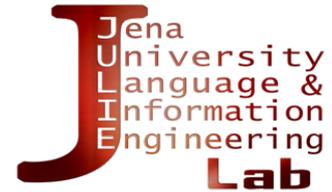




FRIEDRICH-SCHILLER-  
UNIVERSITÄT  
JENA



# Annotationsleitlinien für deutschsprachige Medizintexte.

## Teil 4: Annotation von Medikationsangaben

Franz Matthies, Christina Lohr, Luise Modersohn & Udo Hahn

Jena University Language & Information Engineering Lab (JULIE Lab)

Friedrich-Schiller-Universität Jena

**JULIE Lab – SMITH Technischer Report 4**

<https://doi.org/10.5281/zenodo.7707947>

Dezember 2022

# Inhaltsverzeichnis

<b>1</b>	<b>Vorwort</b>	<b>2</b>
<b>2</b>	<b>Einleitung</b>	<b>3</b>
<b>3</b>	<b>Daten</b>	<b>3</b>
<b>4</b>	<b>Annotationsleitlinien</b>	<b>4</b>
4.1	Allgemeine Hinweise . . . . .	4
4.2	Typensystem für Medikationsangaben . . . . .	4
4.2.1	Medikation / Medikament . . . . .	5
4.2.2	Dosierung . . . . .	7
4.2.3	Modus . . . . .	8
4.2.4	Häufigkeit . . . . .	9
4.2.5	Dauer . . . . .	10
4.2.6	Grund . . . . .	12
4.3	Einige komplexe Annotationsbeispiele . . . . .	13
<b>5</b>	<b>Das Annotationswerkzeug BRAT</b>	<b>14</b>
5.1	Hinweise zur Konfiguration von BRAT . . . . .	14
5.2	Hinweise zur Nutzung von BRAT . . . . .	14
<b>A</b>	<b>Zusammenfassung Typensystem</b>	<b>18</b>
	<b>Referenzen</b>	<b>19</b>

# 1 Vorwort

Die hier dokumentierten Annotationsleitlinien für deutschsprachige Medizintexte sind im Rahmen des SMITH-Projekts<sup>1</sup> im Zeitraum zwischen 2016 bis 2022 entstanden.

Der vorliegende Teil 4 hat die Kennzeichnung von Medikationsangaben in klinischen Dokumenten zum Gegenstand. Diese semantische Kategorie zählt neben Diagnosen, Befunden und Symptomen (s. hierzu Teil 3 der Annotationsleitlinien) zu den zentralen inhaltlichen Beschreibungen eines Patienten<sup>2</sup> in Arztbriefen. Medikationsangaben stehen im klinischen Kontext zwischen klassischen semantischen Entitäten (wie Diagnosen, Befunde und Symptome), da sie neben dem Namen des jeweiligen Medikaments (eine weitere zentrale semantische Entität) üblicherweise auch ergänzende Angaben zu seinen Attributen (Vergabeform, Dauer der Vergabe, Vergabemenge, zeitliche Angaben zur Vergabe und den Grund sowie das Ende der Vergabe) enthalten. Damit werden letztlich semantische Relationen zwischen dem jeweiligen Medikament und seinen Attributen beschrieben. Solche, jedoch wesentlich allgemeinere semantische Relationen werden in Teil 5 der Annotationsleitlinien ausführlicher behandelt.

---

<sup>1</sup>Das SMITH-Projekt (<https://www.smith.care/de/smith-konsortium/ueber-smith/>) wurde im Rahmen der Medizininformatik-Initiative (MII – <https://www.medizininformatik-initiative.de>) vom Bundesministerium für Bildung und Forschung (BMBF) gefördert (Fördernummer: 01ZZ1803G).

<sup>2</sup>Wir verwenden dem sprachlichen Usus der Mehrheit der Sprechergemeinschaft des Deutschen und linguistischen Argumenten folgend das generische Maskulinum zur Referenzierung aller Geschlechter.

## 2 Einleitung

Die hier vorliegende Annotationsleitlinie beschreibt, wie mithilfe des *BRAT rapid annotation tool* (Kapitel 5) für deutschsprachige Medizintexte Medikamente zusammen mit deren Dosierung, Modus, Häufigkeit, Dauer und Gründen zur Verabreichung annotiert werden (Kapitel 4). Hierfür wurden die Leitlinien der *i2b2 Challenge* für englischsprachige klinische Dokumente von 2008 adaptiert (1).

Durch die Bereitstellung dieser Leitlinien soll es der Forschungsgemeinschaft ermöglicht werden, einerseits unsere Daten zu reproduzieren, andererseits unsere Vorschläge in zukünftigen Projekten zu nutzen bzw. weiterzuentwickeln.

Die vorliegende Annotationsleitlinie ist Grundlage der folgenden Veröffentlichung:

(2) Hahn et al., *3000PA - Towards a National Reference Corpus of German Clinical Language*; MIE 2018 (<https://doi.org/10.3233/978-1-61499-852-5-26>)

## 3 Daten

Die Annotationen, für die die hier dokumentierten Leitlinien entwickelt wurden, wurden an realen klinischen Arztbriefen durchgeführt. Diese Briefe sind Teil des 3000PA-Textkorpus (2). Das Korpus wurde im Rahmen der Pilotphase des SMITH-Projekts an den Standorten Jena, Leipzig und Aachen nach folgenden Kriterien zusammengestellt:

- Alle Briefe entstanden zwischen 2010 bis 2015.
- Alle in den Briefen beschriebenen Patienten waren mindestens fünf Tage in einem der Universitätsklinika Jena, Leipzig oder Aachen stationär in Behandlung.
- Die Behandlung fand auf einer intensivmedizinischen oder internistischen Station statt.
- Alle Patienten waren zum Zeitpunkt der Datenerhebung im September 2016 bereits verstorben.

Das Jenaer Segment von 3000PA besteht insgesamt aus 1106 deutschsprachigen Arztbriefen, die sich aufteilen in:

- 383 Kurzberichte,
- 103 Verlegungsbriefe und
- 620 Entlassbriefe.

## 4 Annotationsleitlinien

### 4.1 Allgemeine Hinweise

Dieser Annotationsleitlinie werden einige Begriffsklärungen vorangestellt, die zentrale Konzepte von Annotationsprozessen klären sollen:

- **Token:** kleinster Bestandteil eines Textes, der für die Annotation relevant ist, z.B. einzelne Wörter und Satzzeichen voneinander getrennt.
- **Annotat:** inhaltlich markiertes Textstück (ein einzelnes Token oder eine Spanne von Token) in einem Text, hier in einem Arztbrief.
- **Entität:** semantischer Typ eines Annotats; dieser Typ ist Teil grundlegender Taxonomien, die den Gegenstandsbereich der Annotationsaufgabe inhaltlich (kategorisch) strukturieren. Das jeweilige Annotat kann als konkrete Ausprägung (Instanz) eines Typs betrachtet werden.
  - Beispiel: Das Annotat [*Insulin*] wird der Entität *Medikation* zugeordnet.
- **Relation:** inhaltliche Beziehung zwischen zwei Instanzen von Entitäten; dies ist die Grundlage für die Repräsentation medizinischer Fakten.
  - Beispiel: Das Annotat [*Ringer-Acetat*], das der Entität *Medikament* zugeordnet ist, wird mit dem Annotat [*1000 ml*], das der Entität *Dosierung* zugeordnet ist, verknüpft. Damit wird die (hier implizite) Relation *hat-Dosierung* etabliert.<sup>3</sup>

### 4.2 Typensystem für Medikationsangaben

Das Ziel der Annotationsarbeiten besteht darin, folgende sechs Medikationskategorien zu kennzeichnen (die hier eingeführten Farbkodierungen werden im weiteren Verlauf für entsprechende Beispiele verwendet; für eine Zusammenfassung, siehe Appendix A):

- **Medikation**
- **Dosierung**
- **Modus**
- **Häufigkeit**
- **Dauer**
- **Grund**

---

<sup>3</sup>Die Unterscheidung zwischen impliziten und expliziten Relationen hat historische Gründe, die aus den Annotationsvorgaben des *i2b2 Challenge* von 2008 herrühren und die wir für die vorliegende Annotationsleitlinie übernommen haben. Bei einer impliziten Relation wird ein Link von einer Instanz zu einer anderen Instanz einer Entität gesetzt, ohne diesen Link zu etikettieren (die im Beispiel angesprochene Relation *hat-Dosierung* folgt lediglich implizit aus der Verknüpfung einer *Dosierungs*-Instanz mit einer *Medikamenten*-Instanz). Bei einer expliziten Relation wird dieser Link ausdrücklich etikettiert, d.h. mit einem eigenen, inhaltscharakterisierenden Namen versehen. Solche expliziten Relationen sind Gegenstand des Teils 5 der Annotationsleitlinien.

Für die Annotation sind nur **positive** Medikationsindikationen relevant, die ausschließlich dem Patienten, von dem im Arztbrief berichtet wird, **direkt zugeordnet** werden können.

Die Annotation *Medikament* ist die Basisannotation, worauf sich die anderen fünf (*Dosierung, Modus, Häufigkeit, Dauer, Grund*) beziehen und im hier verwendeten Annotationswerkzeug BRAT auch entsprechend mit Relationspfeilen markiert werden. Demzufolge existieren keine Annotationen für die letztgenannten, wenn nicht eine entsprechende Medikamentenannotation vorhanden ist, auf die diese sich beziehen können.

#### 4.2.1 Medikation / Medikament

Unter Medikationen bzw. Medikamenten sind nur Substanzen zu verstehen, die ein Patient

- eingenommen / zugeführt bekommen hat,
- einnimmt / zugeführt bekommt,
- einnehmen / zugeführt bekommen soll.

**Was wird annotiert?** Medikationen werden nur dann markiert, wenn die Einnahme durch den Patienten explizit im Text angegeben ist. Jegliche andere Erwähnung (bspw. eine Einnahme eines Medikaments durch Familienangehörige) soll nicht annotiert werden.

Demzufolge gelten als **zulässige Medikamente:**

- Rezeptpflichtige Substanzen:
  - Handelsname (z.B. Vancocin®, Vancomycin)
  - Generika
  - Inhaltsstoffe (z.B. Furosemid)
  - Kollektiver Name für eine Gruppe von Medikamenten (z.B. Diuretika)
- Freiverkäufliche Substanzen:
  - Handelsname (z.B. Aspirin®)
  - Inhaltsstoffe (z.B. Vitamin E)
  - Kollektiver Name für eine Gruppe von Medikamenten (z.B. Vitamine)
- Biologische Substanzen (vom Arzt vorgeschlagen/verordnet):
  - Totale Parenterale Ernährung (TPE)
  - Inhaltsstoffe der TPE, wenn gesondert gelistet
  - Blut für Transfusionen und dessen Zusammensetzung (so gesondert gelistet)
  - (Intravenöse) Infusionen & Kochsalzlösungen
- Kontrastmittel, Betablocker, etc.
- Pronomen, die auf eine der oben erwähnten Substanzen verweisen

**Keine zulässigen Annotationen für Medikamente sind:**

- Nahrung bzw. Nahrungsbestandteile
- (nicht-)alkoholische Getränke
- Zigaretten, Tabak etc.
- andere (il)legale Drogen (Haschisch, Kokain, Designer-Drogen wie Ecstasy usw.)

**Wie werden Medikamente annotiert?** Für die Annotation soll die komplette Nominalphrase markiert werden, die zur Erwähnung eines Medikaments zählt. Wörter wie *Tabletten*, *Spray*, *Salbe* o. ä. werden mitmarkiert, sofern sie direkt nach der Medikamentenbezeichnung vorkommen (z.B. *Turixim®-Salbe*, *Polyspektran-Augentropfen*). Zahlenangaben werden mit zur Medikationsannotation gezählt, wenn sie inhärenter Teil dessen sind (z.B. *D-Fluoretten® 500 I.E.*). (Ausnahmen werden später erläutert.)

### Beispiele:

1. Eine einzelne Erwähnung eines Medikaments ist als solche zu markieren
  - „Der Patient erhielt **Vancomycin Saft**.“
  - „Wir initiierten daraufhin eine antimykotische Behandlung mit **Fluconazol**.“
2. Eine einzelne Erwähnung eines Medikaments zusammen mit etwas anderem (tritt oft als Nominalkompositum auf) wie
  - „**Insulin**therapie“
3. Ein einzelnes Medikament, das durch zwei Namen referenziert erwähnt wird, indem etwa einer davon in Klammern steht, resultiert nur in **einem** Eintrag
  - „**Eisen(III)-natrium-gluconat-Komplex (B. Ferrlecit)** 30 mg Mo. + Mi. + Fr.“
4. Koreferenzen (ein sprachlicher Ausdruck, der meist in Form von Pronomen oder definiten Nominalphrasen auftritt) verweisen auf ein im Text zuvor bereits eingeführtes Diskursobjekt und müssen entsprechend aufgelöst werden, solange das zur Identifizierung einer Medikation notwendig/förderlich ist. Zur *Auflösung* muss in BRAT ein entsprechender Relationspfeil gesetzt werden.
  - „Aus diesem Grund begannen wir die Therapie mit **Codein-Tabletten**<sup>Medikament</sup>. **Diese**<sup>Anapher</sup> bitten wir im weiteren Verlauf schnellstmöglich zu reduzieren.“
5. Aufzählungen werden einzeln aufgelöst.
  - „Trotz kurzzeitiger antibiotischer Therapie mit **Tazobac** + **Vancomycin** bzw. **Meropenem** + **Vancomycin** trat intermittierend Fieber auf.“
6. Inhaltsstoffe von Kombipräparaten werden ebenfalls einzeln aufgelöst.
  - „**Kaliumchlorid** + **Macrogol** + **Natriumchlorid** + **Natriumhydrogencarbonat** (z.B. **Movicol-Beutel**) 1-0-0-0“
7. In diesem Zusammenhang sollte auch darauf geachtet werden, dass jegliches Medikationsattribut (Dosierung, Häufigkeit, ...) mit jedem Wirkstoff verbunden wird bzw. nur mit dem übergeordneten Medikament (im vorherigen Beispiel würde die Häufigkeitsangabe „1-0-0-0“ nur mit „Movicol-Beutel“ verbunden). Während bei folgendem Beispiel die Dosierung „4,5 mg“ sowohl mit „Piperacillin“ als auch mit „Tazobactam“ verbunden würde.
  - **Piperacillin** / **Tazobactam** 4,5 mg
8. **Negativ-Erwähnungen** von Medikamenten (bspw. Medikamentenallergie) werden nicht annotiert:

- Aufgrund einer Amoxicillin-Unverträglichkeit wurde **Erythromycin**<sup>Medikament</sup> ver-schrieben.

9. Wird jedoch die Medikamentenunverträglichkeit durch eine entsprechende Einnahme, die im Bericht protokolliert ist, ersichtlich, muss das Medikament aufgrund der Einnahme annotiert werden.

#### 4.2.2 Dosierung

Mit der Dosierung wird annotiert, wie viel einer Medikation per Verabreichung genommen wird, bspw.: „Eine Tablette, [50 mg], [1 Ampulle].“

**Was wird annotiert?** Zur Annotation der Dosierung werden numerische sowie textuelle Information einschließlich Abkürzungen markiert, die angeben, wie viel eines Medikaments eingenommen/verabreicht wird und beinhaltet demzufolge:

- 100 ml
- 40 mg
- 1 Tablette
- 4 IE
- auch Angaben wie „10mg/die“, „3 Liter/Minute“

#### Es wird nicht annotiert:

- Angaben kumulativer Mengen, die der Patient einnehmen soll

**Wie wird annotiert?** Angelehnt an die Medikament-Annotation werden nur die Dosierungen annotiert, die den im Arztbrief beschriebenen Patienten direkt betreffen (in der Vergangenheit, aktuell, in der Zukunft). Unabhängig davon, ob Dosierungsangaben direkt aufeinanderfolgen oder nicht, werden sie als je ein einzelnes Attribut annotiert. Unterschiedliche Dosierungen des gleichen Medikaments werden auch als je eine Dosierung annotiert.

#### Beispiele:

1. Allgemein:

- „Vorthherapie mit **Prednisolon** **50mg/d**<sup>Dosierung</sup> über wenige Tage“
- „... **2l O2/min**, aktuelle Betriebsstunden 3135h, in Beatmungspausen **2,5l O2** via Nasensonde“
- „**Miflunide** **200 µg** 2 x 1 **Hub**“

2. Wenn ein Bereich für eine Dosierung gegeben ist, wird dieser als **ein** Attribut annotiert:

- „**Oxycodon** **5-10 mg**“

3. Dosierungsangaben, die **im** konkreten Medikamentnamen stehen, werden mit als Medikament annotiert. Angaben, auf die kein weiterer Teil des Medikamentnamens folgt, werden als Dosierung annotiert:

- „**Dilzem 90 mg retard** 1-0-1“ ← keine Dosierung
- „**D-Fluoretten®** **500 I.E.**“

4. Eine Dosierung für mehrere Medikamente kann und soll in BRAT auch entsprechend mehrfach verknüpft werden.
5. Wenn eine Dosierungsangabe sich eventuell auf mehrere Medikamente beziehen kann, die jedoch *nicht eindeutig* ist, wird diese dem *nächststehenden* zugeordnet.
6. Weitere Beispiele:
  - „**Valsartan** **80/12,5 mg**“
    - „80/12,5 mg“ als eine Dosierung annotieren
  - „**Sitagliptin** / **Metformin** 50/1000“
    - Die beiden Inhaltsstoffe werden einzeln annotiert (s. Annotation von Medikamenten). Die beiden Zahlenangaben werden weder zum Medikament noch als Dosierung annotiert, wenn nicht genau bekannt ist, ob es sich insgesamt um eine Medikamentenbezeichnung (z.B. aus der „Roten Liste“ o.ä.) handelt.

### 4.2.3 Modus

Durch den *Modus* wird annotiert, wie die Medikation eingenommen/verabreicht wurde/wird.

**Was wird annotiert?** Es werden die Stellen im Text annotiert, die die Art und Weise der Medikationsverabreichung angeben; aber nicht, wenn die Textstelle Teil des Medikationsnamens ist. Beispiele zu annotierender Modi sind demnach:

- *oral*
- *intravenös*
- *lokaltherapeutisch*
- *sublingual* (unter der Zunge liegend)
- alle Formen von Abkürzungen solcher Modi

**Es werden nicht annotiert:**

- Angaben, die die Beschaffenheit des Medikaments beschreiben und somit Teil desselben sind (**Transtec pro Pflaster**) ← Kein Modus

**Wie wird annotiert?** Für jede Medikation wird die entsprechende Verabreichungsform markiert.

**Beispiele:**

1. Jede Erwähnung wird mit einer Annotation versehen. Verschiedene Arten, die die gleiche Dosierung ausdrücken, werden mit je einer Annotation versehen; ebenso wenn eine Änderung der Art der Dosierung erwähnt ist:
  - „Aufgrund der ausgeprägten Dyspnoesymptomatik begannen wir zunächst mit einer **intravenösen**<sup>Modus</sup> **Prednisolon** therapie.“
  - „Auch die bisher **oral** erfolgte Einnahme von **Theophyllin** wurde kurzzeitig auf die **intravenöse** Gabe umgestellt.“

- „Zudem wurde bei paraklinisch erhöhten Entzündungswerten eine antibiotische Therapie mit **Ceftriaxon** begonnen, die im Verlauf zügig **oralisiert** werden konnte.“

2. Auch für die Annotation des Modus' gilt das Gleiche wie für die Dosierung: ein Modus kann auf mehrere Medikamente verweisen und soll entsprechend in BRAT verknüpft werden.

#### 4.2.4 Häufigkeit

Mit der Häufigkeit werden numerische/textuelle Einzelwörter bzw. Wortgruppen oder entsprechende Abkürzungen annotiert, die angeben, wie oft jede Dosierung administriert wird/wurde.

**Was wird annotiert?** Jeder Ausdruck, der angibt, wie oft eine Dosis eines Medikaments verabreicht/ingenommen wird/wurde, wird durch die Häufigkeit annotiert und beinhaltet:

- Frequenzen, z.B.:
  - *zweimal am Tag*
  - *jeden Tag*
  - *täglich*
  - *x 1*
  - *einmal im Monat*
  - *x 3 täglich*
- Ausdrücke der Form „je nach Bedarf“, wenn diese alleine erscheinen oder direkt in Zusammenhang mit einer Häufigkeitsangabe stehen, z.B.:
  - *im Bedarfsfall*
  - *dreimal am Tag, wenn nötig*
  - *im Bedarfsfall dreimal am Tag*
  - z.B. auch dann, wenn die Häufigkeit an Bedingungen geknüpft wird, wie etwa in „bei Schmerzen“
- konkrete(re) Zeitangaben (oft Präpositionalphrasen), z.B.:
  - *nach dem Essen*
  - *vor dem Abendessen / Schlafengehen*
  - *um 15 Uhr*
  - *mit jedem Essen*
- Kombinationen aus den oben genannten Möglichkeiten:
  - *vor dem Schlafengehen, wenn nötig*
  - *dreimal vor dem Essen*
  - *jeden Tag um 16 Uhr*
- Ausdrücke wie
  - *1-0-0*
  - *½-0-1 IE*
  - *1MB-0-0*
  - *eine ... Gabe*
  - *Abendgabe*

### Die Annotation der **Häufigkeit** beinhaltet nicht:

- „im Bedarfsfall“ / „wenn notwendig“, wenn es nach einer Dauer (nicht der Häufigkeit) steht:
  - „Der Patient soll täglich Lasix 40 mg einen Monat lang, wenn notwendig, nehmen.“

**Wie wird annotiert?** Für die Annotation der Häufigkeit gilt die gleiche grundlegende Herangehensweise wie bisher: nur Angaben, die sich direkt auf den im Arztbrief beschriebenen Patienten beziehen, werden annotiert, wobei die gesamte Phrase markiert wird.

### Beispiele:

1. Die erwähnten „*Ausdrücke des Bedarfs*“ und eventuelle weitere Häufigkeitsangaben werden einzeln annotiert:
  - „**Ambroxol-Saft** **5 ml** **bei Bedarf** **morgens und abends**“ ← 2x Häufigkeit
  - „**Metamizol** (z.B. **Novalgin-Tropfen**) **bei Bedarf** (**4 x 15 Trpf.**)“
2. In Medikationslisten stehende Häufigkeiten sind auch zu annotieren, wie folgendes Beispiel zeigt:
  - „**ASS** **100 mg** **1-0-0**“
3. Es muss unbedingt darauf geachtet werden, dass Angaben mit „IE“ als Dosierung und nicht als Häufigkeit annotiert werden:
  - „**Actrapid** nach Schema  
Blutzucker < 8 mmol/l = **2 IE**  
8 – 10 mmol/l = **4 IE**  
[. . .]“
4. Gaben:
  - „Damit kann **eine Toujeo-Gabe** zur **Abendgabe** des **Apidra** erfolgen.“

### 4.2.5 Dauer

Durch die *Dauer* werden Angaben annotiert, die ausdrücken, für wie lange (über welchen Zeitraum) eine Medikation verabreicht oder eingenommen wird/werden soll.

**Was wird annotiert?** Mit der Angabe der *Dauer* werden Ausdrücke annotiert, die die Gesamtdauer einer zeitlichen Spanne beschreiben, in der die Medikation in einer bestimmten Dosierung verabreicht oder eingenommen wird/werden soll. Für den Fall, dass die Medikationseinnahme in der Vergangenheit liegt, wird annotiert, wie lange die Beendigung her ist. Die Annotation der Dauer beinhaltet:

- Präzise zeitliche Angaben, z.B.:
  - 10 Tage lang
  - für die Dauer von 10 Tagen
  - für einen Monat
  - konkrete Daten („seit 10/2011“, „am 11.02.2010“ etc.), aber nur, wenn diese in Verbindung mit einer Präposition stehen (z.B. „seit“, „für“, „am“, etc.)
- Vage zeitliche Angaben, z.B.:
  - bis die Symptome verschwinden
  - so lange, wie nötig
  - für kurze Zeit
  - passager
  - zeitweise
- Kumulative Mengen einer Medikation (wurden bei *Dosierung* ausgeschlossen), z.B.:
  - 10 Packungen
  - Anzahl an Dosierung notwendig: 10
- Angaben wie „beendet“ und „pausiert“

#### Die Annotation der Dauer beinhaltet nicht:

- Zeitliche Ausdrücke, die angeben, wann eine Dosis eingenommen werden soll (unter Häufigkeit geführt)
  - „Das Medikament soll während physischer Aktivität eingenommen werden.“
  - „Wann immer körperliche Anstrengung unternommen wird, soll das Medikament eingenommen werden.“
- Kumulative Dosierungen für eine Verabreichung (entgegen der für eine Behandlung)
  - „Nehmen Sie bis zu drei Tabletten und verständigen Sie Ihren Arzt, wenn diese nicht anschlagen.“
- Aussagen wie „Langzeit“, z.B. „Langzeit-Therapie“ (beschreibt keine direkte Dauer für eine Medikationsgabe)

**Wie wird annotiert?** Für die Annotation der Dauer gelten die gleichen Grundprinzipien wie für die Annotation der Häufigkeit. Die komplette Präpositionalphrase, inklusive Wörter wie „ab“, „bis“, „für“ etc., wird annotiert.

#### Beispiele:

- „Wir bitten um Wiederbeginn der **Xarelto**-Therapie ab dem 31.12.2017<sup>Dauer</sup> mit **10 mg/Tag** für 3 Tage<sup>Dauer</sup>, anschließend **15 mg** für weitere 3 Tage<sup>Dauer</sup>, danach **Xarelto** **20 mg** (wie gehabt).“
- „**Alprostadil**-Infusionstherapie über insgesamt 3 Tage<sup>Dauer</sup>“
- „**Methylprednisolon** **20 mg** **1-0-0** (bitte innerhalb der nächsten **2 Wochen**<sup>Dauer</sup> ausschleichen)“

#### 4.2.6 Grund

Der medizinische *Grund*, der Anlass für die Verschreibung des Medikaments ist, wird ebenfalls als Medikationsinformation annotiert. Indikatoren, aufgrund derer das Medikament normalerweise verschrieben wird, die jedoch nicht konkret im Text als Grund vorkommen, werden nicht markiert.

#### Was wird annotiert?

Angegebene medizinische Gründe, die die Ursache für die Verschreibung des Medikaments sind, werden annotiert und beinhalten z.B.:

- *Schwindel*
- *Fieber*
- *Diabetes*

#### Die Annotationen für den medizinischen *Grund* beinhalten nicht:

- Nicht-medizinische Gründe, aufgrund derer das Medikament appliziert wird
- Finanzielle oder Versicherungsgründe
- Alles, was als Laborwert gilt

**Wie wird annotiert?** Für den Grund soll nach Möglichkeit nur die kürzeste aussagekräftige Adjektivphrase (ein oder mehrere aufeinanderfolgende Adjektive) bzw. die längste Nominalphrase (ein oder mehrere Nomen mit eventuell vorangestelltem Adjektiv und voran- oder zwischengestelltem Artikel) markiert werden.

#### Beispiele:

##### 1. Der Standardfall

- „Bei Zustand nach Bronchitis besteht aktuell ein **anhaltender, nicht produktiver**<sup>Grund</sup> Husten. Aus diesem Grund begannen wir die Therapie mittels **Codein-Tabletten**.“
- „Es erfolgte zur **Stabilisierung der Kreislaufsituation**<sup>Grund</sup> der Patientin die **intravenöse** bilanzierte **Flüssigkeit**ssubstitution.“
- „Aufgrund der **ausgeprägten Dyspnoesymptomatik**<sup>Grund</sup> begannen wir zunächst mit einer **intravenösen** **Prednisolon**therapie“

2. Unterschiedliche Verweise, die den gleichen Grund meinen, sind als zwei verschiedene Gründe zu annotieren. Es sollen keine komplexen Phrasen annotiert werden; wenn möglich, sollen aufeinanderfolgende Token (Wörter) annotiert werden. Koordinationen (z.B. mit „und“ verknüpfte Wortgruppen) sollen nicht aufgelöst werden. Das kann zur Folge haben, dass es mehrere *Grund*-Annotationen für eine *Medikament*-Annotation gibt.

- „[. . .] und verordneten bei **schwerer respiratorischer Partialinsuffizienz** und **restriktiver Ventilationsstörung** eine Langzeit-Sauerstofftherapie.“

### 4.3 Einige komplexe Annotationsbeispiele

Die folgenden Beispiele illustrieren das gemeinsame Austreten der unterschiedlichen Attribute von Medikamenten in einzelnen Sätzen (und durchaus auch die Komplexität der Annotationsaufgabe):

- „zunächst **Levetiracetam** **1 g** als Kurz**infusion**.“
- „Die stationäre Aufnahme des Patienten erfolgte aufgrund eines **entgleisten Diabetes mellitus** zur Optimierung der **Insulin**therapie.“
- „**Insulin** **s.c.**: **Actraphane** **0-0-4-0 IE** (reduziert)“
- „**anhaltender, nicht produktiver Husten**“
- „PTA mit **paclitaxel**-freisetzendem Ballon (5x40 mm)“
- „Wir stellten die Patientin nach Empfehlung durch die Kollegen der Neurologie auf **Levetiracetam** ein.“
  - „Empfehlung . . . der Neurologie“ ist hier eher kein Grund, da es nur ein impliziter medizinischer Hinweis ist.
- „**Sauerstoff**langzeittherapie seit ca. 1999 mit **1 Liter/Minute**; seit 08/2010 **3 Liter/Minute**“
  - „Langzeit“ wird nicht als Dauer annotiert. Angaben wie „.../h, .../min.“ etc. sind als Dosierung und nicht separat als Dauer zu annotieren.
- „**2 Erythrozytenkonzentrate**“
  - In solchen Fällen wird die Zahl als Häufigkeit annotiert.
- „**Levetiracetam** **500 mg** **1-0-1** ( ab 20.01.2014 **750 mg** **1-0-1**)“
  - Die Angaben in den Klammern werden auch mit dem Medikament verknüpft.
- „**Alprostadil**-**Infusion**stherapie“
  - Annotationen können auch nur Teile von Wörtern/Token umfassen

## 5 Das Annotationswerkzeug BRAT

Die folgenden Abschnitte beschreiben die Installation bzw. Konfiguration sowie die Verwendung des BRAT-Annotationswerkzeugs für die manuelle Annotation von Medikationsangaben in deutschsprachigen Arztbriefen.

### 5.1 Hinweise zur Konfiguration von BRAT

Für die manuelle Annotation der Texte wird das Werkzeug BRAT RAPID ANNOTATION TOOL – BRAT (3) verwendet.<sup>4</sup> BRAT wird über einen Internet-Browser verwendet und kann entweder lokal oder auf einem Server installiert werden.

Vorab muss BRAT konfiguriert werden. Die notwendigen Konfigurations-Dateien werden unter <https://doi.org/10.5281/zenodo.7707947> bereitgestellt.

### 5.2 Hinweise zur Nutzung von BRAT

Im Browser muss sich jeder Annotator vorab oben rechts mit den zugewiesenen individuellen Login-Daten einloggen. Im Anschluss daran wählt man aus der *Collection* eine Aufgabe bzw. einen Text aus und kann anschließend mit der Bearbeitung beginnen.

1. Durch einen Doppelklick auf ein Token bzw. eine (dis-)kontinuierliche Spanne von Token markiert, und ein Menü zur Auswahl aus den sechs Annotationskategorien, die eine Medikation beschreiben (Medikament, Dosierung, Modus, Häufigkeit, Dauer, Grund), erscheint (s. Abb. 1). Die Markierung kann auch durch Klicken und Ziehen durchgeführt werden, wodurch die Markierung von Wortteilen oder mehreren Wörtern (Token) möglich ist.

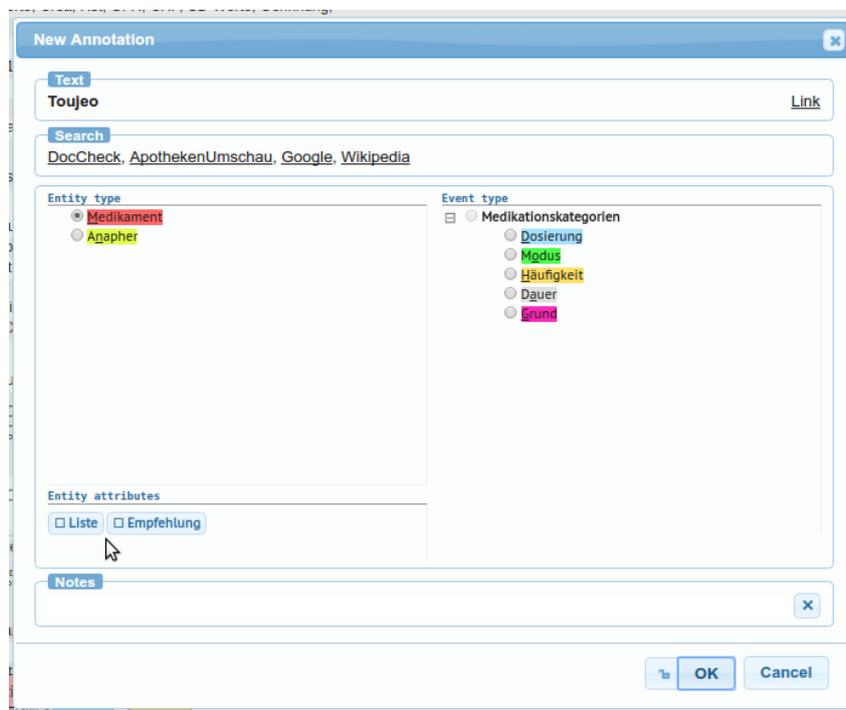


Abbildung 1: Allgemeine Auswahl nach Markierung

<sup>4</sup><http://brat.nlplab.org/index.html>

Ergänzend zu diesen inhaltlichen Kategorien erlaubt die Annotation auch die Kennzeichnung von (Medikament-)Anaphern. Hierbei handelt es sich meist um Pronomen bzw. definite Nominalphrasen, die im Text auf ein bereits eingeführtes Medikament rückverweisen (s. Abb. 2). Die Einteilung in die Kategorien *Entity* und *Event* beruht auf der Annahme, dass ein *Medikament* die Basisannotation ist (*Entität*) und die weiteren fünf inhaltlichen Kategorien diese nur ergänzen bzw. ohne diese gar nicht existieren würden (*Ereignis*).

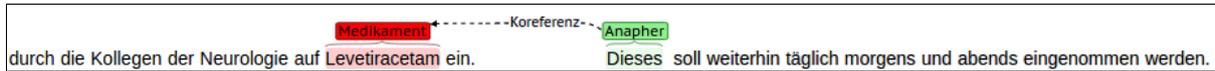


Abbildung 2: Anapher-Relation

2. Um die Produktivität während der Annotation zu erhöhen, hat jede Annotationskategorie zur komfortableren Auswahl ein Tastatur-Kürzel (entsprechender Buchstabe in Abb. 1 unterstrichen, siehe dazu auch Appendix A).
3. Aufgrund der Annahme aus (1.) ist jedes *Event* einzeln betrachtet unvollständig und bedarf einer Verknüpfung mit einem Medikament. BRAT weist darauf hin, indem die Annotationskategorie nicht ausgefüllt, sondern nur umrandet ist (s. Abb. 3).

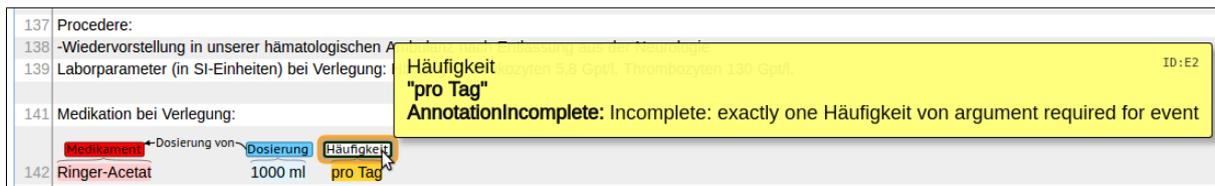


Abbildung 3: Fehlende Entity-Event-Verknüpfung

4. Um diese Verknüpfung herzustellen, wird das *Event* durch Klicken und Ziehen auf die Medikament-Annotation gesetzt (s. Abb. 4). Es erscheint ein entsprechender Pfeil und beim Loslassen der Maus ein Auswahlménü (s. Abb. 5). Im Fall der Event-Medikament-Annotation ist dieses Auswahlménü auf eine Kategorie beschränkt („Event von“).

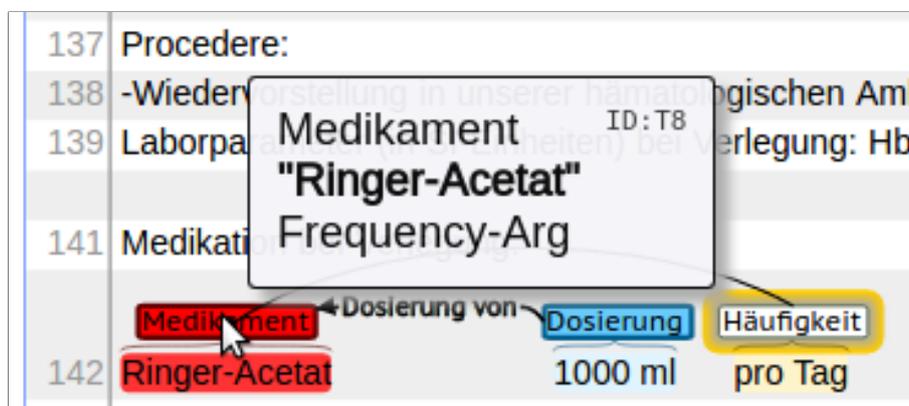


Abbildung 4: Ausführen der Verknüpfung

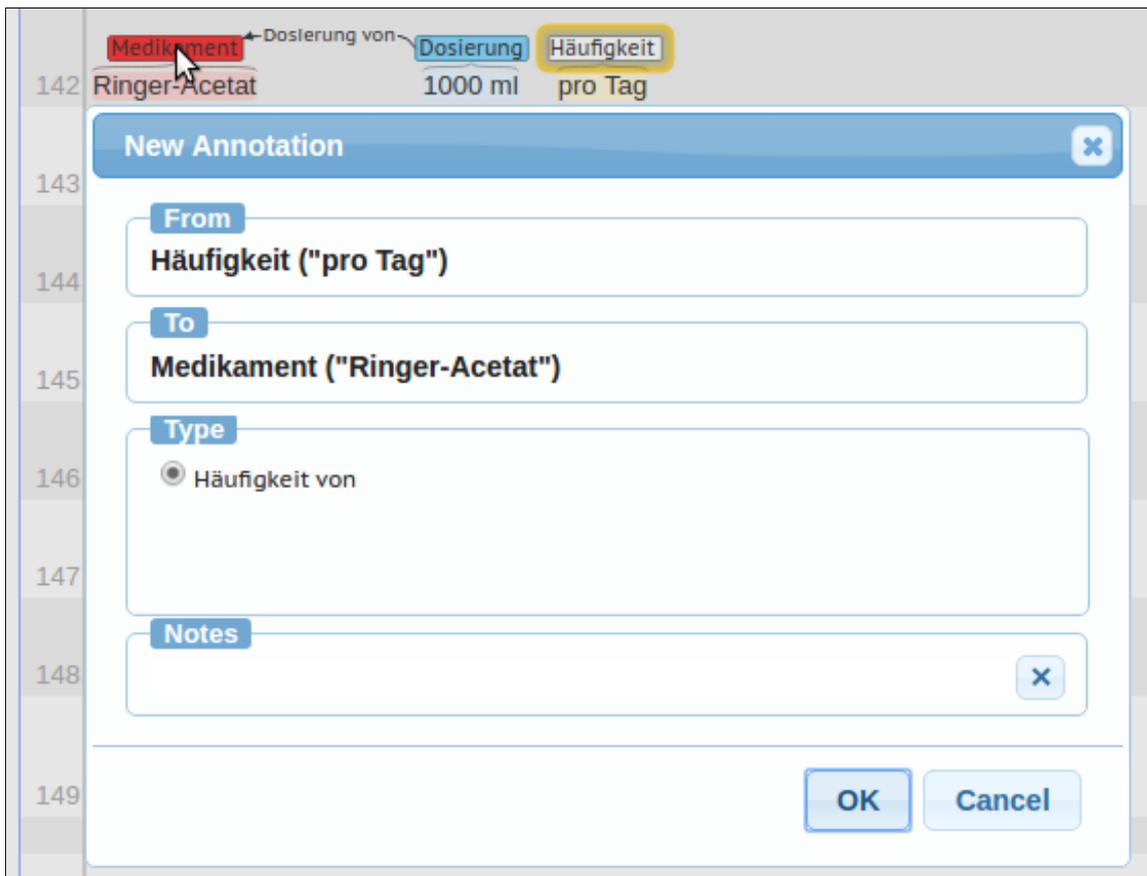


Abbildung 5: Verknüpfung - Auswahlmennü

5. Eine Beziehung zwischen Medikament und *Event* wird nur gesetzt, wenn letzteres nicht mehr als drei Zeilen vor oder nach dem Medikament erscheint.
6. Neben *Events* und *Entitäten* können zudem zwei *Relationen* annotiert werden: **Äquivalenz** und **Koreferenz**. Das Auszeichnen dieser Relationen geschieht auf die gleiche Weise, wie das der *Events* durch Klicken und Ziehen. Der Unterschied besteht lediglich in den potenziellen Argumenten

Relationen	
<b>Äquivalenz</b>	Medikament – Medikament (Äquivalenz zwischen Text-Elementen)
<b>Koreferenz</b>	Anapher → Medikament

7. Jede Medikament-Entität kann außerdem zwei Attribute annehmen: *Liste* oder *Empfehlung*. Diese erscheinen als Auswahlmöglichkeit, wenn *Medikament* ausgewählt wurde (s. Abb. 1). Das Attribut *Liste* wird gesetzt, wenn sich das entsprechende Medikament in einer Liste (etwa *Entlassmedikation*) befindet. Das Attribut *Empfehlung* wird gesetzt, wenn das Medikament nur zur Einnahme empfohlen wird, die Einnahme jedoch nicht gesichert ist.

8. Es kann mitunter notwendig sein, *mehrere Token* als zu *einer Annotation* *zusammengehörig* zu markieren, wenn diese nicht direkt aneinander angrenzen, zum Beispiel bei „... eine Toujeo-Gabe ...“, wenn „eine Gabe“ als *Dosierung* angenommen wird (s. Abb. 6). Diese Verknüpfung zwischen „eine“ und „Gabe“ kann erreicht werden, indem eines der beiden Fragmente als eine entsprechende Annotation markiert wird (hier: „Gabe“ als „*Dosierung*“). Nach erneutem Doppelklick auf die Annotation erscheint der abgebildete Dialog, bei dem mit der Schaltfläche *Add Frag.* (s. Abb. 8) ein weiteres Token als zugehöriges Fragment markiert werden kann (s. Abb. 7), um die Verknüpfung zu erreichen.

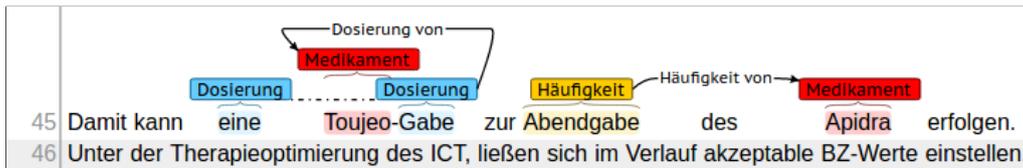


Abbildung 6: Beispiel für nicht-kontinuierliche Annotation

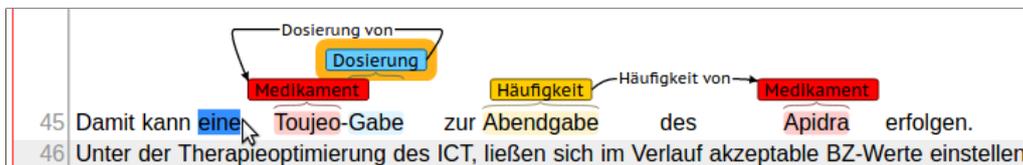


Abbildung 7: Markierung des zugehörigen Tokens

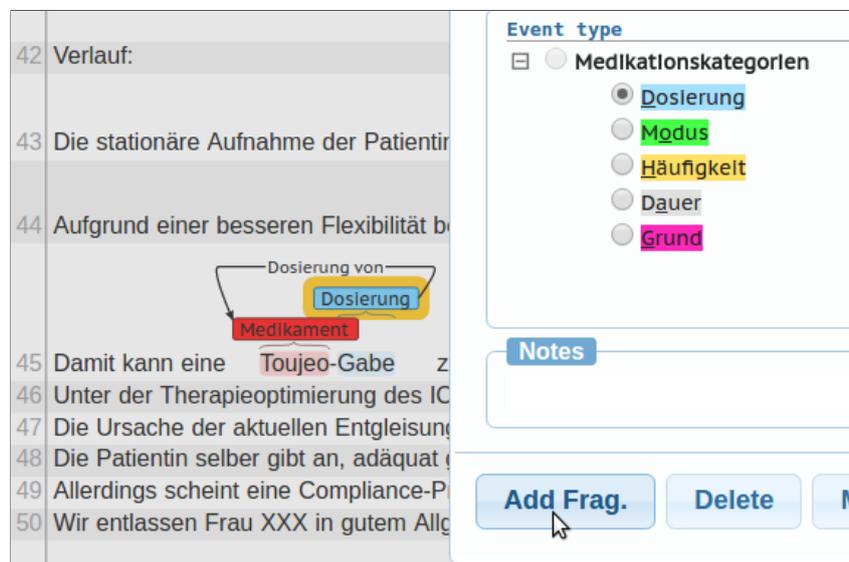


Abbildung 8: „Add Frag.“-Funktionalität

9. Anmerkung: Wenn eine Annotation über mehrere Zeilen angelegt werden soll (nicht die Zeilen, die durch Veränderung der Browser-Größe entstehen, sondern die, die durch Nummern am Anfang einer Zeile angegeben werden), so muss die „*Add Frag.*“-Funktion benutzt werden, um den Teil in der neuen Zeile als zugehörig zu markieren. Geschieht dies nicht, kann BRAT das Dokument eventuell nicht mehr richtig anzeigen oder stürzt sogar bei der Anzeige der Dokumentdatei ab.

## A Zusammenfassung Typensystem

Typ	Beschreibung	Short-Cut
<b>Medikation</b> / <b>Medikament</b>	Substanzen, die ein Patient einnimmt / zugeführt bekommt	M
<b>Dosierung</b>	<i>wie viel</i> einer Medikation eingenommen / verabreicht wurde	D
<b>Modus</b>	<i>wie</i> die Medikation eingenommen / verabreicht wurde	O
<b>Häufigkeit</b>	wie oft jede Dosierung vom Arzt administriert wurde	H
<b>Dauer</b>	wie lange Medikation eingenommen / verabreicht wurde	A
<b>Grund</b>	Anlass für Verschreibung des Medikaments	G

Table 1: Übersicht der Typensystems für Mediaktionsannotationen sowie deren Short-Cuts der BRAT-Konfiguration

## Referenzen

- [1] Özlem Uzuner, Imre Solti, Fei Xia, and Eithon Cadag. Community annotation experiment for ground truth generation for the i2b2 medication challenge. *Journal of the American Medical Informatics Association*, 17(5):519–523, 2010.
- [2] Udo Hahn, Franz Matthies, Christina Lohr, and Markus Löffler. 3000PA: towards a national reference corpus of German clinical language. In Adrien Ugon, Daniel Karlsson, Gunnar O. Klein, and Anne Moen, editors, *Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth - Proceedings of MIE 2018, Medical Informatics Europe, Gothenburg, Sweden, April 24-26, 2018*, volume 247 of *Studies in Health Technology and Informatics*, pages 26–30. IOS Press, 2018.
- [3] Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. BRAT: A Web-based tool for NLP-assisted text annotation. In Frédérique Segond, editor, *EACL 2012 — Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics: Demonstrations. Avignon, France, April 25-26, 2012*, pages 102–107, Stroudsburg/PA, 2012. Association for Computational Linguistics (ACL).