

Guidelines für die Annotation von Diskursrelationen in LeiKo

Sarah Jablotschkin

Stand: 09.05.2023

Grundlage: Guidelines der Penn Discourse Treebank (PDTB), Versionen 2.0 (Prasad *et al.*, 2007) und 3.0 (Webber *et al.*, 2019)

Ergänzendes Dokument: Senses nach den PDTB-Guidelines 3.0 (Jablotschkin, 2023)

Die vorliegenden Guidelines übertragen zum einen die PDTB-Guidelines ins Deutsche. Zum anderen beinhalten sie Spezifikationen für bestimmte Strukturen, die häufig in vereinfachten Texten des Deutschen auftreten, wie z.B. längere Absätze wiedergegebener Rede (vgl. S. 8-9, dieses Dokument) oder Subtexte in Form von Worterklärungen (vgl. S. 9, dieses Dokument).

Sofern nicht anders angegeben, stammen alle Sprachbeispiele aus LeiKo 1.6 (Jablotschkin und Zinsmeister, 2020).

Inhaltsverzeichnis

1 Vorbereitung.....	1
1.1 Anlegen der Ordnerstruktur für die Annotation.....	1
1.2 Herunterladen des Annotationstools.....	1
1.3 Beim Öffnen des Annotators.....	2
2 Durchführung der Annotation.....	2
2.1 Annotationsschritte.....	2
2.2 Grundannahmen und allgemeine Konventionen.....	5
2.2.1 Konnektoren.....	5
2.2.2 Formen von Argumenten.....	6
2.2.3 Lokalisierung von Argumenten.....	7
2.2.4 Ausdehnung von Argumenten.....	8
Anhang.....	11
Quellen.....	12
Stichwortverzeichnis.....	13

1 Vorbereitung

1.1 Anlegen der Ordnerstruktur für die Annotation

1. Lege einen Ordner an und gib ihm z.B. den Namen *Diskursrelationen*.
2. Lade dir das Verzeichnis *LeiKo_core_text.zip* herunter, speichere es in dem Ordner *Diskursrelationen* ab und entpacke es.
3. Erstelle im Ordner *Diskursrelationen* einen Ordner, in dem deine Annotationen gespeichert werden. Nenne ihn z.B. *Annotationen*.
4. Erstelle im Ordner *Diskursrelationen* einen weiteren Ordner für deine Kommentare im Zuge der Annotation. Nenne ihn z.B. *Kommentare*.

Deine Ordnerstruktur sollte anschließend wie folgt aussehen:

Diskursrelationen

- LeiKo_core_text
 - NDR
 - tazleicht
 - SR
 - nachrichteleicht
 - standard_taz
- Annotationen
- Kommentare

1.2 Herunterladen des Annotationstools

1. Folge den Installationsanweisungen unter <https://www.seas.upenn.edu/~pdtb/pdtb3-tools.shtml> im Abschnitt „Prerequisites and Download“ (siehe Abb. 1).

Home
People
Publications
Tools
Bibliography
Events

PDTB 3.0 TOOLS AND DOC

THE PDTB Annotator

The PDTB Annotator is a tool for annotating and adjudicating discourse relations using the PDTB Annotation F simply needs to organize their text files into the correct directory structures before loading them into the Annota

Prerequisites and Download

The latest version of the PDTB Annotator is Annotator version 4. The Annotator runs locally on the user's comp installed on their computer.

Check if you have Java installed [here](#).

Once Java is running on your computer, download the Annotator using the following link:

[PDTB Annotator.](#)

A config file named Options.cfg is also necessary to run the tool:

[Download Options.cfg.](#)

Abbildung 1: Lade die grün eingekreisten Dateien herunter.

1.3 Beim Öffnen des Annotators

1. RawRoot: Navigiere in das Verzeichnis *LeiKo_core_text*.
2. WorkRoot: Navigiere in das Verzeichnis *Annotationen*.
3. WorkComment: Navigiere in das Verzeichnis *Kommentare*.
4. Die übrigen Felder können erst einmal leer bleiben. Klicke „OK“.

2 Durchführung der Annotation

2.1 Annotationsschritte

1. **Einen Text im Annotator laden:** Oben links im Dropdown-Menü das entsprechende Subkorpus auswählen. Rechts daneben den zu annotierenden Text auswählen und „Load“ klicken.
2. **Eine neue Relation hinzufügen:** „Add new Relation“ anklicken.
3. **Die Elemente einer expliziten Diskursrelation im Text markieren:** Einen Konnektor mit der Maus markieren und auf +Conn Span klicken. Anschließend die beiden Argumente des Konnektors markieren (in der Regel sind diese satzförmig) und auf Arg1 Span bzw. Arg2 Span klicken. Ist der Konnektor eine subordinierende Konjunktion, ist Arg2 immer der untergeordnete Satz und Arg1 der Matrixsatz. Bei allen anderen Konnektoren ist das erste Argument Arg1 und das zweite Argument Arg2. **Technischer Hinweis:** Möchtest du diskontinuierliche Textspannen markieren, halte die Strg-Taste gedrückt.
4. **Den Relationstypen angeben:** Im Dropdown-Menü „Relation Type“ den Wert „Explicit“ auswählen.
5. **Einen Sense zuweisen:** Im Feld SClass1A diejenige Bedeutungskategorie auswählen, die am ehesten der Relation entspricht. Falls mehrere Lesarten möglich sind, kann im Feld SClass1B ein zweiter Sense angegeben werden.
6. **Die Annotation abspeichern:** „Save Relation“ anklicken.
7. Nachdem du Schritt 6 ausgeführt hast, sollte die Annotation bei dir ähnlich wie in Abb. 2 aussehen. Schritte 1-6 ausführen, bis alle expliziten Diskursrelationen markiert sind.

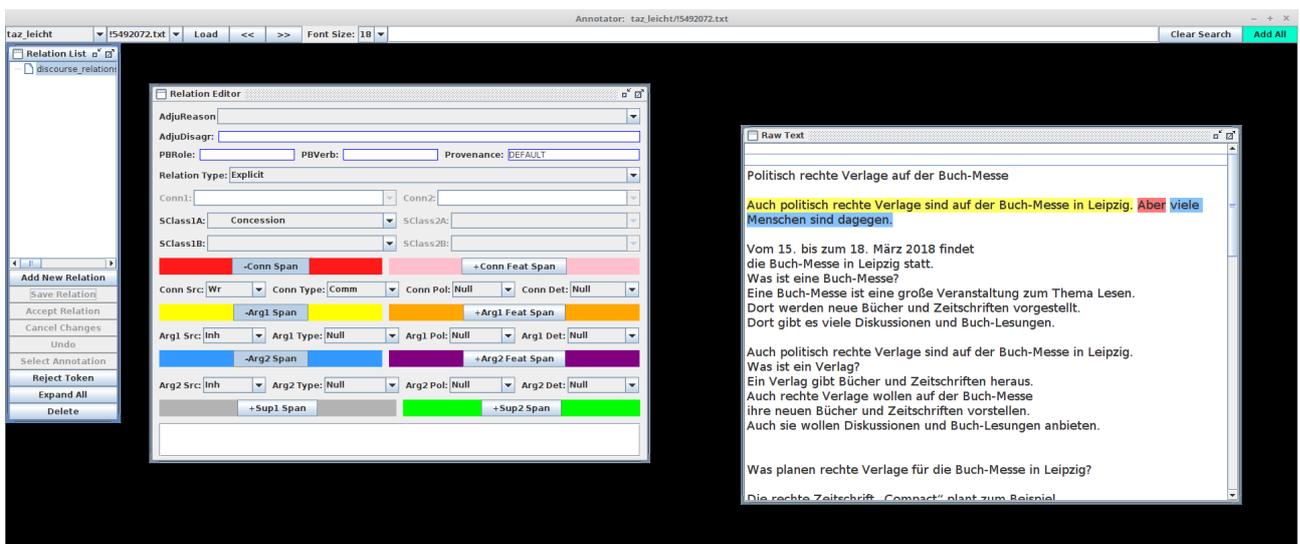


Abbildung 2: Beispiel für eine annotierte Diskursrelation

8. **Weitere Typen von sprachlich markierten Diskursrelationen annotieren:** Gehe den Text erneut durch und konzentriere dich in diesem Durchgang auf die Relationen *Hypophora*, *AltLex* und *AltLexC*.
- Hypophora: Die Argumente stellen ein Frage-Antwort-Paar dar (vgl. Bsp. (w)).
 - AltLex: Die Relation ist nicht durch einen Konnektor markiert. Einen Konnektor einzufügen, wäre aber redundant, weil anderes lexikalisches Material vorhanden ist, das die Relation ausdrückt, z.B. bestimmte Kombinationen aus Anapher und Verb (vgl. (a)), sogenannte Shell Nouns (vgl. (b)) oder andere Ausdrücke mit anaphorischer Komponente (vgl. (c) und (d)). Auch exakte Wortwiederholungen wie in (e) können manchmal bestimmte Relationen vermitteln.
- (a) *Robert Habeck findet das linke Thema „Umverteilung“ wichtig für die Politik der Grünen. **Das bedeutet: Ärmere Menschen sollen mehr Geld bekommen. Reichere Menschen sollen mehr Geld abgeben.***
- (b) *Kranke sterben oft, weil *Medikamente nicht wirken*. **Ein Grund sind bestimmte Keime.***
- (c) *In Hamburg war ein Unfall. **Bei diesem Unfall ist ein Auto aus einem Parkhaus gestürzt.***
- (d) *Bei den Olympischen Spielen sind auch 5 Sportler aus Norddeutschland. **Der bekannteste von den 5 Sportlern ist Arnd Pfeiffer.***
- (e) *Einige Hühner liegen tot herum. **Einige Hühner haben offene Wunden.***
- AltLexC: Zur Markierung der Relation trägt nicht allein lexikalisches Material, sondern auch (oder ausschließlich) die Syntax bei. AltLexC-Relationen sind „lexico-syntactic constructions that are unambiguous signals of particular discourse relations“, das heißt, in der Regel gibt es feste Verbindungen zwischen lexiko-syntaktischer Konstruktion und dem dazugehörigen Sense. Gerade wenn die Markierung durch den Satzbau geschieht, lässt sich keine eindeutige textuelle Einheit als konnektorähnliche Spanne markieren. Behelfsmäßig markieren wir in solchen Fällen den gesamten (Teil-)Satz (vgl. (f)).
- (f) ***Würden Rechte die Macht übernehmen**, wäre nach Auffassung der Gruppe keine freie Debatte mehr möglich. (Inversion zur Markierung einer Relation mit dem Sense Contingency.Condition.Arg2-as-cond) (Beispiel aus Jablotschkin und Zinsmeister, 2020b)*
- (g) *Würden Rechte die Macht übernehmen, wäre nach Auffassung der Gruppe keine freie Debatte mehr möglich. **Daraus folge, „dass wir mit unserem normativ-politischen Bekenntnis zur Meinungsfreiheit aus strategischen Gründen vorläufig und in einem bestimmten Punkt – rechte sexistische Ideologie – brechen müssen“.** (*daraus folge+dass*-Satz zur Markierung einer Relation mit dem Sense Contingency.Cause+Belief.Result+Belief) (Beispiel aus Jablotschkin und Zinsmeister, 2020b)*
- (h) *Ich war so müde, **dass ich im Sitzen eingeschlafen bin.** (*so+Adjektiv/Adverb+dass*-Satz zur Markierung einer Relation mit dem Sense Contingency.Cause.Result; konstruiertes Beispiel)*
- (i) *Ich bin zu hungrig, **um mich konzentrieren zu können.** (*zu+Adjektiv/Adverb+um-zu*-Satz zur Markierung einer Relation mit dem Sense Contingency.Cause.negResult; konstruiertes Beispiel)*
9. **Implizite Diskursrelationen annotieren:** Nachdem du alle sprachlich markierten Diskursrelationen im Text annotiert hast, konzentriere dich auf diejenigen benachbarten Sätze, zwi-

schen denen noch keine Diskursrelation annotiert ist. Überprüfe, ob du dort einen Konnektor oder einen AltLex-Ausdruck einfügen könntest, der die Bedeutung nicht verändern, sondern lediglich paraphrasieren würde. Markiere Arg1 und Arg2, wähle den Relation Type „Implicit“ aus und weise einen Sense zu. Im Feld Conn1 kannst du denjenigen sprachlichen Ausdruck eintippen, der deiner Meinung nach die Bedeutung der impliziten Relation am besten wiedergeben würde.¹ Falls du dich zwischen mehreren möglichen Konnektoren, die eindeutig verschiedene Relationen ausdrücken, nicht entscheiden kannst, gib im Feld Conn2 einen weiteren Konnektor ein. Die Annotation einer Diskursrelation des Typs Implicit sollte wie in Abb. 3 aussehen. In Ausnahmefällen, also wenn du das Gefühl hast, dass sonst eine wichtige Lesart verloren geht, darf auch da eine implizite Relation annotiert werden, wo bereits eine explizite Relation ist. **Damit aber bitte eher zurückhaltend umgehen. Beachte bitte auch, dass Konnektoren häufig ambig sind und du beispielsweise auch eine temporale Relation annotieren kannst, wenn dort und steht.**

10. **EntRel und NoRel annotieren:** Wenn du auch alle impliziten Relationen annotiert hast, entscheidest du bei jeweils benachbarten Argumenten, zwischen denen noch keine Relation annotiert ist, ob auf jeweils (einen) gemeinsame(n) Referenten verwiesen wird. Dann gibst du beim Relation Type *EntRel* an (vgl. (j) und (k)). Falls in kein gemeinsamer Referent vorhanden ist, annotierst du *NoRel* (vgl. (l)). NoRel-Relationen treten häufig zwischen dem letzten Satz eines und dem ersten des folgenden Absatzes auf.

(j) *Besonders viele solcher Erreger stammten aus dem Fluss Hase, in den geklärtes Abwasser aus Osnabrück geleitet wurde. **Multiresistente Keime gelangen meist über zwei Wege in die Gewässer.*** (Beispiel aus Jablotschkin und Zinsmeister, 2020b)

(k) *Auch politisch rechte Verlage sind auf der Buch-Messe in Leipzig. **Was ist ein Verlag?***

(l) *Deutschlands erste „Plastic Attack“: Kund*innen entsorgen die Verpackungen ihrer Einkäufe im Supermarkt – und sorgen für Verwirrung. **Um fünf vor zwölf heißt es an diesem Samstag: Bühne frei.*** (Beispiel aus Jablotschkin und Zinsmeister, 2020b)

11. **Mehrere Diskursrelationen auf denselben Argumentspannen:** In bestimmten Fällen darfst du auch mehr als eine Relation auf der exakt gleichen Argumentspanne annotieren:

- Ein impliziter oder expliziter Konnektor hat verschiedene Lesarten. Die verschiedenen Lesarten gibst du dann unter SClass1A und SClass1B an.
- Es können mehrere implizite Relationen in ein Argumentpaar hineingelesen werden, die nicht durch ein und denselben Konnektor versprachlicht werden können. Die verschiedenen Konnektoren gibst du dann in den Feldern Conn1 und Conn2 an.
- In einer Relation sind mehrere explizite Konnektoren enthalten. Hier annotierst du erst eine Relation mit dem einen expliziten Konnektor, wie unter 2-6 beschrieben, und legst dann durch Klick auf „Add new relation“ eine weitere Relation für den zweiten expliziten Konnektor an. **Da diese Fälle am Ende ggf. separat ausgewertet werden müssen, notiere dir bitte genau den Dateinamen und die Textspanne, an der du diese Art von Doppeltannotation vorgenommen hast.**
- Du hast zwar bereits eine explizite Relation annotiert, bist aber der Meinung, dass eine weitere, implizite Relation enthalten ist, die sich semantisch von der expliziten unterscheidet. Hier annotierst du erst eine Relation mit dem expliziten Konnektor, wie unter 2-6 beschrieben, und legst dann durch Klick auf „Add new relation“ eine weitere Relati-

¹ Bei epistemischen Relationen (z.B. Reason+Belief) reicht ein Konnektor oft nicht aus, um die Bedeutung zu paraphrasieren. In diesem Fall kannst du im Kommentarfeld eine ausführlichere Paraphrase eintippen (z.B. „Deshalb bin ich der Meinung, ...“)

on mit implizitem Konnektor an. **Da diese Fälle am Ende ggf. separat ausgewertet werden müssen, notiere dir bitte genau den Dateinamen und die Textspanne, an der du diese Art von Doppeltannotation vorgenommen hast.**

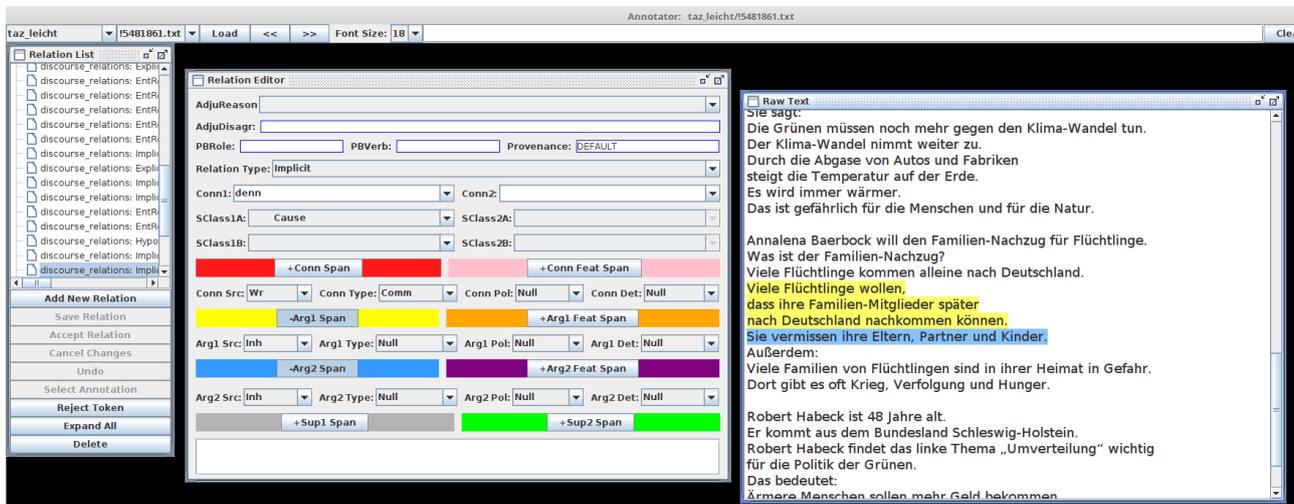


Abbildung 3: Beispiel für die Annotation einer impliziten Diskursrelation

2.2 Grundannahmen und allgemeine Konventionen

Notationskonventionen in den Beispielen:

Arg1 ist *kursiv*.

Arg2 ist **fett**.

Der Konnektor bzw. der AltLex-Ausdruck ist unterstrichen.

Supplemente sind in Klammern mit einem tiefgestellten Index. (_{Sup1} Supplement zu Arg1)

Implizite Konnektoren sind folgendermaßen gekennzeichnet: `Implicit=Konnektor`

2.2.1 Konnektoren

- Konnektoren sind subordinierende Konjunktionen (z.B. *weil, obwohl*), koordinierende Konjunktionen (z.B. *und, aber*), Adverbien (*jedoch, dann*), Pronominaladverbien (*trotzdem, deshalb, dadurch*) und bestimmte Präpositionalphrasen (*zum Beispiel*).
- Konnektoren können auch modifiziert sein (*zum Beispiel wenn, erst dann*).
- Alle lexikalischen Ausdrücke, die keiner der obengenannten Wortklassen angehören, aber trotzdem eine Diskursrelation markieren, werden als AltLex annotiert. Bist du dir nicht sicher, ob ein Ausdruck (z.B. eine Präpositionalphrase) zu den Konnektoren zählt, nimm grammis (Leibniz-Institut für Deutsche Sprache) als Referenz: <https://grammis.ids-mannheim.de/konnektoren/>
- Achtung: Koordinierende Konjunktionen (*und, aber, sowohl ... als auch ...*) werden nur dann als Konnektoren annotiert, wenn sie Sätze bzw. Verbalphrasen koordinieren. In Bei-

spielen (m), (n) und (o) markieren diese Ausdrücke also **keine** Diskursrelationen und werden nicht annotiert. Zu koordinierten Verbalphrasen siehe auch Abschnitt 2.2.2.

(m) Das Gericht in Paderborn hat 2 Menschen verurteilt.

- Einen Mann.
- Und eine Frau.

(n) Das ist wolkig, aber schlau. (Beispiel aus Jablotschkin und Zinsmeister, 2020b)

(o) Zwar ist es richtig, dass sowohl Habeck als auch Baerbock zum Realo-Flügel gehören. (Beispiel aus Jablotschkin und Zinsmeister, 2020b)

2.2.2 Formen von Argumenten

Ein Argument hat in der Regel die Form eines Satzes (vgl. (p) und (q)), Nebensatzes (vgl. (s) und (t)) oder Matrixsatzes (vgl. (s)).

Ein Ausnahmefall ist Beispiel (r): Hier liegt ein elliptischer Matrixsatz „so Hoffmann“ vor, welcher kein finites Verb enthält. Funktional stellt dieses Fragment eine Attribution dar, weil es eine Äußerung einer Person zuordnet. Der gesamte Satz „[...] so Hoffmann“ darf in diesem Fall als Argument annotiert werden, obwohl er elliptisch ist, um den Relation Sense zwischen Arg1 und Arg2 annotieren zu können. Wenn die Attribution aber nicht zur Bedeutung des jeweiligen Sense beiträgt, wird sie bei der Annotation der Markables ausgelassen. Dies ist z.B. dann der Fall, wenn die Relation zwischen zwei Propositionen bzw. *abstract objects* innerhalb der Äußerung besteht (vgl. Prasad *et al.*, 2007, p. 40, Bsp. 144).

In manchen Fällen kann auch ein Relativsatz ein Argument darstellen (vgl. (u)). (Für weitere Konventionen zu Relativsätzen siehe Abschnitt 2.2.4.)

Koordinierte Verbalphrasen werden ebenfalls als Argumente annotiert. Ist der explizite Konnektor *und*, gilt dies allerdings nur, wenn sie zusätzlich zur koordinierenden Konjunktion einen weiteren (impliziten oder expliziten) Konnektor enthalten (vgl. (v) und Erklärung zur Conjunction-Relation in Jablotschkin (2023) sowie Entscheidungsbaum im Anhang dieses Dokuments). → Abweichung von Webber *et al.* (2019)

Wenn der Relationstyp Hypophora² vorliegt, sind die Argumente die Frage und die Antwort darauf (vgl. (w)).

In den folgenden Beispielen ist Arg1 kursiv, Arg2 fett und der Konnektor unterstrichen. Andere syntaktische Strukturen, z.B. von Präpositionen abhängige NPs (vgl. (x)) oder Aufzählungen (vgl. (y)), annotieren wir nicht als Argumente.

(p) *Viele Menschen wollen den Paragraphen 219a abschaffen.* Deshalb **beschäftigt sich jetzt auch der Bundestag mit dem Paragraphen 219a.**

(q) *Und auch in den Lehrerzimmern herrscht Skepsis.* **Spricht man mit erfahrenen Lehrkräften, hört man oft die Sorge, dass die „Alten“ nun künftig auch noch für alle**

2 Die der Hypophora-Relation vorausgehende Relation wird i.d.R. entweder als EntRel annotiert, wenn sie eine einfache Wortklärung einführt (vgl. Bsp. (k)), oder als implizite Contingency.Cause+SpeechAct.Result+Speech Act-Relation, wenn sich aus dem vorangehenden Text eine Informationslücke ergibt, die geschlossen werden soll (siehe Jablotschkin, 2023). Der zweite Fragentyp tritt häufig in Form von Zwischenüberschriften auf und verknüpft größere Textabschnitte.

Fragen der „Neuen“ zuständig sein sollen. (Beispiel aus Jablotschkin und Zinsmeister, 2020b)

- (r) *Lehrerverbandschef Meidinger und GEW-Frau Hoffmann fordern die Kultusminister deshalb auf, für Qualitätsstandards zu sorgen. „Nötig ist ein Masterplan für gemeinsame Standards und die Qualifizierung der Seiteneinsteiger“, so Hoffmann.* (Beispiel aus Jablotschkin und Zinsmeister, 2020b)
- (s) *Sie müssen vielleicht vor Gericht, weil sie öffentlich gesagt haben: Wir machen Schwangerschafts-Abbrüche.*
- (t) **Spricht man mit erfahrenen Lehrkräften, hört man oft die Sorge, dass die „Alten“ nun künftig auch noch für alle Fragen der „Neuen“ zuständig sein sollen.** (Beispiel aus Jablotschkin und Zinsmeister, 2020b)
- (u) *Zum vierten Mal besuchen die beiden den Workshop „Ick mach Welle!“, den die Lebenshilfe im Frühjahr gemeinsam mit dem Berliner Elektroniklabel Killekill initiiert hat.* (Beispiel aus Jablotschkin und Zinsmeister, 2020b)
- (v) [...] *über die Gülle werden sie auf die Felder gebracht und implicit=dann mit dem Regen in die Gewässer geschwemmt.*
- (w) *Was sind Keime? Keime sind Krankheitserreger.*
- (x) *wegen der Untersuchung*
- (y) *Hühner mit gebrochenen Beinen, Wunden und Infektionen, dazu verdurstete Tiere: Die Tierrechtsorganisation Peta wirft drei polnischen und vier niederländischen Hühnerfarm-Betreibern massive Tierquälerei vor.* (Beispiel aus Jablotschkin und Zinsmeister, 2020b)

2.2.3 Lokalisierung von Argumenten

- Überschriften werden nicht mitannotiert. Zwischenüberschriften werden annotiert, wenn sie den formalen Vorgaben für Argumente entsprechen (also z.B. satzförmig sind). Dabei ist aber zu beachten, dass Zwischenüberschriften häufig eher auf der Ebene der globalen und nicht der lokalen Textkohärenz operieren. Zwischenüberschriften verknüpfen häufig größere Textabschnitte und die verknüpften Argumente liegen häufig weiter auseinander (siehe Fußnote 2 zu Fragen als Zwischenüberschriften).
- Die Argumente von expliziten Relationen können irgendwo im Text stehen, also zum Beispiel im selben Satz wie der Konnektor, in den angrenzenden Sätzen oder an anderer Stelle im Text. Dasselbe gilt für AltLex.
- Die Argumente von impliziten Relationen werden in benachbarten Sätzen annotiert. Dabei ist es aber aufgrund des Minimalitätsprinzips nicht immer notwendig, jeweils den gesamten Satz als Argumentspanne zu annotieren (vgl. Bsp. (af)).
- Auch innerhalb von Supplementen dürfen Relationen annotiert werden.
- Ist der Konnektor eine subordinierende Konjunktion, ist Arg2 immer der untergeordnete Satz und Arg1 der Matrixsatz. Bei allen anderen Konnektoren ist das erste Argument Arg1 und das zweite Argument Arg2.

- Argument 1 und Argument 2 können unter bestimmten Bedingungen auch die Konjunkte einer VP-Koordination sein (vgl. Abschnitt 2.2.2).

2.2.4 Ausdehnung von Argumenten

- Satzzeichen an den Grenzen von Argumenten werden nicht mitannotiert.
- Die Konnektoren expliziter Relationen sind *nicht* Teil der Argumente (vgl. (z)), AltLex- und AltLexC-Marker allerdings schon (vgl. (aa)).

(z) *Kranke Menschen sterben oft, weil **Medikamente nicht wirken.***

(aa) *Ein Grund dafür ist: **Deutsche Ärzte verschreiben große Mengen Antibiotika an kranke Menschen.***

- Bei impliziten Relationen beinhalten die Argumente ggf. auch Konnektoren von angrenzenden expliziten Relationen (vgl. (ab)). Dasselbe gilt für explizite Relationen: Die Argumentspannen beinhalten auch Konnektoren, die eigentlich zu einer angrenzenden expliziten Relation gehören (vgl. (ac)).

(ab) *Die Länder machen zum Beispiel gemeinsame Einsätze mit ihren Soldaten. implicit=deshalb **Die Länder haben auf dem Treffen in der Stadt Brüssel beschlossen: Wir geben mehr Geld für unser Militär aus.***

(ac) *Doch ansonsten ist Wagenknechts Vorwurf unterkomplex. Denn **Habeck zielt mit seiner Rede auf einen gesellschaftlichen Trend, um den sich linkes Denken kümmern muss.** (Beispiel aus Jablotschkin und Zinsmeister, 2020b)*

- Minimalitätsprinzip: Die Argumentspannen sollen so groß wie nötig und so klein wie möglich annotiert werden. Das bedeutet, nur diejenigen Textspannen, die unbedingt zur Interpretation der Relation notwendig sind, werden als Argument markiert. Relativsätze und Appositionen können ggf. ausgespart werden, wenn sie für die Interpretation der jeweiligen Relation nicht notwendig sind (siehe Beispiele (ad) und (ae)). Das gleiche gilt für Adverbiale und Attribute nach satzbeendenden Interpunktionszeichen (siehe Beispiele (af) und (ag)). Die Argumente gelten dann trotzdem als benachbart, so dass auch eine implizite Relation annotiert werden kann. Weitere Hinweise zu „sub-sentential arguments“ finden sich in Abschnitt 3.3.1 in Prasad et al. (2007).

(ad) *Der Dritte im Bunde, Danilo Amalique, spielt ein paar Synthesizertöne dazu, **es wird nun laut im improvisierten Proberaum im Richardkiez in Berlin Neukölln.** (Beispiel aus Jablotschkin und Zinsmeister, 2020b)*

(ae) *Als Beweis präsentieren die Tierschützer Bildmaterial aus den betroffenen Unternehmen, das ihnen zugespielt wurde. **Die Fotos zeigen unter anderem Hühner, die zwischen Gitterstäben eingequetscht sind oder halb gerupft mit Entzündungen herumlaufen.** (Beispiel aus Jablotschkin und Zinsmeister, 2020b)*

(af) *In vielen Lebensmitteln sind Eier.* Zum Beispiel in Nudeln. **Wo kommen diese Eier her?**

(ag) *Die Gruppe bietet auf der Buch-Messe viele Veranstaltungen an.* Zum Beispiel Veranstaltungen zum Thema Rassismus. **Damit zeigt die Gruppe: Wir sind gegen rechte Hetze. Wir sind gegen rechte Verlage auf der Buch-Messe.**

- Bestimmte Doppelpunktkonstruktionen werden jeweils als *ein* Argument markiert, weil das einleitende Segment an sich nicht als *abstract object* (z.B. im Sinne eines Ereignisses, eines Zustands oder einer Proposition) interpretiert werden kann. Dies gilt bei Konstruktionen, bei denen über den Doppelpunkt hinweg eine syntaktische Dependenz besteht (z.B. Objekt nach dem Doppelpunkt). Wird nach dem Doppelpunkt Redewiedergabe eingeleitet, wird die komplette Rede als Teil des Arguments markiert, sofern eindeutig ist, wo sie aufhört (z.B. durch Einrückung markiert; vgl. Bsp. (ai)).

(ah) Viele Menschen glauben zwar: Annalena Baerbock und Robert Habeck setzen sich nicht für die Politik von der linken Gruppe ein.

(ai) Viele Politiker in Hamburg haben nämlich gesagt:

In Hamburg gibt es jedes Jahr nur 9 Feier·tage.

Aber in vielen anderen Bundes·ländern gibt es mehr Feier·tage.

Deshalb soll es in Hamburg bald einen neuen Feier·tag geben.

- Supplemente: Wenn du findest, dass eine Textspanne nicht unbedingt zur Interpretation des Arguments nötig ist, aber trotzdem weiterführende Informationen zu einem Argument (oder einem Teil aus dem Argument) liefert, kannst du diese als Supplement des Arguments annotieren (Sup1 Span bzw. Sup2 Span). Argumente, zwischen denen ein Supplement steht, gelten trotzdem als benachbart (das heißt, du kannst hier eine implizite Relation annotieren). Beispiel für eine implizite Relation, zwischen deren Argumenten ein Supplement steht:

(aj) *Keime können Menschen krank machen.* (_{Sup1} Zum Beispiel, wenn die Keime in eine Wunde gelangen). *Implicit = jedoch **Keime können mit Medikamenten behandelt werden.***

(ak) *Das Gericht in Paderborn hat 2 Menschen verurteilt.*

- (_{Sup1} Einen Mann.

- Und eine Frau.)

Diese 2 Menschen haben nämlich 2 Frauen getötet.

→ Hier wird die Referenz einer NP genauer beschrieben. Das Supplement ist nicht notwendig, um die Relation zu verstehen.

- Auch die Antezedenten von anaphorischen Ausdrücken werden als Supplemente annotiert.

(al) Bisher *war es bei den Grünen so:* (_{Sup1} Ein Vorsitzender kam von den Linken. Und ein Vorsitzender kam von den Realos. Sie waren sich bei vielen Themen nicht einig). **Mit**

den neuen Vorsitzenden ist es bei den Grünen so: (_{Sup2} Annalena Baerbock und Robert Habeck sind beide von den Realos. Sie sind sich bei vielen Themen einig).

- Bestimmte Textspannen werden nicht als Teil des Haupttextes annotiert, sondern wie kleine separate Texte behandelt. In (am) ist der eingerückte Teil eine Worterklärung. Es wird keine Relation zwischen dem Haupttext und der Worterklärung annotiert, sondern nur innerhalb der Worterklärung/dem eingerückten Textblock und zwischen den Sätzen vor und nach der Einrückung. In (an) wird eine Aussprache und in (ao) ein „schwieriges Wort“ eingeführt. Die entsprechenden Textsegmente werden bei der Annotation wie der eingerückte Block in (am) behandelt.

(am) *In Nordrhein-Westfalen war ein Prozess.*

Ein Prozess ist ein Streit vor Gericht.

Bei einem Prozess entscheidet das Gericht: [...]

Der Prozess war in Paderborn.

(an) Die rechte Zeitschrift „Compact“ plant zum Beispiel eine Lesung mit Akif Pirinçci. Pirinçci spricht man so aus: Pi-rin-tschi. Wer ist Akif Pirinçci?

(ao) Oft leben die Hühner dort in sehr kleinen Käfigen. Das nennt man: Käfig-Haltung. In Deutschland gibt es sehr wenig Käfig-Haltung.

Anhang

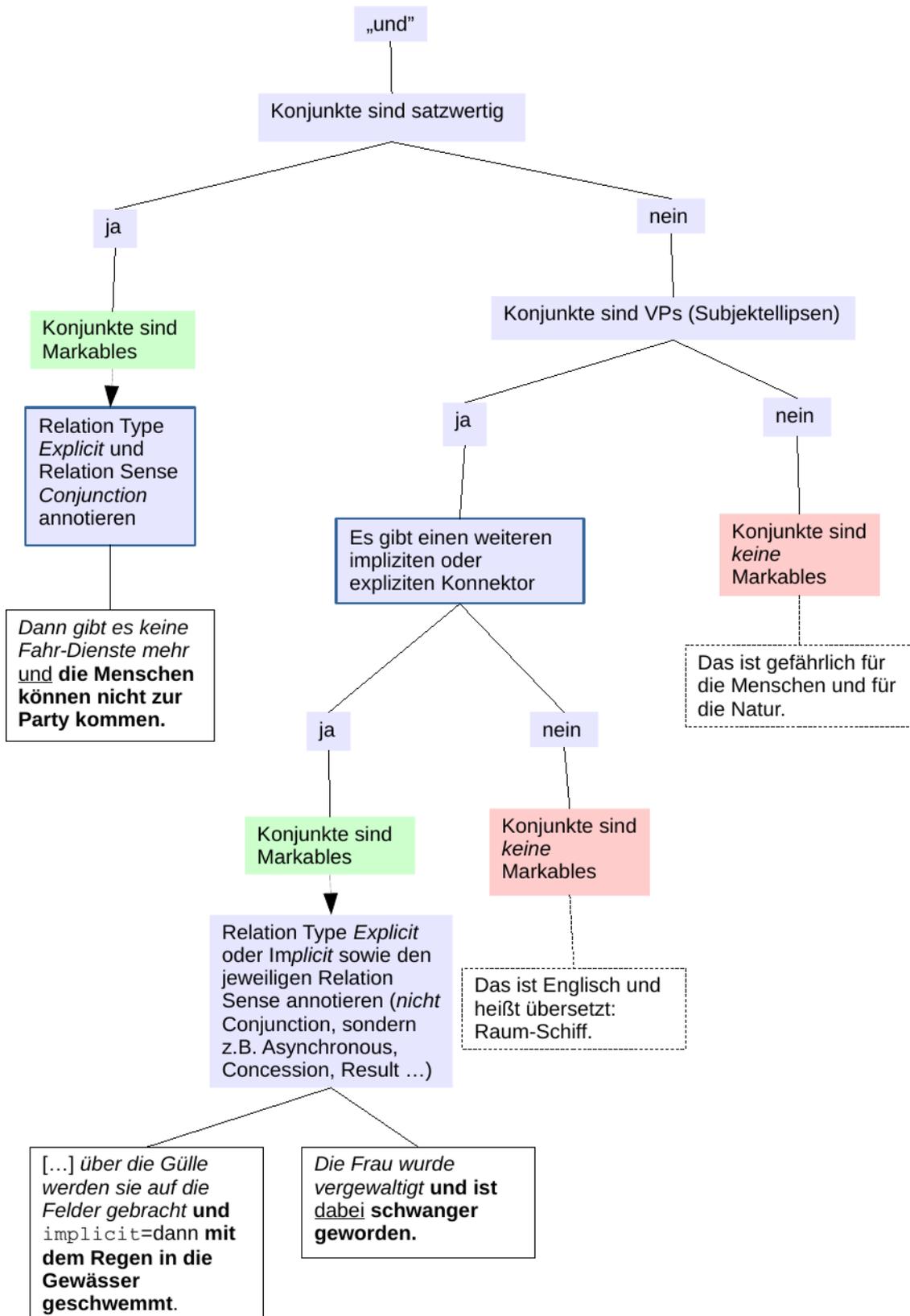


Abbildung 4: Entscheidungsbaum zum Umgang mit der Konjunktion "und" bei der Annotation

Quellen

Jablotschkin, S. (2023) ‘Senses nach den PDTB-Guidelines 3.0’. Available at: <https://doi.org/10.5281/zenodo.6362739>.

Jablotschkin, S. and Zinsmeister, H. (2020a) ‘LeiKo: A corpus of easy-to-read German’. Hamburg. Available at: <https://zenodo.org/record/3923917>.

Jablotschkin, S. and Zinsmeister, H. (2020b) ‘taz-Korpus. Standardsprachliches Referenzkorpus zu LeiKo’. Unveröffentlicht.

Leibniz-Institut für Deutsche Sprache (no date) *Wörterbuch der Konnektoren, grammis (Grammatisches Informationssystem)*. Available at: DOI: 10.14618/wb-konnektoren.

Prasad, R. *et al.* (2007) ‘The Penn Discourse Treebank 2.0 Annotation Manual’. Available at: <https://www.cis.upenn.edu/~elenimi/pdtb-manual.pdf>.

Webber, B. *et al.* (2019) ‘The Penn Discourse Treebank 3.0 Annotation Manual’. Available at: <https://catalog.ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf>.

Stichwortverzeichnis

Adverbial.....	8
AltLex.....	3ff., 7f.
AltLexC.....	3, 8
Appositionen.....	8
Arg1.....	2, 4ff.
Arg2.....	2ff.
Argumentspanne.....	4, 8
Attribut.....	8
Attribution.....	6
Conjunction.....	6
Conn1.....	4
Conn2.....	4
diskontinuierliche Textspannen.....	2
Doppelpunktkonstruktion.....	9
eingerrückt.....	10
EntRel.....	4
Explicit.....	2
Frage.....	3, 6
Hypophora.....	3, 6
implicit.....	7f.
Implicit.....	4f., 9
Konnektor.....	2ff.
koordinierend.....	5f.
Koordinierend.....	5
Koordinierte Verbalphrasen.....	6
Matrixsatz.....	2, 7
Minimalitätsprinzip.....	8
NoRel.....	4
Relation Type.....	2, 4
Relativsatz.....	6
Relativsätze.....	6, 8
Satzzeichen.....	8
SClass1A.....	2, 4
SClass1B.....	2, 4
Sense.....	2ff.
Shell Nouns.....	3
subordinierend.....	2, 5, 7

Sup1.....	5, 9
Sup2.....	9f.
Supplement.....	5, 7, 9
überschrift.....	7
Überschrift.....	7
VP-Koordination.....	8
Wortklärung.....	10
Zwischenüberschrift.....	7