sollten in erster Linie **nach der Bedeutung der Argumente im Zusammenspiel mit dem gegebenen Prädikat und im vorliegenden Satzkontext** annotiert

¹⁷ Die Datei ist herunterladbar unter: http://www.cl.uni-heidelberg.de/projects/GNVN_semanno/material/SemAnno_RoleInventory.pdf

werden. Ein Argument in Subjektposition ist des Öfteren ein *Agent*, jedoch kann es mit einem bestimmten Prädikat und im gegebenen Kontext auch ein *Instrument* oder eine *Time*-Angabe sein. Beispiele:

- [*Agent* **Ich**] kann mit diesem Hammer keine Nägel in die Wand **{schlagen}**.
- [*Instrument* **Der Hammer**] **{schlug}** die Nägel in die Wand. (Der (implizite) Agent ist die Person, die den Nagel mit dem Hammer (Instrument) in die Wand schlägt.)
- [*Time* **Die Stunden**] **{vergingen}** bis Mitternacht ohne dass ich mich gelangweilt hätte.



Abbildung 5: Einbindung von semantischen Argumenten in Prädikat-Instanzen

Abbildung 5 zeigt folgende Eigenschaften der Annotationsaufgabe:

- Freie Ergänzungen wie *Nach einer Umfrage* oder *im Mai* werden nicht als semantische Argumente annotiert.
- Argumente wurden hier nur am Kopfwort markiert – z.B. bei Präpositionalphrasen an der Präposition (siehe Annotationsvarianten in Kapitel 1.1).
- Für das Argument „*Ross for President*“ wurde die semantische Rolle *Theme* annotiert, in einer zusätzlichen *Note*-Annotation wurde jedoch angemerkt, dass diese Annotation nicht die einzig mögliche ist (siehe Kapitel 3.3.5).
- Das Reflexivpronomen *sich* wird als Teil des Prädikats (SemPredPart) annotiert, nicht jedoch als Argument.

3.2 Wie wird annotiert?

Prädikate (und ggf. Argumente) sind als Grundlage für die Rollenannotation **vorannotiert**. Im Rahmen der Annotation semantischer Rollen muss man die Argumente mit ihrem Prädikat durch eine semantische Rolle verbinden.

In den Dateien findet man folgende Vorannotation vor:

- Für **Prädikate**: **SemPred**-Instanzen, ausgezeichnet mit Ihrer **GermaNet-Lesart**.
- Für **Argumente** (falls vorannotiert): **SemArg**-Instanzen werden durch eine generische Annotation ausgezeichnet, mit einer generischen „(SemArg)“-Beschriftung.

Bei der SRL-Annotation werden alle SemArg-Instanzen mit der zugrunde liegenden SemPred-Instanz verbunden, wobei die Relation mit einer Rolle versehen wird. Die SemArg-Instanzen selbst bleiben generisch.

Es wird empfohlen, **zuerst alle SemArg-Instanzen** zu einem gegebenen Prädikat zu **identifizieren**. Sind die SemArg-Instanzen generisch **vorannotiert**, muss in erster Linie nur die Zugehörigkeit dieser Instanzen zum Prädikat geprüft werden. Eventuell fehlt die Annotation für einzelne Argumente. In diesem Fall sollten die Argumente nachannotiert werden.

Das **Anbinden** der identifizierten Argumente ans gegebene Prädikat erfolgt in folgenden Schritten:

1. Man **klickt doppelt** auf die vorhandene **SemPred**-Instanz und aktiviert dadurch das rechte Annotationsfeld.
2. Man wählt nacheinander die entsprechenden **SemArg**-Instanzen **für das Feature „RoleSet“**:
 1. Aus der Drop-down-Liste wird eine der **angegebenen semantischen Rollen** ausgewählt.
 2. Gegebenenfalls gibt es eine Vorauswahl an semantischen Rollen, wobei die ersten Rollen in der Liste als zum Prädikat passende Rollen betrachtet werden. Diese werden visuell von den weiteren Rollen getrennt. Falls dies der Fall ist, kann man die oben platzierten Rollen als beste Kandidaten für die Auswahl der Rollen betrachten. Falls man unter diesen oberen Rollen keine passende findet, wählt man eine aus den unteren Rängen der Liste aus.
3. Erscheint keine der angegebenen Rollen als angemessen, sollte man sich für die **am besten passende** Rolle entscheiden. Zusätzlich kann man in einem Kommentar (siehe Kapitel 3.3.5) angeben, woran die Schwierigkeit liegt.
4. Hält man den Mauszeiger über dem Rollennamen in der Drop-down-Liste der Rollen, erscheint eine kurze **Beschreibung** zur gegebenen Rolle als Entscheidungshilfe in Form eines Tooltips (siehe Abbildung 6).
5. Klicken Sie auf „**Add**“. Dadurch wird die Rolle übernommen, und rechts davon erscheint ein weiteres, **orange** markiertes Textfeld mit der Überschrift „**<Select to fill>**“. Klicken Sie umgehend auf die entsprechende **SemArg**-Instanz in der Annotationsoberfläche doppelt. Die verbindende Kante mit der angegebenen semantischen Rolle erscheint zwischen dem Prädikat und seinem Argument.
6. In dieser Weise werden alle weiteren Argumente nacheinander abgebunden.

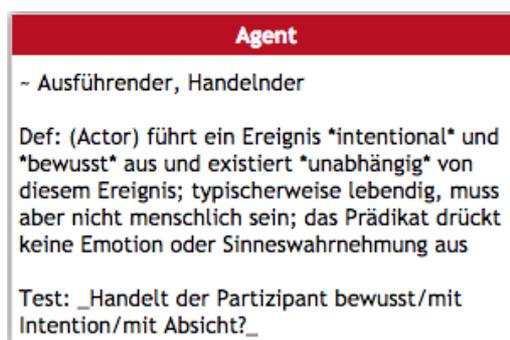


Abbildung 6: Tooltip-Beispiel für die semantische Rolle "Agent"

Änderungen an den bereits annotierten Rollen kann man nach Aktivierung der SemPred-Instanz vornehmen. Anschließend wählt man in der rechten Annotationsoberfläche die zu ändernde Rolle aus. Das Textfeld zur Rolle wird daraufhin orange markiert.

- Um die aktuelle Rolle durch ein anderes Argument zu besetzen, doppelklickt man einfach anschließend auf das andere SemArg. Die neue Wahl wird übernommen.
- Zum Löschen der ganzen Rollenverbindung klickt man auf „**Del**“.

- Möchte man den **Rollennamen ändern**, wählt man von der Drop-down-Menüliste den passenden Rollennamen aus.

Eventuell muss man **außerhalb des aktuellen Satzes** nach einem geeigneten Argument suchen; dieses sollten Sie auch annotieren.

Nicht vorannotierte, aber im Text vorhandene Argumente sollten als SemArg-Instanz nachannotiert werden. Argumente, die im gegebenen Kontext nicht explizit erscheinen, werden nicht annotiert.

3.3 Schwierige Fälle

3.3.1 Mehrfachannotation bei mehrmaliger Referenz

Gehört ein Zielelement (Prädikat oder Argument) zu mehreren Prädikat-Argument-Strukturen, sollte diese mehrfache Zugehörigkeit auch durch **Mehrfachannotation** widerspiegelt werden.

3.3.1.1 Überlappungen in Prädikat-Argument-Strukturen

Gehört ein **Argument** zu mehreren Prädikat-Argument-Strukturen, wird das Argument **mehrfach** als Rollenträger ausgezeichnet.

Beispiel: *Fritz sitzt im Garten und liest.* (siehe auch Abbildung 4)

Mit Annotation: $[_{Agent(sitzen)} [_{Agent(lesen)} \mathbf{Fritz}]] \{sitzt\} [_{Location(sitzen)} \text{im Garten}] \text{ und } \{liest\}$.

Fritz ist der Agent für *sitzen* und für *lesen*, sodass dieses Subjekt zweimal annotiert wird.

im Garten ist eine Lokation (Location) für *sitzen* – dieses Argument wird für *lesen* nicht annotiert, da lesen die Angabe des Ortes nicht benötigt.

3.3.1.2 Elidiertes Prädikat

Wenn ein Prädikat im Satz weggelassen (elidiert) wird, aber mindestens ein Argument des Prädikats, und ein dazu passendes (in der Regel das gleiche) Prädikat im Kontext vorhanden sind, wird dieses overte Prädikat für das elidierte annotiert. Das overte Prädikat besitzt typischerweise auch eine eigene Argumentstruktur, und somit wird es so oft annotiert, wie sich Argumentstrukturen auf dieses Prädikat beziehen.

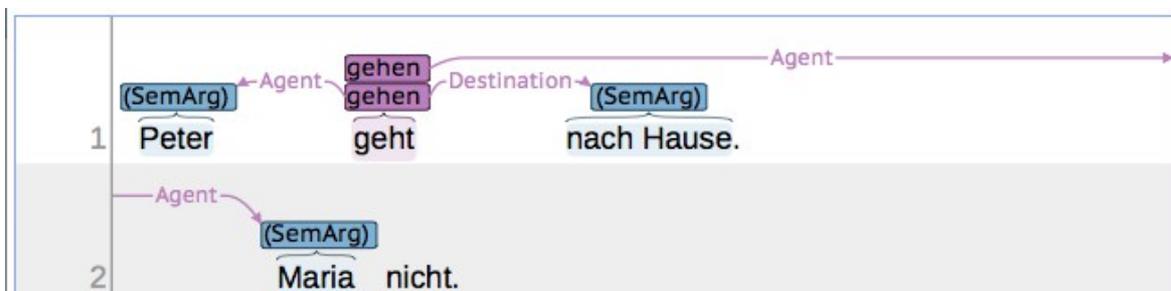


Abbildung 7: Weggelassene, aber erschließbare Prädikate werden je Prädikat-Argument-Struktur annotiert.

Das Verb *gehen* wird im Beispiel der Abbildung 7 einmal für den Rollenträger *Peter* und *nach Hause* als SemPred annotiert, und noch einmal für *Maria*. Beachten Sie, dass *nach Hause* hier nicht doppelt annotiert werden muss, ausschließlich das Prädikat wird ergänzend markiert.

3.3.1.3 Multiple Lesarten

Wird ein **Prädikat** durch **mehrere Lesarten** ausgezeichnet, d.h. ist das Verb bereits mit mehreren Lesarten annotiert, sollten eventuell identische Rollenträger für alle Lesarten separat (mehrfach) annotiert werden. Diese Annotation ist in den zugrunde liegenden Dateien womöglich bereits enthalten. Andernfalls sollte die Annotation durch passende Mehrfachannotation der Argumente ergänzt werden.

3.3.2 Nicht lokal realisierte Argumente

Falls ein **Argument nicht im selben Satz (lokal)** erwähnt wird, z.B. wenn das Prädikat in Passiv steht, dann muss man das fehlende Argument ggf. im Textkontext annotieren.

Beispiele:

- *Man hört* [_{Agent} *ihn*] *schon wieder* {reden}. *Und wieder* [_{Topic} **über unwichtige Sachen**]. – Worüber der Agent redet, steht im nächsten Satz. Solche nicht-lokalen, aber im Kontext explizit erwähnten Argumente werden annotiert.
- *Fritz fährt nach* [_{Destination} **London**]. [_{Agent} *Er*] {kommt} [_{Time} *um 15 Uhr*] {an}.
- [_{Experiencer} *Finanzminister Paleokrassas*] *kann* {sich} {freuen} : [_{Theme} **Gut eine Milliarde Mark brachte ihm allein der Verkauf des einst von den Sozialisten verstaatlichten Unternehmens Herakles ein, einer der größten Zementhersteller Europas**].
- Gegenbeispiel: [_{Theme} *Alle Motoren diesen Typs*] *wurden* {verkauft}. – Wird nicht einmal im Kontextfenster erwähnt, wer (Agent) die Motoren verkauft hat, dann ist nur die Theme-Rolle explizit vorhanden. Die nicht-overtete Agent-Rolle muss (kann) nicht annotiert werden.

Gegebenenfalls werden nicht alle (realisierten) Argumente als SemArg-Instanz vorannotiert und müssen von Hand nachannotiert werden. Dabei soll man je nach Aufgabenstellung entweder die ganze Phrase (Aufgabenvariante 3.a), siehe Kapitel 1.1), oder ausschließlich das Kopfelement einer Phrase (Variante 3.b)) auszeichnen.

3.3.3 Annotationsvarianten für Argumente

Argumente können ein einziges Wort, eine Phrase, bis zu einem ganzen Satz abdecken. Die Annotation der Argumente kann also die ganze zugehörige (ggf. diskontinuierliche) Wortsequenz umfassen, bei einer vorhandenen Dependenzannotation der Sätze kann jedoch auch nur das Kopfwort des jeweiligen Arguments annotiert werden, wobei durch die Dependenzstruktur sichergestellt wird, dass die ganze Wortsequenz für ein Argument transparent und nachvollziehbar bleibt.

3.3.3.1 Annotation ganzer Phrasen

Bei der Auszeichnung aller zugehörigen Wörter zu einer SemArg-Instanz wird eine Annotation über die ganze Wortsequenz gespannt. Man geht hier bei der Auswahl der Wortsequenz in der Regel intuitiv vor, wobei einige Regeln konsequent verfolgt werden sollten:

- Ein anschließendes Satzzeichen (Komma, Punkt oder schließende Klammer, evtl. Anführungszeichen) soll nicht in die Annotation mit aufgenommen werden. Beispiel: [*Viele*] {meinen}, [**daß Perot ... laufen würde**]. – Der abschließende Punkt gehört nicht zur SemArg-Instanz.

- Zweigliedrige Satzzeichen wie Klammern sollte man vollständig in die Annotation einbeziehen: **[Electronic Data Services (EDS)]**
- Diskontinuierliche Einheiten sollten ggf. nachannotiert werden. Bsp. (Topikalisierung der Argumentteile): **[Eier] hat sie [nur drei] gekauft.**
 - Die Annotation erfolgt mithilfe der SemArgPart-Annotationsebene parallel zur Annotation der mehrteiligen Prädikate, zur Beschreibung des Vorgangs siehe Kapitel 2.4.1.

3.3.3.2 Annotation des Kopfelements

Falls eine Dependenzannotation vorliegt, kann man die Annotation der Argumente auf die Annotation deren Kopfwörter beschränken. Dabei folgt man der Dependenzstruktur der jeweiligen Wortsequenz für das aktuelle Argument. Gegebenenfalls können durch unterschiedliche Dependenzstrukturen unterschiedliche Kopfelemente für eine Phrase ausgezeichnet werden.

Wenn man sich auf die Auszeichnung des jeweiligen Kopfelements ohne eine unterliegende Dependenzannotation beschränkt, kann man einer konsequenten Richtlinie folgen, welche die Auszeichnung möglichst relevanter Wörter erzielt. Empfohlen werden folgende Regeln:

- In einer Phrase, die eine Einworteinheit ist, wie die meisten Pronomina (**[er]**, **[uns]**), Determinativa ohne nominalen Kopf (**[dieser] wird auch noch einbezogen**) wird dieses einzelne Wort annotiert.
- In einer Nominalphrase wird das Kopfnomen ausgezeichnet: *der kleine **[Prinz]**; vorsichtige **[Maßnahmen]** mit niedrigen Kosten.*
- Bei Präpositionalphrasen wird die Präposition als Kopf des semantischen Arguments betrachtet: **[nach] Hause, [für] Clinton**
- Bei mehrteiligen Namen das wichtigste Element (z.B. Nachnamen) oder das letzte Wort: *Eike **[Fischer]**; General **[Motors]**; Electronic Data Services (**[EDS]**)*
- Bei eingebetteten Sätzen als Argument wird das Element, das vom annotierten Prädikat direkt eingebettet ist, annotiert:
 - das Korrelat-es oder das **Präpositionaladverb**, falls vorhanden: *Ich wusste **[es]** schon, dass ...; Denke gar nicht **[daran]**, dass...*
 - das **finite Verb** des eingebetteten Satzes (nicht die Konjunktion): *Er sagte laut und deutlich, daß er nicht **[mitgeht]**; / ..., daß er nicht mitgehen **[will]**; ich habe versprochen dir zu **[helfen]***

Anderweitige Zieleinheiten und ihre Auszeichnung müssen jeweils explizit vereinbart werden.

3.3.4 Mehrteilige semantische Prädikate und Argumente

Wenn **Prädikate aus mehreren Wörtern** bestehen und diese Teile nicht als zusammengehörig ausgezeichnet sind, werden diese wie in Kapitel 3.3.4 für syntaktische Prädikate erklärt **nachannotiert**.

Auch semantische Argumente (SemArg) können **mehrteilig** sein oder **Diskontinuitäten** enthalten. Einige Beispiele:

- Einbettungen innerhalb der Wortsequenz des Arguments: Bsp.: **[Ich]**, sagte er leise, **[ich] wusste nicht davon.**
- Topikalisierung der Argumentteile: Bsp.: **[Eier] hat sie [nur drei] gekauft.**

- Argumentteile in Extraposition: Bsp.: **[Geschäftemachen] ist seine Welt [und nicht die Politik].**

Diese Fälle werden durch eine vorliegende Abhängigkeitsstruktur in der Regel als zusammengehörig gekennzeichnet. Somit kann diese Art der Diskontinuität in der Annotationsvariante Kopfelement-Annotation unannotiert bleiben. Ob die Abhängigkeitsstruktur die Diskontinuität erfasst hat, kann man erkennen, wenn man mit der Maus auf den POS-Tag des Elements zeigt. Im erscheinenden kleinen Fenster werden unter „Yield of relation“ alle vom gegebenen Element abhängigen Satzteile angezeigt. Drei Punkte machen eine Diskontinuität sichtbar wie in Abbildung 8 gezeigt wird.

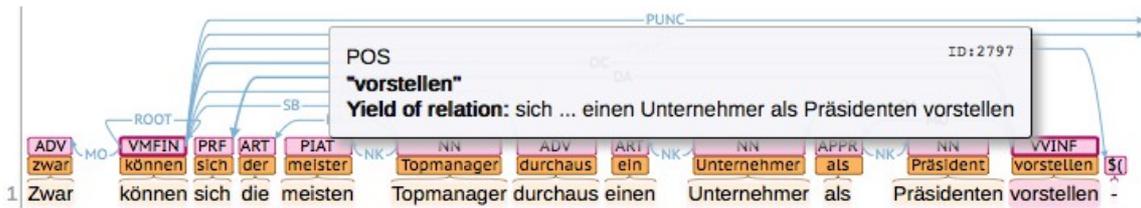


Abbildung 8: Fährt man über die POS-Tag-Annotation mit der Maus, erscheint die Textsequenz, die durch die Abhängigkeitsbaumstruktur abgedeckt ist. Eine Diskontinuität dieser Sequenz wird durch „...“ markiert.

3.3.5 Kommentar-Annotation: Note-Layer

Erscheint eine Annotation als nicht sicher oder möchte man eine Anmerkung zu einer Annotation hinterlassen, wird dafür die Annotationsebene *Note* zur Verfügung gestellt.

Note gehört zum span-Annotationstyp (siehe Kapitel 4.3.1). Man kann das fragliche oder diskussionsbedürftige Token markieren und einen beliebigen Text mit der Beschreibung bzw. Fragestellung formulieren.

4 Verwendung von WebAnno

Das Tool wird mit online-Guidelines zur aktuellen Version unterstützt, die Guidelines sind auf der offiziellen Webseite des Tools unter <https://webanno.github.io/webanno/documentation/> zu finden.

Die folgende Beschreibung setzt voraus, dass die dort vorhandenen Informationen bekannt sind, und fokussiert auf die spezifischen Gegebenheiten der Annotationsaufgaben WSD und SRL.

4.1 Allgemeine Hinweise

Das Tool wurde mit den Browsern Safari und Chrome getestet. Es wurden auch gute Erfahrungen mit Mozilla Firefox gemacht.

Jeder Schritt der Annotation und jede Änderung wird im Hintergrund sofort gespeichert, sodass man sich jederzeit ausloggen kann, wenn man die Annotation zeitweise abrechnen muss. Will man die Annotationsarbeit wieder aufnehmen, kann man die Annotation nach erneutem Einloggen von diesem letzten Stand aus weiterführen.

4.2 Einstellungen des Annotationsprozesses (Settings)

Die Annotationsoberfläche und die Annotationsdetails können durch Anklicken des Knopfs „Settings“ in der oberen Leiste der Annotationsoberfläche vorgenommen werden. Die hier angegebenen Einstellungen werden für alle Dateien, die zum aktuellen Projekt gehören, übernommen. Man kann die Einstellungen vor und auch während des Annotationsvorgangs jederzeit ändern.

Es wird empfohlen, unter *Annotation Layers* alle anderen Annotationsebenen als die der Hilfs- bzw. Zielannotation durch das Entfernen des Hakens vor dem Namen der Annotationsebene auszublenden. Relationale Annotationsebenen, die mit einer anderen Ebene verknüpft sind, sollten mit diesen zusammen ein- oder ausgeblendet sein.

- Das Layer *SemPredPart* ist eine Relation, die zwei Instanzen des Layers *SemPred* verbindet. Wird das Layer *SemPred* ausgeblendet, können die Annotationen zu *SemPredPart* auch nicht angezeigt werden.
- Das Layer *Dependency* kann nur bei angezeigtem *POS*-Layer eingeblendet werden.

Wenn die Annotation lemmaweise erfolgt, und das aktuelle Prädikat im Satz in einem Kontext von mehreren Sätzen vor und/oder nach dem Satz mit dem Zielprädikat gezeigt wird, soll diese Kontextfenstergröße für die Anzahl der aktuell angezeigten Sätze (*Number of Sentences*) eingestellt werden.

- In wenigen Fällen kommt das Zielprädikat in den ersten Sätzen des Korpus vor. In diesem Fall hat das erste Zielprädikat ggf. weniger als die vorgesehenen Anzahl an Sätzen als linken Kontext.
- Bei lemmaweiser Annotation mit Kontextfenster ist es außerdem sinnvoll, das automatische Weiterscrollen des angezeigten Textbereiches während des Annotationsvorgangs (*Auto-scroll document while annotating*) zu deaktivieren.

Die Einstellung „*Use the same color for all tags in a layer*“ stellt sicher, dass die Annotationen zur selben Annotationsebene durch die gleiche Farbe markiert werden. Bei einem Neuladen der Datei

oder nach einem erneuten Einloggen können die verwendeten Farben unterschiedlich ausfallen, eine Unterscheidung der Farben unterschiedlicher Annotationsebenen wird jedoch immer gewährleistet.

4.3 Annotation verschiedener Annotationstypen in WebAnno

Für die vorliegende Annotationsstudie werden mehrere Annotationsebenen annotiert. Diese gehören auch unterschiedlichen Typen an: *span*, *slot* und *relation*. Im Folgenden wird der Umgang mit diesen verschiedenen Ebenentypen vorgestellt.

4.3.1 „span“-Annotation

Die „span“-Annotation, die Annotation von Zeichen- bzw. Tokenfolgen, wird in dieser Annotationsstudie für die Layers *SemPred* und *SemArg* verwendet.

Für ein Wort (Token): Die Auswahl eines Wortes erfolgt durch Doppelklick auf das Wort mit der linken Maustaste automatisch. Für die Annotation **mehrerer Wörter** als eine Annotationsinstanz wählt man mit der Maus bei gedrückter linker Maustaste mehrere, nebeneinander stehende Wörter aus.

Ist man sich bezüglich des Werts eines Features nicht sicher, oder möchte man eine unterspezifizierte Annotation erstellen, können die Featurewerte ggf. leer gelassen werden. Wird jedoch eine vorhandene Liste mit nicht erweiterbaren Werten zur Annotation für das Feature zur Verfügung gestellt, dann ist der Annotator gezwungen, einen der angebotenen Werte zu wählen.

Möchte man eine **bestehende Annotation ändern oder löschen**, klickt man auf die Annotationsinstanz mit der linken Maustaste doppelt. In der rechten Annotationsoberfläche erscheinen die annotierten Eigenschaften der Instanz, an denen man jetzt ändern kann. Die ganze Instanz kann man mit Anklicken des Knopfs *Delete* löschen. Es ist zu beachten, dass das Löschen einer Span-Annotation, mit der andere Relationsannotationen verknüpft sind, auch das Löschen der Relationsannotationen nach sich zieht.

4.3.2 „relation“-Annotation

Als Relation zwischen zwei Span-Annotationen werden die Layers *SemPredPart* und *SemArgPart* definiert.

Die Relation wird durch eine gezogene Kante zwischen zwei Instanzen der Grundlagenebene für das jeweilige Relation-Layer repräsentiert. Die Grundlagenebenen sind die folgenden:

- für *SemPredPart*: *SemPred* – Man annotiert die einzelnen, diskontinuierlichen Bestandteile eines Prädikats durch einzelne *SemPred*-Instanzen, und verbindet sie mit einer Kante des Layers *SemPredPart*.
- für *SemArgPart*: *SemArg* – Sind Argumente eines semantischen Prädikats diskontinuierlich, werden diese Einheiten durch *SemArgPart* verknüpft.

Generell gilt, dass die Instanzen, die durch eine „*Part*“-Kante verbunden werden, eine **hierarchische** Beziehung abbilden. Die Richtung der Kante soll von der übergeordneten Instanz zur untergeordneten Instanz zeigen. Weitere Annotationen auf diesem Instanzkomplex sollen auf der übergeordneten Instanz aufgesetzt werden.

Die Änderung oder das Entfernen der Relation-Instanzen erfolgt nach dem Anklicken der Kante, was die Annotationsdetails zu der gegebenen Instanz aktiviert. Featurewerte können hiermit

gegebenenfalls auch geändert werden. Im rechten Kasten „Actions“ erscheinen die Knöpfe „Reverse“ und „Delete“. Ersterer dient der Umkehrung der Kantenrichtung, falls notwendig. Die Betätigung des Knopfes „Delete“ entfernt die ganze Relation.

4.3.3 „slot“-Annotation für semantische Rollen

Dieser Annotationstyp wird bei der Annotation semantischer Rollen durch das Einbinden des Layers *SemArg* in das Layer *SemPred* realisiert.

Der Annotationsvorgang ist Teil der Annotation des Layers *SemPred*, wobei das Feature *RoleSet* mit den möglichen *SemArg*-Instanzen gefüllt wird.

Als Erstes definiert man die semantische Rolle der aktuell betrachteten *SemArg*-Instanz und wählt eine der angebotenen Werte in der erscheinenden Liste aus. Die empfohlenen Rollen sind in der Auswahlliste oben platziert und von den übrigen Rollen visuell (durch Fettdruck) getrennt. Diese sind bevorzugt einzusetzen, jedoch kann sich auch eine weitere Rolle als am besten geeignet erweisen. Die Verwendung der Default-Rolle "XXX" sollte man nur in Ausnahmefällen nutzen, und möglichst eine Note-Annotation dazu erstellen.

Die Rollenauswahl wird durch das Anklicken dem Knopf „Add“ bestätigt, wobei ein zusätzliches Feld mit orangener Farbe für den Inhalt des Arguments erscheint. Klickt man die betrachtete *SemArg*-Instanz an, wird diese in das orangefarbene Feld eingefügt, und somit die Annotation für die aktuelle semantische Rolle komplettiert. Gleichzeitig erscheint ein neues Textfeld für weitere Rollen, sodass man Schritt-für-Schritt die ganze Prädikat-Argument-Struktur annotieren kann.

Gegebenenfalls kommen nicht alle möglichen Rollen für ein Prädikat im aktuellen Satz vor. Findet man im Kontext des Satzes ein geeignetes Argument, sollte man dieses für das gegebene Prädikat ggf. annotieren. Wird das Argument im Satz oder Kontextumgebung nicht explizit ausgedrückt (implizite Argumente), wird das Annotieren der entsprechenden Rolle unterlassen.

Das Ändern des Inhalts einer Slot-Annotation kann man nach der Aktivierung der Annotation durch Anklicken vornehmen. Der Text wird orange angezeigt. Klickt man dann auf die korrekte *SemArg*-Instanz doppelt, wird daraufhin der alte Inhalt durch den der neu gewählten Instanz ersetzt. (Die alte *SemArg*-Instanz bleibt erhalten, diese muss man zusätzlich löschen.)

Man kann die ganze Rolle löschen, wenn man nach dem Aktivieren dieser Rolle (orangene Färbung des Textes) den Knopf „Del“ drückt.

5 Anhang

5.1 Übersicht zu GermaNet-Lesarten

Ein Auszug relevanter Informationen für GermaNet-9.0-Lesarten wird während des Annotationsvorgangs im Annotationstool als Tooltips zur Verfügung gestellt. Ein Tooltip-Beispiel für die Lesart *zustimmen-1* ist in Abbildung 9 dargestellt.

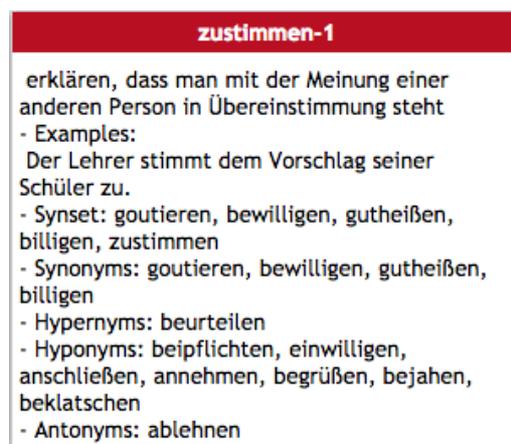


Abbildung 9: Tooltip mit Beschreibung der Lesart "zustimmen-1". Das Tooltip wird beim Annotationsprozess bei Bedarf auf der Annotationsoberfläche von WebAnno eingeblendet.

5.2 Übersicht der semantischen Rollen

Das anliegende Rolleninventar¹⁸ enthält die Definition und beispielhafte Verwendungen für die gegebenen semantischen Rollen, die für diese Annotationsaufgabe ausgearbeitet wurden.

Die Rollen sind in größeren Gruppen *Actor*, *Undergoer*, *Place*, *Time* und *Circumstance* geordnet. Eine Übersicht der Hierarchie ist in Abbildung 10 dargestellt.

¹⁸ Rolleninventar: http://www.cl.uni-heidelberg.de/projects/GNVN_semanno/material/SemAnno_RoleInventory.pdf

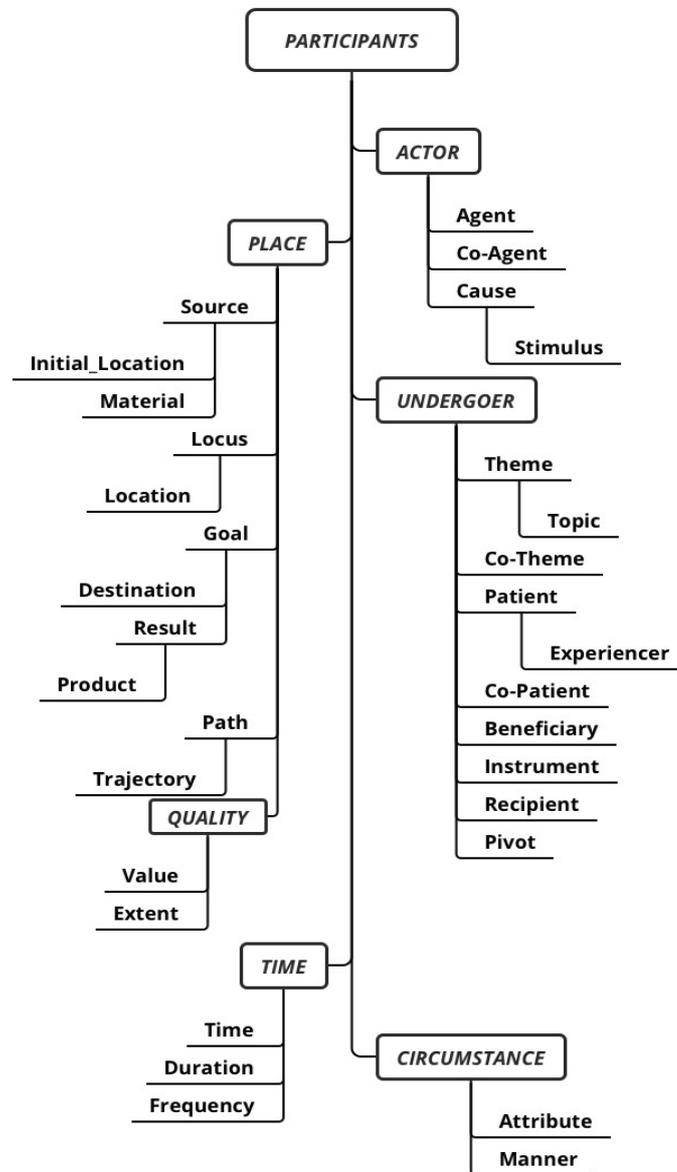


Abbildung 10: Hierarchie der semantischen Rollen wie in [Mújdricza-Maydt et al. 2016] dargestellt.

6 Literaturverzeichnis

[Burchardt et al. 2009] Aljosha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Padó, Manfred Pinkal (2009): FrameNet for the semantic analysis of German: annotation, representation, and automation, Kapitel 8. Hrsg.: Hans C. Boas: *Multilingual FrameNets in Computational Lexicography: Methods and Applications*. Mouton de Guyter.

[Mújdricza-Maydt et al. 2016] Éva Mújdricza-Maydt, Silvana Hartmann, Iryna Gurevych, Anette Frank (2016): Combining Semantic Annotation of Word Sense & Semantic Roles: A Novel Annotation Scheme for VerbNet Roles on German Language Data. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC2016)*. Portorož, Slovenia, S. 3031-3038.

[Pittner und Berman 2006] Karin Pittner, Judith Berman (2006): *Deutsche Syntax. Ein Arbeitsbuch*. Tübingen: Narr.