

# Allgemeine Semantische Annotation

## GeMTeX Annotationsrichtlinie

Justin Hofenbitzer<sup>1</sup>,  
Christina Lohr<sup>2</sup>, Andrea Riedel<sup>3</sup>, Jutta Romberg<sup>4</sup>,  
Jakob Faller<sup>3</sup>, Sarah Riepenhausen<sup>5</sup>, Miriam Schechner<sup>6</sup>,  
Markus Wolfien<sup>7,8</sup>, Peter Klügl<sup>9</sup>, Jazia Omeirat<sup>10</sup>,  
Jacqueline Lammert<sup>1</sup>, Stefan Schulz<sup>9</sup>, Martin Boeker<sup>1</sup>,  
Luise Modersohn<sup>1</sup>

<sup>1</sup>Technische Universität München,

<sup>2</sup>Universität Leipzig, <sup>3</sup>Friedrich-Alexander Universität Erlangen-Nürnberg,

<sup>4</sup>Charité Berlin, <sup>5</sup>Universität Münster, <sup>6</sup>Ludwig-Maximilian-Universität München,

<sup>7</sup>Technische Universität Dresden, Medizinische Fakultät,

<sup>8</sup>Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Dresden,

<sup>9</sup>Averbis, <sup>10</sup>Universitätsklinikum Essen

justin.hofenbitzer@tum.de

Version 2.3.0: 06.11.2025

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung und Motivation</b>	<b>2</b>
<b>2</b>	<b>Grundbegriffe und Annotationsmaximen</b>	<b>4</b>
2.1	Grundbegriffe . . . . .	4
2.2	Annotationsmaximen . . . . .	7
2.3	Linguistische Besonderheiten . . . . .	11
<b>3</b>	<b>Annotation von Konzepten</b>	<b>13</b>
3.1	Semantik der Konzeptannotation . . . . .	13
3.2	SNOMED CT Semantic Tags . . . . .	17
3.3	Konzepte in SNOMED CT . . . . .	18
3.3.1	Core-Konzepte . . . . .	18
3.3.2	Modifier-Konzepte . . . . .	23
3.3.3	Qualifier-Konzepte . . . . .	26
3.4	Familiäre Erkrankungen . . . . .	34
<b>4</b>	<b>Das No-Human Layer: Markierung von ausschließlich maschinell annotierten Bereichen</b>	<b>35</b>
<b>5</b>	<b>Das Hinzufügen von Relationen</b>	<b>36</b>
5.1	Inhaltliche Relationen . . . . .	36
5.2	Relationen bei familiären Erkrankungen . . . . .	37
5.3	Kausalrelationen . . . . .	37
5.4	Anaphorische Relationen . . . . .	37
5.5	Messwertrelationen . . . . .	38
5.6	Schrankenrelationen . . . . .	38
5.7	Koordinationsrelationen . . . . .	39
5.8	Faktualitätsrelationen . . . . .	39
5.9	Temporalrelationen . . . . .	40
<b>6</b>	<b>Annotationspragmatik: Komplette Beispielannotationen und Grenzfälle</b>	<b>42</b>
<b>7</b>	<b>Kurationshinweise</b>	<b>45</b>

# 1 Einleitung und Motivation

Im klinischen Alltag wird nach wie vor eine Vielzahl von Texten erstellt und verwaltet. Da Texte als *unstrukturierte Daten* gelten, ist die statistische Auswertung oder Weiterverarbeitung der enthaltenen Daten äußerst aufwändig.

Das Forschungsgebiet, das sich mit dieser Problematik beschäftigt, nennt sich *Computerlinguistik* oder auch *Natural Language Processing* (NLP) und fokussiert die automatische Verarbeitung natürlicher Sprache. Methoden in diesem Forschungsgebiet reichen von regelbasierter Neurosymbolik bis hin zu maschinellem Lernen. Zur Entwicklung und Evaluation dieser Methoden im klinischen Kontext werden geeignete Trainingsdaten gebraucht, insbesondere im deutschen Sprachraum.

Zu diesem Zweck wurde im Rahmen der Medizininformatikinitiative (MII) das Projekt **GeMTeX**<sup>1</sup> (German Medical Text Corpus) ins Leben gerufen. Ziel des Projektes ist es, klinische Texte aus der Routineversorgung mit Metadaten anzureichern und für die Forschung zur Verfügung zu stellen. An dem Projekt sind zahlreiche Forschungseinrichtungen und Universitätskliniken aus ganz Deutschland beteiligt. Sechs Universitätskliniken sind annotierende Standorte: Die Charité in Berlin, das Universitätsklinikum Carl Gustav Carus Dresden, das Universitätsklinikum Erlangen, das Universitätsklinikum Essen, das Universitätsklinikum Leipzig und das TUM Klinikum München. Jeder dieser sechs Standorte extrahiert und annotiert nach den gleichen Richtlinien bis zu 10.000 klinische Dokumente. Auf diese Weise soll eine variantenreiche Sammlung unterschiedlichster klinischer Dokumente entstehen, die mit diversem medizinischem Hintergrund und hoher Varianz in Formulierung, Krankheitsverläufen und anderen Merkmalen die deutsche klinische Sprache abbildet.

Grundlage für die Annotation der klinischen Dokumente ist die derzeit jeweils aktuelle Vollversion der englischsprachigen International Edition der umfassendsten klinischen ontologiegestützten Terminologie SNOMED CT<sup>2</sup>. Mit Hilfe von SNOMED CT werden Textspannen in den GeMTeX-Texten direkt *gegroundet*, also mit einer standardisierten Bedeutung versehen. So wird beispielsweise dem Wort *Tilidin* das SNOMED CT Konzept “SCTID: 373562008 | Tilidine (substance)” zugeordnet und ist direkt wieder auffindbar. Relationen werden anschließend zwischen semantisch verflochtenen Konzepten gezogen<sup>3</sup>.

Die praktische Umsetzung der semantischen Annotation findet maßgeblich in dem Annotationstool *INCEpTION*<sup>4</sup> statt<sup>5</sup>. Die Annotationskräfte bekommen von der Averbis Health Discovery<sup>6</sup> maschinell vorannotierte, reale Klinikdokumente mit Patientenbezug, die zuvor de-identifiziert wurden. Ein Recommender-System von ID Logik schlägt für jede annotierte Textspanne alternative SNOMED CT Konzepte vor. Die Annotationskräfte müssen entsprechend der Qualität der Vorannotationen diese korrigieren oder löschen und den Text nach nicht erkannten Spannen durchsuchen. Zur Findung des korrekten

---

<sup>1</sup><https://www.medizininformatik-initiative.de/de/gemtex-medizinische-texte-fuer-die-forschung-automatisiert-erschliessen>

<sup>2</sup>[https://www.bfarm.de/DE/Kodiersysteme/Terminologien/SNOMEDCT-CT/\\_node.html](https://www.bfarm.de/DE/Kodiersysteme/Terminologien/SNOMEDCT-CT/_node.html)

<sup>3</sup>Wichtig: Es werden **nicht** die SNOMED CT inhärenten Relationen annotiert (s. Abschnitt 5).

<sup>4</sup><https://inception-project.github.io>

<sup>5</sup>Nähere Informationen zu INCEpTION befinden sich im GeMTeX De-Identifizierungsguide: <https://zenodo.org/records/11502329>

<sup>6</sup><https://averbis.com/health-discovery/>

SNOMED CT Konzepts, wird empfohlen, über den Webbrowser den SNOMED CT International SNOMED CT Browser<sup>7</sup> für den Zugriff auf die englischsprachige International Edition zu verwenden.

---

<sup>7</sup><https://browser.ihtsdotools.org/>

## 2 Grundbegriffe und Annotationsmaximen

### 2.1 Grundbegriffe

Begriff	Definition
<b>Adjazenz</b>	Direkte Nachbarschaft zweier Einheiten (z.B. zweier Tokens oder SNOMED CT Konzepte).
<b>Ambiguität</b>	Eine sprachliche Mehrdeutigkeit, die sich nicht oder nur durch Kontext- oder Hintergrundwissen auflösen lässt.
<b>Anapher</b>	Ein zurückverweisendes sprachliches Element innerhalb eines Textes.
<b>Annotat</b>	Eine mit einer SNOMED CT Concept ID bzw. einem SNOMED CT Konzept versehene Textspanne.
<b>(Semantische) Domäne</b>	<i>Technisch-formal:</i> Die Menge aller zur Annotation zulässigen Konzepte. <i>Pragmatisch:</i> Grundsätzlich alle SNOMED CT Konzepte mit Ausnahme von Konzepten unter "SCTID: 243796009   Situation with explicit context (situation)". Fokussiert werden sollen jedoch die in diesem Richtlinie hervorgehobenen Konzeptgruppen.
<b>Faktualität</b>	Faktualität beschreibt den Wahrheitsgehalt von Aussagen. Oft sind Trigger-Wörter, wie z.B. "vermutlich", für einen modifizierten Wahrheitswert einer Aussage verantwortlich.
<b>Koordination</b>	Syntaktisches Phänomen, wo zwei oder mehr sprachliche Einheiten durch eine Konjunktion oder Listenelemente getrennt sind, dabei aber dieselbe syntaktische Funktion erfüllen.
<b>Koordinierte Einheiten</b>	Sprachliche Einheiten, die dem syntaktischen Phänomen "Koordination" unterliegen.
<b>Koordinationsprädikat</b>	Eine Textspanne, die sich strukturell wie inhaltlich auf koordinierte Einheiten bezieht oder diese näher definiert.
<b>Negation</b>	Negation ist eine spezielle Art der Faktualität. Der Wahrheitswert von Aussagen wird invertiert: Wahre Aussagen werden falsch, falsche Aussagen werden wahr.
<b>Relation</b>	Inhaltliche Verknüpfung zwischen genau zwei Annotaten.
<b>Satz</b>	Eine Aneinanderreihung von Tokens, die einen abgeschlossenen Gedanken ausdrückt und zumindest aus einem Subjekt und einem Prädikat besteht.
<b>SCTID</b>	In diesem Richtlinie wird die Abkürzung <i>SCTID</i> verwendet, um auf die Kombination aus SNOMED CT Concept ID und SNOMED CT FSN zu referenzieren.
<b>SNOMED CT Concept ID</b>	Der numerische Identifikator eines SNOMED CT Konzepts.
<b>SNOMED CT</b>	Eine ontologie-basierte Terminologie, die eine systematische Beschreibung klinischer und medizinischer Prozeduren, Substanzen oder sonstiger Vorgänge, inklusive vieler Synonyme in verschiedenen Sprachen beinhaltet und miteinander in Beziehung setzt.
<b>SNOMED CT Fully Specified Name (FSN)</b>	Der SNOMED CT FSN ist eine eindeutige Beschreibung des zugrundeliegenden klinischen Konzepts bestehend aus einem Deskriptor und dem SNOMED CT Semantic Tag.
<b>SNOMED CT Konzept</b>	Eine in SNOMED CT vorhandene Einheit, die eine konkrete Semantik hat. Ein Konzept kann durch die Annotation einer Textspanne zugewiesen werden. Ein SNOMED CT Konzept setzt sich aus einer SNOMED CT Concept ID und einem SNOMED CT FSN zusammen. SNOMED CT Konzepte sind <i>präkoordiniert</i> , wenn sie sich aus mehreren Konzepten zusammensetzen.
<b>SNOMED CT Semantic Tag</b>	Alle SNOMED CT FSNs enden mit einem <i>semantic tag</i> in Klammern hinter dem Deskriptor. Sie geben an, zu welcher inhaltlichen Gruppe ein Konzept gehört.
<b>Textspanne</b>	Eine Sequenz von im Text aneinandergereihten Tokens, mindestens jedoch genau ein Token.
<b>Token</b>	Kleinste annotationsrelevante Einheit. Ein Token ist eine durch Leerzeichen oder sonstige Trennzeichen (z. B. ",", "-", "<", ":", etc.) limitierte aneinandergereihte Zeichenkette.

Tabelle 1: Definitionen wichtiger Begriffe

Symbol	Bedeutung	Hinweis
$\{x P(x)\}$	Menge	Der Ausdruck beschreibt eine Menge, deren Elemente x alle die Eigenschaft P(x) besitzen.
$A \subseteq B$	A ist eine Teilmenge von B	Jedes Element in A ist auch in B.
$A \subset B$	A ist eine strikte Teilmenge von B	Jedes Element in A ist auch in B, aber A und B sind nicht dieselbe Menge.
$A \cap B$	Die Schnittmenge zwischen A und B	Die Menge der Elemente, die sowohl in A als auch in B sind.
$\emptyset$	Die leere Menge	Die Menge ohne Elemente.
$A \neq B$	A ist ungleich B	Die Mengen A und B sind nicht identisch.
$ A $	Kardinalität der Menge A	Die Anzahl von Elementen in A
$x \in A$	x ist ein Element in A	x ist ein Element, A ist eine Menge und x ist Teil von A
$x < y$	Ungleichheitszeichen (kleiner als)	x ist kleiner als y.
$\forall$	Universalquantor	$\forall x \in A(P(x))$ bedeutet: "Für alle Elemente x in der Menge A gilt die Eigenschaft P(x)."
$\wedge$	Logisches UND	$P \wedge Q$ bedeutet, dass sowohl P als auch Q wahr sind.
$\vee$	Logisches ODER	$P \vee Q$ bedeutet, dass mindestens eines der beiden wahr ist.
$\neg$	Logische NEGATION	$\neg P$ bedeutet, dass P falsch ist.
$\rightarrow$	Logisches KONDITIONAL	$P \rightarrow Q$ bedeutet, wenn P wahr ist, ist auch Q wahr.
$\Leftrightarrow$	Logisches BIKONDITIONAL	$P \Leftrightarrow Q$ bedeutet, dass P genau dann wahr ist, wenn Q auch wahr ist.
1 oder 0	Wahrheitswerte	1: Eine Aussage ist wahr; 0: Eine Aussage ist falsch.
$\llbracket P \rrbracket$	Denotation/Interpretation	$\llbracket P \rrbracket$ ist die Bedeutung oder Interpretation von P.

Tabelle 2: Definitionen formaler Symbole, die in dieser Richtlinie verwendet werden.

Annotationsbeispiele werden in dieser Richtlinie wie in (1) dargestellt: Eine annotierte Textspanne wird mindestens im Satzkontext eingebettet und durch eckige Klammern umrahmt sowie fett geschrieben. Hinter dieser Textspanne befindet sich ein Buchstabe, der auf das genaue SNOMED CT Konzept hinweist, mit dem die Textspanne annotiert wurde. Das SNOMED CT Konzept wird wiederum unter dem Satz dargestellt und besteht aus der SNOMED CT Concept ID und dem SNOMED CT Fully Specified Name (FSN). Das konkrete Beispiel (2) veranschaulicht, dass die Konzeptbezeichnung “SCTID: 386661006 | Fever (finding)” aus der SNOMED CT Concept ID “386661006” und dem SNOMED CT FSN “Fever (finding)” besteht.

- (1) Ich bin ein [**Annotationsbeispiel**](a.).  
 a. SCTID: SNOMED CT Concept ID | Deskriptor (SNOMED CT Semantic Tag)
- (2) Der Betroffene hat [**Fieber**](a.).  
 a. SCTID: 386661006 | Fever (finding)

Relationen zwischen zwei Konzepten werden durch einen Pfeil zwischen den beiden Konzepten dargestellt und mit dem Label  $\vec{r}_{x,y}$  versehen. Der erste tiefgestellte Buchstabe meint das Start- und der zweite das Zielkonzept der Relation. Das ist beispielhaft dargestellt in (3).

- (3) Die [**Körpergröße**](a.) beträgt [**155**](b.) [**cm**](c.).  
 a. SCTID: 1153637007 | Body height (observable entity)  
 b. **155.0**  
 c. SCTID: 258672001 | Centimeter (qualifier value)

#### Hinweis

In dieser Richtlinie werden teilweise formale Ausdrücke verwendet. Die logischen Formeln sind zum Lesen des Guides nicht notwendig, sondern dienen lediglich der Verständnisunterstützung. Eine Übersicht über die in dieser Richtlinie verwendeten formalen Symbole findet sich in Tab. 2.

## 2.2 Annotationsmaximen

Um die Komplexität der semantischen Konzeptannotation mit SNOMED CT abzufedern, werden sechs Annotationsmaximen eingeführt, die beim Annotationsprozess stets umzusetzen sind. Folgende Annotationsmaximen werden eingeführt:

1. Concept before Structure!
2. Better Together!
3. Identify the Meaning!
4. Stack with Care!
5. In doubt? Be broad!
6. Three Minutes Max!

**Annotationsmaxime I. Concept before Structure!** Die erste Maxime besagt, dass die zu annotierende Textspanne immer durch die verfügbaren SNOMED CT Konzepte bestimmt wird. So entsprechen beispielsweise die Textspannen “von der Leiter gefallen”, “Arthrose im rechten Knie”, “Diabetes mellitus Typ 1” oder “Ascaris lumbricoides” jeweils einem einzelnen SNOMED CT Konzept und nicht einer Kombination aus verschiedenen Codes (s. (4), (5), (6) und (7)). Zumeist wird ein Konzept also eine größere Anzahl von Tokens umspannen.

- (4) Bei Pt. ist [**Arthrose am rechten Knie**](a.) bekannt.  
a. SCTID: 323321000119100 | Osteoarthritis of right knee joint (disorder)
- (5) Fr. Schubert ist [**von der Leiter gefallen**](a.).  
a. SCTID: 86591008 | Fall from ladder (event)
- (6) [**Diabetes mellitus Typ 1**](a.) wird vermutet.  
a. SCTID: 46635009 | Diabetes mellitus type 1 (disorder)
- (7) Wir gehen von einem [**Ascaris lumbricoides**](a.) aus.  
a. SCTID: 19061001 | Ascaris lumbricoides (organism)

**Annotationsmaxime II. Better Together!** Die Konzeptannotation kann hin und wieder tückisch sein, da oft nicht klar ist, wo die inhaltlichen und textuellen Grenzen liegen. Die zweite Maxime **Better Together!** gibt hierfür klare Regeln an die Hand:

**Adjazent und Sinnhaft:** Jede annotierte Textspanne sollte eine *möglichst spezifische Bedeutung* haben sowie eine *sinnhafte, abgeschlossene textuelle Einheit* darstellen. In diesem Zusammenhang spielt *Adjazenz* oder *textuelle Nachbarschaft*, also die relative Position mehrerer Tokens im Text, eine erhebliche Rolle. Nebeneinanderstehende sprachliche Einheiten (**Adjazenz**), deren Bedeutung mit einem SNOMED CT Konzept darstellbar ist (**Sinnhaftigkeit**), sollten **wenn möglich** gemeinsam annotiert werden. Sind zwei sprachliche Einheiten nicht zusammenhängend, sondern durch dritte Konzepte oder Satzgrenzen getrennt, so sollen sie nicht gemeinsam, sondern einzeln annotiert werden.



Idealerweise sollte der Satz *Patientin kommt mit verletzter linker Hand* also wie in (8) annotiert werden.

- (8) Patientin kommt mit **[verletzter linker Hand](a.)**.  
a. SCTID: 11791321000119108 | Injury of left hand (disorder)

**Pragmatik: Sinnhaftigkeit vor Adjazenz:** Alle Texte, die im Rahmen von GeMTeX annotiert werden, unterziehen sich einer maschinellen Vorannotation, d.h., dass die Texte schon SNOMED CT Annotate enthalten. Sollte die Vorannotation inhaltlich korrekte, aber kurzspanniger Konzeptannotationen liefern, so sollen diese angenommen werden. Dies gilt auch, wenn eigentlich eine längerspannige Annotation möglich gewesen wäre (z.B. aufgrund von Adjazenz). Demnach wären die Beispiele (9)-(11) äquivalent zur Musterlösung (8).

- (9) Patientin kommt mit **[verletzter](a.) [linker Hand](b.)**.  
a. SCTID: 417163006 | Traumatic or non-traumatic injury (disorder)  
b. SCTID: 85151006 | Structure of left hand (body structure)
- (10) Patientin kommt mit **[verletzter](a.) [linker](b.) [Hand](c.)**.  
a. SCTID: 417163006 | Traumatic or non-traumatic injury (disorder)  
b. SCTID: 7771000 | Left (qualifier value)  
c. SCTID: 85562004 | Hand structure (body structure)
- (11) Patientin kommt mit **[verletzer [linker](b.) Hand](a.)**.  
a. SCTID: 125599006 | Injury of hand (disorder)  
b. SCTID: 7771000 | Left (qualifier value)

Zusammengefasst sagt Maxime II aus, dass Spannen möglichst zusammenhängend annotiert werden sollen, wenn sie *adjacent* und *sinnhaft* sind. Sollte dies nicht möglich sein, so darf auf kurzspannige Konzepte zurückgegriffen werden. Gleiches gilt, wenn die Vorannotation äquivalente, aber kurzspannigere Annotationen vorschlägt: Diese sind dann zu akzeptieren.

**Annotationsmaxime III. Identify the Meaning!** Textspannen sollen **ausschließlich** anhand ihrer tatsächlichen Bedeutung annotiert werden. Der Kontext **muss** bei jeder Annotation berücksichtigt werden. Anaphern sind ebenfalls mit ihrer intendierten Bedeutung zu annotieren (s. Beispiel (12)).

- (12) Die Rhythmusstörungen wurden mit **[Amiodaron](a.)** behandelt. Der Pat. hat **[das Medikament](a.)** wegen NW eigenmächtig abgesetzt.  
a. SCTID: 774551007 | Product containing only amiodarone (medicinal product)

Ambiguitäten / Mehrdeutigkeiten sind möglichst nach dem Kontext oder medizinischem Wissen aufzulösen. Falls das nicht möglich ist, sollen alle plausiblen Konzepte zu dieser Textspanne vergeben werden, (sog. *gestapelte Annotationen* oder *Stacked Annotations*

entstehen). In (13) wird eine lexikalische Ambiguität veranschaulicht: Die Abkürzung *HWI* kann gleichermaßen für *Harnwegsinfekt* wie *Hinterwandinfarkt* stehen. Sollte die Ambiguität nicht durch den Kontext auflösbar sein, so sollen alle **relevanten**, möglichen Konzepte als *Stacked Annotations* annotiert werden.

- (13) Wir stellen einen **[[HWI](a.)](b.)** fest.
- a. SCTID: 68566005 | Urinary tract infectious disease (disorder)
  - b. SCTID: 233838001 | Acute posterior myocardial infarction (disorder)

#### Hinweis

Grundsätzlich gilt: Determinierer (z.B. “der”, “die”, “das”) sind nicht in die Annotation einzuschließen. Ausnahmen können gemacht werden, sofern ein Determinierer einen Bedeutungsbeitrag innerhalb des SNOMED CT Konzepts leistet, z.B. im Falle von Anaphern (s. z.B. (12)), oder als Teil einer Präpositionalphrase (s. z.B. (24)). In Bsp. (12) wird deutlich, dass der Determinierer “das” eine referentielle Funktion übernimmt (in etwa: “Es handelt sich um DAS, also letztgenanntes, Medikament”)

**Annotationsmaxime IV. Stack with Care!** Gestapelte Annotationen sind grundsätzlich so minimal wie möglich zu halten. Sie sind lediglich in den folgenden Situationen erlaubt:

- **Vollständige Überlappungen** sind erlaubt bei getrennten Partikelverben und vergleichbaren Einschüben (s. Maxime II).
- **Exakte Stapel** sind erlaubt, wenn Textspannen nicht eindeutig auflösbar sind, sie mehrdeutig oder ambig sind (s. Maxime III).

Ausnahmen gelten dann, wenn z.B. auf äquivalente Vorannotationen zurückgegriffen wird (z.B. (11)).

**Annotationsmaxime V. In doubt? Be broad!** Die fünfte Annotationsmaxime soll eine pragmatische Hilfestellung für schwierig zu annotierende Textstellen sein. Grundsätzlich gilt bei der semantischen Annotation, dass Textspannen mit solchen Konzepten annotiert werden sollen, die am exaktesten für die gegebene Textstelle sind (s. Maxime III). Sollte jedoch bei der Annotation nicht klar werden, ob und wenn ja, welches spezifischere Konzept zu wählen ist, so sollte auf das nächsthöhere *Elternkonzept* zurückgegriffen werden (s. Abschnitt 3.1 für nähere Erläuterungen).

**Annotationsmaxime VI. Three Minutes Max!** Die sechste Annotationsmaxime bezieht sich auf die Arbeitszeit der Annotationskräfte. Es wird im Verlauf dieses Dokuments deutlich, dass die semantische Annotation mit SNOMED CT ein aufwändiges Unterfangen sein kann. Um zu verhindern, dass sich die Annotationspersonen in größten Detailfragen von SNOMED CT verlieren, soll die Recherchezeit reguliert werden. Die Identifizierung und Annotation einer medizinischen Aussage darf nicht länger als drei Minuten dauern. Die Annotationskräfte sollten daher zunächst die Hauptinformation einer medizinischen Aussage identifizieren. Anschließend sollte der SNOMED CT Browser herangezogen werden, um Konzept(e) für diese Hauptinformation zu finden, um

sie dann in INCEpTION zu annotieren. Befinden sich noch wichtige Faktualitäten in der Aussage (z.B. Negation), so sollten diese ebenfalls annotiert werden (s. Kapitel 3.3.3.5). Sollten sich nach Abschluss der drei Minuten noch weitere Informationen in der jeweiligen Textpassage befinden, die keine Hauptinformationen sind, so werden diese nicht annotiert.

#### Hinweis

Es ist davon auszugehen, dass Hauptinformationen zumeist **Core Konzepte** sind (s. Kapitel 3.3.1).

Dieses Prozedere soll einmal anhand von Beispiel (14) illustriert werden. Zunächst sind die Kernaussagen des Satzes zu identifizieren: (i) Es liegt eine Hypertrophie am linken Ventrikel vor und (ii) die Funktion ist nicht beeinträchtigt. Für Aussage (i) sollten wir also ein Konzept finden, das die Hypertrophie am linken Ventrikel ausdrückt, wie in (15) dargestellt. Gleichermaßen suchen wir ein Core Konzept, was die Funktion beschreibt wie in (16). Als nächstes werden weitere Konzepte gesucht, die im Text die Core Konzepte modifizieren (s. (17) & (18)). Für die Einbettung der Spanne “konzentrisch” können zwei Alternativen gewählt werden: In (19) wird die Spanne “konzentrisch hypertropher” mit dem SNOMED CT Konzept “SCTID: 34344004 | Concentric hypertrophy (morphologic abnormality)” überlappend mit der Spanne “hypertropher linker Ventrikel” annotiert. In (20) wird hingegen kurzspannig “konzentrisch” mit dem Konzept “SCTID: 255465008 | Concentric (qualifier value)” annotiert.

- (14) Normalgroßer, konzentrisch hypertropher linker Ventrikel mit noch guter LV-Funktion.
- (15) Normalgroßer konzentrisch [**hypertropher linker Ventrikel**](a.) mit noch guter LV-Funktion.
  - a. SCTID: 55827005 | Left ventricular hypertrophy (disorder)
- (16) Normalgroßer konzentrisch [**hypertropher linker Ventrikel**](a.) mit noch guter [**LV-Funktion**](b.).
  - a. SCTID: 55827005 | Left ventricular hypertrophy (disorder)
  - b. SCTID: 366188009 | Finding of left ventricular function (finding)
- (17) Normalgroßer konzentrisch [**hypertropher linker Ventrikel**](a.) mit noch [**guter**](b.) [**LV-Funktion**](c.).
  - a. SCTID: 55827005 | Left ventricular hypertrophy (disorder)
  - b. SCTID: 20572008 | Good (qualifier value)
  - c. SCTID: 366188009 | Finding of left ventricular function (finding)
- (18) [**Normalgroßer**](a.) konzentrisch [**hypertropher linker Ventrikel**](b.) mit noch [**guter**](c.) [**LV-Funktion**](d.).
  - a. SCTID: 53461003 | Normal size (finding)
  - b. SCTID: 55827005 | Left ventricular hypertrophy (disorder)
  - c. SCTID: 20572008 | Good (qualifier value)
  - d. SCTID: 366188009 | Finding of left ventricular function (finding)

- (19) [Normalgroßer](a.) [konzentrisch [hypertropher](b.) linker Ventrikel](c.) mit noch [guter](d.) [LV-Funktion](e.).
- a. SCTID: 53461003 | Normal size (finding)
  - b. SCTID: 34344004 | Concentric hypertrophy (morphologic abnormality)
  - c. SCTID: 55827005 | Left ventricular hypertrophy (disorder)
  - d. SCTID: 20572008 | Good (qualifier value)
  - e. SCTID: 366188009 | Finding of left ventricular function (finding)
- (20) [Normalgroßer](a.) [konzentrisch](b.) [hypertropher linker Ventrikel](c.) mit noch [guter](d.) [LV-Funktion](e.).
- a. SCTID: 53461003 | Normal size (finding)
  - b. SCTID: 255465008 | Concentric (qualifier value)
  - c. SCTID: 55827005 | Left ventricular hypertrophy (disorder)
  - d. SCTID: 20572008 | Good (qualifier value)
  - e. SCTID: 366188009 | Finding of left ventricular function (finding)

## 2.3 Linguistische Besonderheiten

**1. (Syntaktische) Koordination** liegt vor, wenn zwei oder mehr sprachliche Einheiten z.B. durch eine Konjunktion oder eine Aufzählung getrennt positioniert sind, aber dieselbe syntaktische Funktion erfüllen. Koordinierte Einheiten sind durch Trennzeichen oder Junktionen räumlich getrennt, **gelten als nicht-adjazent und sollen daher grundsätzlich getrennt voneinander annotiert werden.**

In Bsp. (21) lässt sich erkennen, dass die letzte koordinierte Einheit (“Ohren”) und das Koordinationsprädikat (“drittgradig verbrannt”) trotz Adjazenz isoliert annotiert werden sollen. Hintergrund für diese Anweisung ist die Gleichbehandlung aller koordinierten Einheiten. Bei der später eingeführten Relationsannotation wird darauf hingewiesen, dass koordinierte Einheiten mit ihrem jeweiligen Koordinationsprädikat über eine Relation verbunden werden sollen, um ihnen inhaltlich gerecht zu werden.

- (21) [Kopf](a.) und [Ohren](b.) [drittgradig verbrannt](c.).
- a. SCTID: 69536005 | Head structure (body structure)
  - b. SCTID: 117590005 | Ear structure (body structure)
  - c. SCTID: 1403192003 | Full thickness burn (disorder)

Koordinierte Einheiten und ihr Koordinationsprädikat können in solchen Fällen ausnahmsweise als eine Spanne annotiert werden, wenn es ein SNOMED CT Konzept gibt, das die Textspanne genau beschreibt, wie beispielsweise in (22).

- (22) Es gibt eine [Verbrennung 3. Grades am Kopf und Nacken](a.).
- a. SCTID: 741070003 | Full thickness burn of head and neck (disorder)

**2. Kopula** sind Wörter oder Phrasen, die zwei Satzglieder syntaktisch miteinander verbinden, aber keine semantische Eigenbedeutung haben. Ein typisches Beispiel sind flektierte Formen des Verbs “sein”. **Aufgrund des fehlenden Bedeutungsbeitrags gelten durch Kopula getrennte Ausdrücke als adjazent und sollen gemeinsam anno-**

**tiert werden.** In (23) kann man sehen, dass mit Ausnahme des Determinierers “der” der gesamte Satz dem Konzept “SCTID: 287139006 | Epidermal burn of head (disorder)” zugeordnet werden soll.

- (23) Der [**Kopf ist erstgradig verbrannt**](a.).  
a. SCTID: 287139006 | Epidermal burn of head (disorder)

**3. Partikelverben und sonstige Einschübe.** Partikelverben sind Verben, die einen Partikel, z.B. eine Präposition, enthalten (z.B. “aufgeben”). In vielen Fällen kann der Partikel bei speziellen syntaktischen Formationen vom Hauptverb getrennt werden. Getrennte Partikelverben und ihre Einschübe sollen überlappend annotiert werden. Es ist jedoch wichtig, dass die Einschübe räumlich *komplett* in der äußeren Annotation enthalten sind. Ein Beispiel findet sich in (24), wo der Einschub (“nach der OP”) komplett in der Matrixannotation (“hörte [...] mit dem Rauchen auf”) enthalten ist.

- (24) Er [**hörte [nach der OP](b.) mit dem Rauchen auf**](a.).  
a. SCTID: 160617001 | Stopped smoking (finding)  
b. SCTID: 262061000 | Postoperative period (qualifier value)

### 3 Annotation von Konzepten

Die Annotation von Textspannen anhand der SNOMED CT Konzepte ist die Grundlage der semantischen Annotation in GeMTeX. Das restliche Kapitel widmet sich der formalen Beschreibung der Konzeptannotation (Kap. 3.1), einem kurzen Abschnitt zu den SNOMED CT Semantic Tags (Kap. 3.2) und der detaillierten Beschreibung relevanter SNOMED CT Konzepte (Kap. 3.3). Abschnitt 3.4 widmet sich dem Umgang mit klinischen Erscheinungen von Familienangehörigen.

#### 3.1 Semantik der Konzeptannotation

Die Annotation von Textspannen folgt einer simplen Regel: Einer möglichst spezifischen Textspanne muss eine möglichst eindeutige und sinnhafte Bedeutung zugewiesen werden. **Bedeutung** definiert sich als in SNOMED CT vorhandenes Konzept. **Eindeutigkeit** bedeutet, dass eine Textspanne  $TS$  nur ein Konzept  $c$  innerhalb der semantischen Domäne  $D_e$  zugeordnet wird, sofern die Textspanne keine Mehrdeutigkeiten aufweist (vgl. Maxime III). Die **semantische Domäne**  $D_e$  definiert die für die Annotation zulässigen Konzepte und denotiert (=bedeutet bzw. beinhaltet) in unserem Fall die Menge aller Konzepte  $C$ , die einer Konzeptgruppe in SNOMED CT angehören, die in dieser Richtlinie diskutiert wird (s. (25)).

- (25)    a.  $C \subseteq \text{SNOMED CT} - \{\text{attributes, meta data, situations with explicit context}\}$   
           b.  $D_e = \{c \mid c \in C\}$

Eine **Textspanne**  $TS$  entspricht der nicht-leeren Menge von sequentiellen Tokens  $t$  aus der Menge aller Tokens  $T$  eines Dokuments (s. (26)). Die kleinste zu annotierende sprachliche Einheit ist stets ein Token - Tokens dürfen nicht “aufgebrochen” werden. Ausnahmen zu dieser Regel bilden z.B. fälschlicherweise zusammengeschriebene Ausdrücke (z.B. “li<re”) oder auch Zahlenwerte und ihre Einheiten. Sollten solche Spannen zusammengeschrieben sein (“3cm”), so darf man sie in ihre Einzeltokens, also Zahlenwert und Einheit teilen (“3” und “cm”).

- (26)    a.  $T = \{t_k \mid k \in \{1, 2, \dots, n\}\}$   
           b.  $TS = \{t_k \mid k \in I\}$  wo gilt:  
               $I = \{s, s+1, \dots, n\} \wedge 1 \leq s \leq n$

Die Konzepte in SNOMED CT sind *hierarchisch* organisiert. Dabei wird von sehr allgemeinen Konzepten in immer spezifischere Konzepte verzweigt. Jedes Konzept, das spezifischere Konzepte unter sich hat, gehört zur Menge der Elternkonzepte  $C_P$ .

Jedes Konzept, das keine weiteren Konzepte unterordnet, gehört zur Menge der Terminalkonzepte  $C_T$ . Beide Mengen sind Teilmengen aller Konzepte  $C$ , die sich gegenseitig ausschließen (s. (27)).

- (27)    a.  $\{C_P, C_T\} \subseteq C$   
           b.  $C_P \cap C_T = \emptyset$

Die Annotationen sollen so spezifisch wie möglich sein, d.h. grundsätzlich sind spezifischere Elternkonzepte vor allgemeineren Elternkonzepten und Terminalkonzepten vor Elternkonzepten vorzuziehen, sofern es der Kontext zulässt und nichts “in die Textspanne hinein” interpretiert wird.

Anhand von Abb. 1 soll die hierarchische Struktur von SNOMED CT (ein gerichteter azyklischer Graph, also eine Polyhierarchie) kurz verdeutlicht werden.

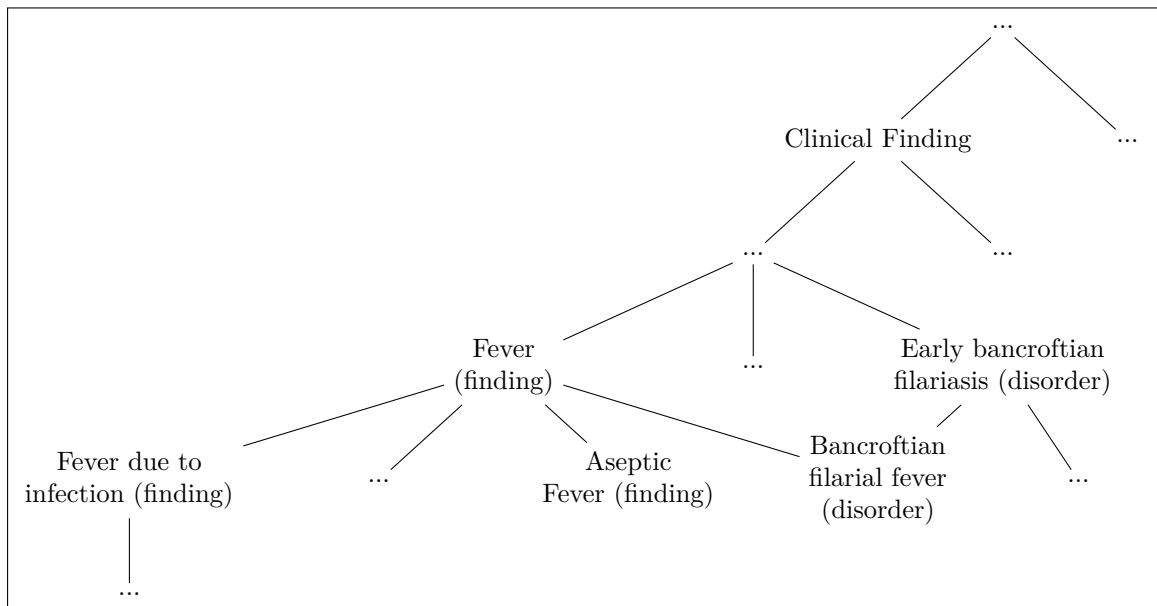


Abbildung 1: Hierarchischer Aufbau von SNOMED CT

Das Elternkonzept “SCTID: 404684003 | Clinical Finding” besitzt ein spezifischeres Tochterkonzept “SCTID: 386661006 | Fever (finding)”. Letzteres ist wiederum ein Elternkonzept, z.B. für das noch spezifischere Konzept “SCTID: 9619006 | Aseptic fever (finding)”. Da letzteres keine weiteren Kinder mehr hat, ist es ein Terminalkonzept. Es wird außerdem deutlich, dass ein SNOMED CT Konzept mehrere Elternkonzepte besitzen kann: “SCTID: 240822009 | Bancroftian filarial fever” ist ein Kind sowohl von “SCTID: 386661006 | Fever (finding)” als auch von “SCTID: 240821002 | Early bancroftian filariasis”.

Auf der Basis von Abb. 1 kann man die in (28) dargestellten Beziehungen zwischen den Konzepten schließen: *Clinical Finding* und *Fever* gehören zu den Elternkonzepten  $C_P$ , während *Aseptic Fever* und *Bancroftian filarial fever* zu der Menge der Terminalkonzepte  $C_T$  gehören. Gleichzeitig sind *Fever* und *Fever due to infection*, *Aseptic fever* und *Bancroftian filarial fever* Kindkonzepte von *Clinical Finding*. Wie schon dargestellt, ist *Bancroftian filarial fever* ein Kind zweier Konzepte: *Fever* und *Early bancroftian filariasis*.

Die Semantik einer annotierten Textspanne  $[TS](C^*)$  definiert sich wie folgt: Zunächst muss eine Textspanne mit mindestens einem Konzept annotiert werden (s. (29)). Weiterhin ist eine Annotation genau dann *zutreffend* oder *korrekt*, wenn das gewählte Konzept  $c$  Teil der semantischen Domäne  $D_e$  ist, inhaltlich mit der markierten Textspanne  $TS$  kompatibel ist und es kein alternatives Konzept  $c'$  in  $D_e$  gibt, das ebenfalls mit  $TS$  kompatibel

und dabei detaillierter ist als  $c$  (s. (30)). Inhaltliche Kompatibilität zwischen einer Textspanne und einem Konzept liegt dann vor, wenn  $TS$  eine Instanz von  $c$  ist, äquivalent zu  $c$  ist oder eine Subklasse von  $c$  darstellt (s. (31)). In allen anderen Fällen, ist die Annotation *unzutreffend* oder *inkorrekt* und somit nicht zu wählen.

- (28) a.  $\{\text{Fever, Clinical Finding, Early bancroftian filariasis}\} \subset C_P$   
 b.  $\{\text{Aseptic Fever, Bancroftian filarial fever}\} \subset C_T$   
 c.  $\{\text{Fever, Fever due to infection, Aseptic Fever, Bancroftian filarial fever}\} \subset \text{Clinical Finding}$   
 d.  $\text{Bancroftian filarial fever} \in \{\text{Fever, Early bancroftian filariasis}\}$
- (29)  $\llbracket [TS](C^*) \rrbracket = \begin{cases} 1 : & C^* \neq \emptyset \\ 0 : & \text{sonst} \end{cases}$
- (30)  $C^* = \{c \in D_e \wedge \text{kompatibel}(TS, c) \wedge \forall c' \in C (\text{kompatibel}(TS, c') \rightarrow \text{detail}(c') \leq \text{detail}(c))\}$
- (31)  $\llbracket \text{kompatibel}(TS, c) \rrbracket = \begin{cases} 1 : & \text{Instanz}(TS, c) \vee \text{Äquivalenz}(TS, c) \vee \text{Subklasse}(TS, c) \\ 0 : & \text{sonst} \end{cases}$   
 a.  $\llbracket \text{Instanz}(TS, c) \rrbracket = TS \in \{x \mid x = \text{spezifisches, individuelles Vorkommen von } c\}$   
 b.  $\llbracket \text{Äquivalenz}(TS, c) \rrbracket = TS \in \{x \mid x \text{ ist bedeutungsgleich zu } c\}$   
 c.  $\llbracket \text{Subklasse}(TS, c) \rrbracket = TS \in \{x \mid x \subset c\}$

#### Hinweis

Die Bedingung in (31-c) erlaubt es, einer Textspanne ein Elternkonzept des (vermutlich) eigentlich korrekten Konzepts zuzuweisen. Somit gilt eine solche Annotation als korrekt, obwohl sie etwas unscharf ist. Es empfiehlt sich, solche Spannen mit der lokalen Annotationsleitung zu diskutieren.

Wäre die Textspanne “Fieber” in Beispiel (32) mit dem Konzept “SCTID: 9619006 | Aseptic fever” annotiert, so wäre die Annotation unzutreffend. Dieses Konzept ist zwar Teil der semantischen Domäne und auch detaillierter als das Konzept “SCTID: 386661006 | Fever (finding)”. Es ist jedoch nicht mit der Textspanne kompatibel, da keine Informationen darüber existieren, ob der Betroffene aseptisches Fieber hat oder nicht. Formal trifft auch keine der in (31) eingeführten Kompatibilitätsbedingungen zu: Die Textspanne “Fieber” ist keine Instanz von “SCTID: 9619006 | Aseptic fever”, da “Fieber” kein spezifisches oder individuelles Vorkommen des Konzepts ist, sondern eine sehr generelle Beschreibung eines medizinischen Zustands. Textspanne und Konzept sind nicht äquivalent, da sie nicht bedeutungsgleich sind. Zuletzt lässt sich noch feststellen, dass “Fieber” auch keine Subklasse bzw. Teilmenge des Konzepts “SCTID: 9619006 | Aseptic fever” darstellt, tatsächlich ist das Gegenteil wahr.<sup>8</sup> Das Beispiel verdeutlicht auch, dass spezi-

<sup>8</sup>Das Beispiel illustriert auch, dass es andersherum weniger problematisch wäre. Würde man für die Textspanne “aseptisches Fieber” das SNOMED CT Konzept “SCTID: 386661006 | Fever (finding)” vergeben, wäre es zwar nicht spezifisch genug. Es würde der Kompatibilitätsbedingung in (31) aufgrund der Definition (31-c) jedoch genügen. Denn: Die Textspanne “aseptisches Fieber” ist eine Subklasse des Konzepts “SCTID: 386661006 | Fever (finding)”.



fischere Konzepte nur dann zu wählen sind, wenn sie auch durch den Text ausgedrückt werden (vgl. Maxime III).

- (32) Der Betroffene hat **[Fieber](a.)**.  
a. SCTID: 386661006 | Fever (finding)

Da die Anzahl an spezifischen Eltern- oder Terminalkonzepten in SNOMED CT immens ist, wird im restlichen Kapitel eine Auswahl an allgemeinen Elternkonzepten bzw. Semantic Tags erläutert und spezifischere Elternkonzepte und Terminalkonzepte werden beispielhaft aufgeführt. Während die in den Richtlinie n aufgeführten Konzepte nicht exhaustiv sind, so gelten die Konzeptgruppen, die eingeführt werden, als Richtschnur für die Annotation. Formal denotieren sie und ihre Tochterkonzepte die semantische Domäne  $D_e$  der GeMTeX-Annotation (vgl. (25)).

## 3.2 SNOMED CT Semantic Tags

Ein elementarer Baustein, um herauszufinden, ob ein SNOMED CT Konzept zur markierten Spanne passt (also *kompatibel* ist), sind die SNOMED CT Semantic Tags. Sie sind immer in Klammern im SNOMED CT FSN zu finden und geben an, zu welcher Konzeptgruppe ein bestimmtes SNOMED CT Konzept gehört. Grundsätzlich kann man hier meist seiner Intuition folgen: Wird beispielsweise im Text über eine vorhandene Krankheit oder eine Diagnose gesprochen, sollte das Konzept in der Regel den Semantic Tag “(finding)” oder “(disorder)” vorweisen. Die Tags “(body structure)” oder “(qualifier)” wären hier eher unangebracht.

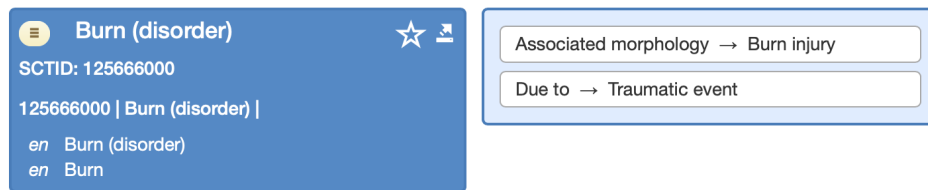


Abbildung 2: Das Konzept “SCTID: 125666000 | Burn (disorder)” im SNOMED CT International SNOMED CT Browser.

Beispiel (33) illustriert, dass Semantic Tags elementar sind, um das korrekte Konzept zu finden: Es gibt in SNOMED CT zwei sehr ähnlich klingende Konzepte zur Verbrennung: “SCTID: 125666000 | Burn (disorder)” und “SCTID: 48333001 | Burn injury (morphologic abnormality)”. Der Semantic Tag “(disorder)” zeigt, dass das Konzept (a.) zur Gruppe der *Disorders* (auf deutsch kommt die Umschreibung *abnormale klinische Zustände / Erkrankungen / Störungen* sehr nah) gehört. Wie im folgenden Kapitel 3.3 näher beschrieben, gehört diese Gruppe von Konzepten zu den *Core Konzepten*, die insbesondere dann verwendet werden sollen, wenn Krankheitsbilder oder Diagnosen beschrieben werden. Zweites Konzept gehört zu der Gruppe “(morphologic abnormality)”, was in der Regel für die Beschreibung von z.B. Gewebeveränderungen als Folge einer Verbrennung verwendet werden kann (z.B. in Radiologie- oder Pathologieberichten). Je nach Briefkontext kann der Semantic Tag also dabei helfen, das korrekte Konzept zu identifizieren.

- (33) Der Patient leidet unter einer [[**Verbrennung**](a.)](b.).
- a. SCTID: 125666000 | Burn (disorder)
  - b. SCTID: 48333001 | Burn injury (morphologic abnormality)

### Hinweis

Natürlich reichen die SNOMED CT Semantic Tags nicht aus, um zu bestimmen, ob ein Konzept das richtige ist oder nicht. Weitere Hilfe bietet der SNOMED CT International SNOMED CT Browser: Schaut man sich den Eintrag des Konzepts “SCTID: 125666000 | Burn (disorder)” an, so wird das Konzept “SCTID: 48333001 | Burn injury (morphologic abnormality)” als Axiom für die assoziierte Morphologie angezeigt (s. Abb. 2). Daraus lässt sich schließen, dass letzteres Konzept wirklich nur in den Fällen verwendet werden sollte, wenn tatsächlich von der morphologischen Veränderung die Rede ist.

### 3.3 Konzepte in SNOMED CT

In diesem Abschnitt werden die für die semantische Annotation in GeMTeX drei relevanten Konzepttypen eingeführt: **Core-Konzepte** (Kap. 3.3.1), **Modifier-Konzepte** (Kap. 3.3.2) und **Qualifier-Konzepte** (Kap. 3.3.3). Core-Konzepte gelten als die relevanteste Konzeptgruppe und ihre Entitäten sollten möglichst vollständig annotiert werden. Modifier-Konzepte sind etwas weniger relevant, kodieren aber dennoch wichtige medizinische Informationen. Qualifier-Konzepte tragen am wenigsten Relevanz, eignen sich oft jedoch sehr gut, um Core- oder Modifier-Konzepte zu spezifizieren. Alle Annotationsbeispiele in dieser Rubrik sind *nicht vollständig*, sondern fokussieren jeweils das Konzept, das illustriert werden soll. Vollständig annotierte Satzbeispiele lassen sich in Kap. 6 finden.

#### 3.3.1 Core-Konzepte

Core-Konzepte umfassen die größten und klinisch relevantesten Konzepte und gliedern sich in folgende vier Kategorien: **Condition**, **Procedure**, **Medikationen & Substanzen** sowie **Observables**. Core-Konzepte sind in der Regel voll definiert, beinhalten somit bereits viele hinreichende und notwendige Bedingungen und sind daher der Grundbaustein für die semantische Annotation.

##### 3.3.1.1 Condition

---

###### SNOMED CT Definition

---

|**Condition**| steht für gesundheitsrelevante Phänomene, Ereignisse, Prozesse oder Zustände im weitesten Sinn. In SNOMED CT sind das die Inhalte der Hierarchien unter “SCTID: 404684003 | Clinical finding (finding)”, “SCTID: 64572001 | Disease (disorder)” und “SCTID: 272379006 | Event (event)”. Ausgeschlossen sind allerdings die, die durch Personen des Gesundheitswesens willentlich ausgelöst werden (s. Procedures).

---

|**Clinical finding**| repräsentiert das Ergebnis einer klinischen Beobachtung, Einschätzung oder Beurteilung. Es enthält normale und abnormale klinische Zustände (z.B. |asthma|, |headache|, |normal breath sounds|). Die Hierarchie unter |clinical finding| enthält Konzepte, die für Diagnosen zu verwenden sind.

---

|**Disorders**| sind permanente oder vorübergehende abnormale klinische Zustände mit zugrundeliegendem pathologischem Prozess, die ggf. unter Behandlung stehen.

---

|**Events**| sind Vorkommnisse, die die Gesundheit oder die Versorgung beeinflussen, allerdings - wie eingangs schon erwähnt - keine Prozeduren oder Interventionen.

---

**Verwendbare Semantic Tags:** (*finding*), (*disorder*), (*event*)

---

Condition umfasst *Clinical Findings*, *Disorders* und *Events*. Damit beinhaltet *Condition* alles, was einen Patienten in seinem Organismus oder seinem Lebensumfeld charakterisiert. Dazu gehören Krankheiten, Symptome, Befunde oder Phänotypen. Zu Condition gehört umgangssprachlich also *alles, was jemand hat oder mit jemanden geschieht* mit der Ausnahme von Procedures (s. 3.3.1.2). Je ein Beispiel für ein Clinical Finding, Disorder und Event finden sich in den Beispielen (34), (35) und (36).

- (34) Sympomatik: [**Anarthrie**](a.).  
 a. SCTID: 48257004 | Anarthria (finding)
- (35) Pat. leidet unter [**spastischer Paraplegie Typ 2**] (a.).  
 a. SCTID: 723622007 | X-linked spastic paraplegia type 2 (disorder)
- (36) Betroffene wurde [**von Hund gebissen**](a.).  
 a. SCTID: 217697000 | Dog bite (event)

### 3.3.1.2 Procedure

---

#### SNOMED CT Definition

---

|**Procedure**| repräsentiert Maßnahmen oder Prozeduren der Krankenversorgung. Hierunter fallen nicht nur invasive Prozeduren, sondern auch die Verabreichung von Arzneimitteln, bildgebende Verfahren, Schulungen, Therapien und administrative Prozeduren (z.B. |appendectomy|, |physiotherapy|, |subcutaneous injection|). In SNOMED CT finden sich Prozeduren unter “SCTID: 71388002 | Procedure (procedure)”

---

**Verwendbare Semantic Tags:** *(procedure), (regime/therapy)*

---

Procedures sind alle Untersuchungen und Maßnahmen, die einem Patienten widerfahren (s. (37)).

- (37) Nach erfolgreicher [**Blutabnahme**](a.) wurde die [**Implantation**](b.) durchgeführt.  
 a. SCTID: 82078001 | Collection of blood specimen for laboratory (procedure)  
 b. SCTID: 782902008 | Implantation procedure (procedure)

### 3.3.1.3 Pharmaceutical/Biological Product & Substance

---

#### SNOMED CT Definition

---

|**Pharmaceutical/biological product**| repräsentiert pharmazeutische Produkte (z.B. |amoxicillin 250mg capsule|, |paracetamol + codeine tablet|)

---

|**Substance**| repräsentiert allgemeine Substanzen, die chemischen Bestandteile von pharmazeutischen und biologischen Produkten, Körpersubstanzen, Nahrungsbestandteile und Diagnostika (z.B. |methane|, |insulin|, |albumin|)

---

**Verwendbare Semantic Tags:** *(product), (medicinal product), (clinical drug), (medicinal product form), (substance)*

---

Medikationen und anderweitige Substanzen werden als Pharmaceutical/Biological Product oder Substance annotiert. Ein Element der Kategorie Pharmaceutical/Biological Product kann informell definiert werden als *alles, was einen Beipackzettel hat*. Pharmazeutische Produkte sollen grundsätzlich anhand ihrer Wirkstoffe mit dem SNOMED CT Semantic Tag “(medicinal product)” annotiert werden. Diese Konzepte finden sich unter “SCTID: 763158003 | Medicinal product (product)” und sind stets nach dem Schema

“Product containing X, Y, ...” aufgebaut. In manchen Fällen ist die Entscheidung hier sehr einfach, s. Bsp. (38).

- (38) Die Pt.in nahm selbstständig [**Aspirin**](a.) ein.  
a. SCTID: 774656009 | Product containing only aspirin (medicinal product)

In anderen Situationen kann die Annotation schwieriger sein. Die Produktnamen *Voltaren* oder *Vokanamet* sind nicht in SNOMED CT enthalten. Auch solche Spannen sollen mit einem Konzept annotiert werden, das den Semantic Tag “(medicinal product)” trägt. Da Voltaren aus dem Wirkstoff *Diclofenac* besteht und Vokanamet aus *Canagliflozin* und *Metformin*, sind die Spannen mit den in (39) und (40) dargestellten Konzepten hinreichend genau annotiert.

- (39) Der Patientin wird [**Voltaren**](a.) verschrieben.  
a. SCTID: 3775563008 | Product containing only diclofenac (medicinal product)
- (40) Wir verschreiben [**Vokanamet**](a.).  
a. SCTID: 775026006 | Product containing only canagliflozin and metformin (medicinal product)

#### Hinweis

Sollte ein Produkt aus mehreren Wirkstoffen bestehen und es gibt kein passendes präkoordiniertes Konzept in SNOMED CT, so werden die beiden Wirkstoffe gestackt. Dabei ist darauf zu achten, dass die gestackten Konzepte **nicht** das Wort *only* im Deskriptor tragen.

Die Annotation von Textspannen mit einem Konzept mit dem Semantic Tag “(substance)” soll lediglich dann erfolgen, wenn eine Textspanne im Briefkontext nicht als Produkt, sondern explizit als Substanz verstanden werden soll (vgl. (41)).

- (41) Die Analyse zeigte, dass [**Methylprednisolon**](a.) im Blut der Pt. war.  
a. SCTID: 116593003 | Methylprednisolone (substance)

Angaben zu Mengen, Dosierungs- oder Vergabeformen sollen **nicht durch die Annotatoren**, sondern lediglich durch die maschinelle Vorannotation annotiert werden. Diese Briefbereiche sollen mit dem No\_Human-Layer in INCEpTION markiert werden (s. Abschnitt 4).

### 3.3.1.4 Observables

---

#### SNOMED CT Definition

---

|**Observable entity**| repräsentiert eine Frage oder Untersuchung, die zu einer Antwort oder einem Ergebnis führen kann, (z.B. |systolic blood pressure|, |color of iris|, |gender|)

---

|**Staging and scales**| repräsentiert Bewertungsschemata und Tumorstadiensysteme (z.B. |Glasgow ComaScale|, |FIGO staging system of gynecological malignancy|)

---

**Verwendbare Semantic Tags:** (*observable entity*), (*staging scale*), (*assessment scale*), (*tumor staging*), (*finding*)

---

Unter Observables fallen alle Untersuchungen oder erfragbaren und beobachtbaren Parameter, die in einem definierten Ergebnis oder Messwerten resultieren und oft auch eine Einheit besitzen. Staging and Scales beschreibt und umfasst die Bewertungen von Krankheitsstadien und Schwere- bzw. Risikograden und so auch verschiedene Testskalen (*assessment scales*). Textspannen, die Scores enthalten, sollen grundsätzlich als *observable entities* und nicht als *assessment scale* annotiert werden, da viele Skalen auf mehr als einen Testscore mappen.

Ein Beispiel ist der *Aachen Aphasie Test*: Das Konzept “SCTID: 273251005 | Aachen aphasia test (assessment scale)” ist eine Skala, die den Skalenparameter “SCTID: 714474007 | Aachen Aphasia Test score (observable entity)” enthält. Nur letzterer wird für die Annotation verwendet, da dieser präziser ist (vgl. auch (42)).

- (42) Das [Geschlecht](a.), die [Sauerstoffkonzentration](b.) und die Ergebnisse des [Glasgow Coma Score](c.) des Patienten werden nachstehend aufgelistet.
- SCTID: 263495000 | Gender (observable entity)
  - SCTID: 250772006 | Oxygen concentration (observable entity)
  - SCTID: 248241002 | Glasgow coma score (observable entity)

Es ist auch zu beachten, dass es zu vielen Observable-Konzepten gleichnamige Finding-Konzepte gibt, z.B. zu “SCTID: 75367002 | Blood pressure (observable entity)” das Konzept “SCTID: 392570002 | Blood pressure finding (finding)”. Finding-Konzepten ist nur dann Vorrang vor Observable Entity Konzepten zu geben, wenn sie den im Text beschriebenen Sachverhalt genauer abbilden. Das kann zum Beispiel der Fall sein, wenn das Finding-Konzept eine Kombination aus Observable Entity Konzept und Qualifier Konzept ist (z.B. “SCTID: 81010002 | Decreased systolic arterial pressure (finding)”). Eine Faustregel ist: Steht die Messung eines bestimmten Wertes im Vordergrund, so wählt man das Observable; Ist das Messergebnis *interpretiert*, so wählt man das Finding.

Allgemein gilt: Für alle Annotationen mit einem Observable-Konzept gibt es im Text in der Regel einen qualitativen oder quantitativen Wert, in letzterem Fall meist auch eine Einheit (vgl. (43)). Für viele Beobachtungen gibt es keine Observable-Konzepte, insbesondere bei Laboruntersuchungen. Hier schreibt SNOMED CT vor, ersatzweise passende Procedure-Konzepte (unter “SCTID: 386053000 | Evaluation procedure (procedure)”) zu verwenden. Beispielsweise: Bei der Textspanne “Leukozyten 8000 / µl” (als Teil des Blutbildes) wird “Leukozyten” mit “SCTID: 767002 | White blood cell count (procedu-

re)” annotiert. Bei der Spanne “Glucose im Urin 15 mg/dl” wird “Glucose im Urin” mit “SCTID: 30994003 | Glucose measurement, urine (procedure)” annotiert.

- (43) Der [**Blutzuckerwert**](a.) des Patienten betrug [**135**](b.) [**mg/dl**](c.) bei der morgendlichen Messung
- a. SCTID: 434912009 | Blood glucose concentration (observable entity)
  - b. **135.0**
  - c. SCTID: 258797006 | Milligram/deciliter (qualifier value)

### 3.3.2 Modifier-Konzepte

Modifikatoren beschreiben Kernkonzepte näher und beantworten damit z.B. welche anatomische Struktur involviert ist, welche Geräte dabei zum Einsatz kommen oder welche Organismen an einer Erkrankung beteiligt sind, aber auch Substanzen und externe Faktoren wie elektrische Spannung oder Strahlung. Für die Annotation sind folgende Modifier-Konzepte definiert: **Body Structures**, **Physical Objects** und **Organisms**.

#### 3.3.2.1 Body Structure

---

**SNOMED CT Definition**

---

|**Body structure**| repräsentiert normale und abnorme anatomischen Strukturen (z.B. |mitral valve structure|, |adenosarcoma|)

---

**Verwendbare Semantic Tags:** (*body structure*), (*morphologic abnormality*)

---

Body Structures sind alle Teile des Körpers, von großen Strukturen wie dem Kopf oder Organen wie dem Gehirn, bis hin zu Gewebe. Wie in den Grundregeln zur Annotation beschrieben, soll die Annotation genau so spezifisch sein, wie es der Text zulässt (s. (44)). Körperteile sollen stets als *Body Structure*, bei pathologischen Befunden ggf. als *Specimen* annotiert werden. Die meisten anatomischen Strukturen finden sich in SNOMED CT doppelt, z.B. “SCTID: 56459004 | Foot structure (body structure)” und “SCTID: 302545001 | Entire foot (body structure)”. Ersteres bezieht sich auf die benannten Strukturen, aber auch beliebige Teile davon. Generell ist bei Lokalisationsangaben der ersteren Sorte von Konzepten (mit “*structure*” im Deskriptor) der Vorzug zu geben. Die Konzepte, die “*entire*” beinhalten, sind nur dann zu verwenden, wenn auch tatsächlich eine gesamte Körperstruktur erfasst wird. Das kann z.B. bei Amputationen wichtig sein. So meint eine Amputation *am* Fuß wohl eine Amputation eines beliebigen Teils der gesamten Fußstruktur (also: “SCTID: 56459004 | Foot structure (body structure)”). Die Amputation *des* Fußes beschreibt jedoch die Amputation des gesamten Fußes (also: “SCTID: 302545001 | Entire foot (body structure)”). In den meisten Fällen erübrigt sich aber die Annotation anatomischer Strukturen, weil diese in Core-Konzepten enthalten sind (z.B.: “SCTID: 723312009 | Amputation of right foot (procedure)”).

- (44) Wir untersuchten [**Kopf**](a.) und [**Haut**](b.).
- SCTID: 69536005 | Head structure (body structure)
  - SCTID: 39937001 | Skin structure (body structure)

#### 3.3.2.2 Physical Object

---

**SNOMED CT Definition**

---

|**Physical object**| repräsentiert natürliche und durch den Menschen hergestellte materielle Gegenstände(z.B. |vena cava filter|, |implant device|, |automobile|)

---

**Verwendbare Semantic Tags:** (*physical object*)

---

Geräte und andere Gegenstände, die transplantiert, eingesetzt oder anderweitig in Diagnostik und Therapie verwendet werden können, fallen unter das Elternkonzept **Physical Object** (s. (45-a)). In gewissen Fällen muss auf die Formulierung geachtet werden:



In (45-c) ist der Herzschrittmacher tatsächlich in der Patientin und sollte deswegen als Finding ausgelegt werden. In (46) handelt es sich nur um eine Empfehlung zu dem Gerät und es sollte daher als *Physical Object* annotiert werden. Bei der Spirale verhält es sich parallel: Die Spirale in Beispiel (45) ist in der Patientin vorhanden und müsste daher als Finding ausgelegt werden. Es gibt allerdings kein solches Konzept in SNOMED CT. Daher sollte auf die SNOMED CT Axiome von verwandten Termen zurückgegriffen werden. Abb. 3 zeigt, dass das Konzept (45-c) u.a. das Axiom “hat Interpretation → Present” aufweist. Klickt man auf dieses Axiom, gelangt man zum Qualifierkonzept “SCTID: 52101004 | Present (qualifier value)”. Dieses Konzept sollte also über die Textspanne als präzisierende Annotation gestackt werden (s. Annotationsmaxime IV; mehr zu Qualifierkonzepten findet sich in Abschnitt 3.3.3).<sup>9</sup>

- (45) Die Patientin besitzt [[**Spirale**](a.)](b.) und [**Herzschrittmacher**](c.).
- a. SCTID: 312082008 | Contraceptive coil (physical object)
  - b. SCTID: 52101004 | Present (qualifier value)
  - c. SCTID: 441509002 | Cardiac pacemaker in situ (finding)
- (46) Dem Patienten wird zu einem [**Herzschrittmacher**](a.) geraten.
- a. SCTID: 424921004 | Permanent cardiac pacemaker, device (physical object)

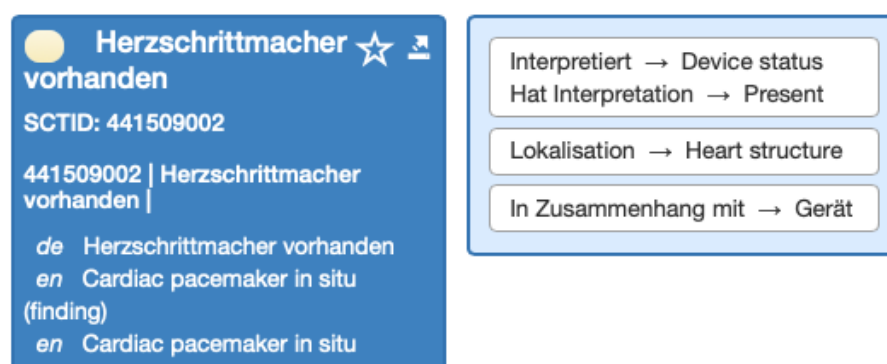


Abbildung 3: Screenshot zu SNOMED CT Axiomen.

### 3.3.2.3 Organismen

---

#### SNOMED CT Definition

---

|**Organism**| repräsentiert Organismen, die in der Human- oder Veterinärmedizin von Bedeutung sind. (z.B. |streptococcus pyogenes|, |beagle|, |texon cattle breed|)

---

**Verwendbare Semantic Tags:** (*organism*)

---

**Organisms** beschreiben alle Arten von Organismen wie Bakterien, Viren, Pilze und Parasiten (s. (47)). Darüber hinaus können, wie in (48), auch größere Organismen vorkommen. Es gilt: Sollten spezifischere Informationen im Text vorhanden sein, wie z.B.

<sup>9</sup>Es wird deutlich, dass Physical Objects, die tatsächlich im Körper vorhanden sind, grds. auch mit dem hier genannten Konzept “SCTID: 52101004 | Present (qualifier value)” zusätzlich annotiert werden sollten, sofern es kein präkoordiniertes Konzept gibt.

*Subtypen* oder *genetische Stämme* der Organismen, so sind diese Konzepte den Allgemeinen vorzuziehen.

- (47)    **[Staphylococcus aureus](a.)** wurde nicht festgestellt.
  - a.    SCTID: 3092008 | Staphylococcus aureus (organism)
- (48)    Vater von Pt. besitzt **[Hund](a.)**. Pt. gibt an gebissen worden zu sein.
  - a.    SCTID: 448771007 | Hund

### 3.3.3 Qualifier-Konzepte

---

#### SNOMED CT Definition

---

|**Qualifier value**| repräsentiert Werte einiger Attribute von SNOMED CT, soweit diese Werte keine Subtypen anderer Top-Level-Konzepte sind. (z.B. |left|, |abnormal result|, |severe|)

---

**Verwendbare Semantic Tags:** (*qualifier value*), (*date*)

---

**Qualifiers** sind eine sehr allgemeine Gruppe von Konzepten in SNOMED CT. Unter anderem *Zahlen*, *Einheiten*, *Zeitangaben* und *Negation* fallen unter diese Kategorie. Es finden sich unter |**Qualifier value**| auch Konzepte, bei denen man erwägen könnte, sie allein zur Annotation von Diagnosen oder Prozeduren zu verwenden, wie “SCTID: 260976002 | Presence of infection (qualifier value)” oder “SCTID: 1055212009 | Losing weight process (qualifier value)”. **Conditions oder Procedures sind diesen Konzepten stets vorzuziehen**, letztere sollen nur dann herangezogen werden, wenn es keine passenden Konzepte unter den Core Concepts gibt.

Bei Qualifiers besteht oft die Qual der Wahl, da sich Konzepte sehr ähneln und keine Definitionen vorliegen. Hierzu sollte man immer den Konzepten den Vorzug geben, die in ähnlichen Zusammenhängen in SNOMED CT -Axiomen verwendet werden.

Beispielhaft wird das Attribut “hoch” näher betrachtet. In der Spanne “hohes Schlaganfallrisiko” könnte “hoch” sowohl mit “SCTID: 75540009 | High (qualifier value)”, “SCTID: 35105006 | Increased (qualifier value)”, als auch mit “SCTID: 281302008 | Above reference range (qualifier value)” annotiert werden. Die Suche nach ähnlichen präkoordinierten Konzepten in SNOMED CT wird bei dem Konzept “SCTID: 161640005 | At high risk for heart disease (finding)” enden. Dieses ist per Definition mit “SCTID: 75540009 | High (qualifier value)” verbunden, also würde man letzteres auch für “hohes” in “hohes Schlaganfallrisiko” verwenden. Nimmt man jedoch die Spanne “hohe Aluminiumkonzentration im Blut”, zeigt die Analyse des ähnlichen, aber präkoordinierten Konzeptes “SCTID: 166689004 | Serum potassium level above reference range (finding)” die Verwendung von “SCTID: 281302008 | Above reference range (qualifier value)”. Das spricht dafür das Wort “hohe” in der Spanne “hohe Aluminiumkonzentration im Blut” mit dem Konzept “SCTID: 281302008 | Above reference range (qualifier value)” zu interpretieren.

#### Hinweis

Grundsätzlich gilt: Wir gehen davon aus, dass Konzepte keine besonderen Merkmale aufweisen. Eine Ausnahme bildet die explizite Nennung, dass etwas an dem Messwert oder Befund besonders ist. Sollte in einem klinischen Text etwas explizit als “zu hoch” bezeichnet werden, so ist dies entsprechend kenntlich zu machen: Die Spanne “zu hoch” wird mit dem passenden Konzept annotiert (s. oben). Gleiches gilt auch, wenn etwas explizit als “unauffällig” oder “normal” gelabelt ist: Dieses Wort ist mit dem SNOMED CT Konzept “SCTID: 17621005 | Normal (qualifier value)” zu annotieren.

### 3.3.3.1 Zeitangaben

Alle **Zeit- und Datumsangaben**, z.B. *Aufnahme- und Entlassdaten, Geburtsdaten* oder Angaben, die sich auf *vorhergehende Diagnosen oder Prozeduren* beziehen, sollen annotiert werden (s. (49)). Alle absoluten Zeitangaben sollen als “SCTID: 410672004 | Date property (qualifier value)” annotiert und anschließend normalisiert werden. Dafür wird in INCEpTION im Feld **literal** das markierte Datum in der Form **yyyy-mm-dd** eingetragen. Sollten Teile des Datums nicht bekannt sein, so werden die betroffenen Stellen durch den Buchstaben ‘x’ ersetzt (s. (49-b)). Bei relativen Datumsangaben sind lediglich die zeitlichen Einheiten (z.B. “Tag” oder “Woche”) zu annotieren. (s. (50)). Eine Normalisierung ist hier nicht vorgesehen.

- (49) Fr. Maier (geb. **[10. März 1965](a.)**) wurde im **[Dezember 2021](b.)** entlassen.
- a. SCTID: 410672004 | Date property (qualifier value) **(1965-03-10)**
  - b. SCTID: 410672004 | Date property (qualifier value) **(2021-12-xx)**
- (50) Laut Patient bestehen die Sehstörungen seit **[drei](a.) [Wochen](b.)** und **[6](c.) [Tagen](d.)**.
- a. **(3.0)**
  - b. SCTID: 258705008 | week (qualifier value)
  - c. **(6.0)**
  - d. SCTID: 258703001 | day (qualifier value)

Sollte eine Entität in einer zeitlichen Relation zu einer anderen Entität stehen, so ist dies durch Kindkonzepte von “SCTID: 272112001 | Relative times (qualifier value)” auszudrücken. Wie in Tab. 3 dargestellt, sollen nur solche Temporalrelationen annotiert werden, die auch explizit ausgedrückt werden (d.h., dass wir Nachzeitigkeit nur dann annotieren, wenn es einen entsprechenden sprachlichen Trigger vofinden, z.B. “nachdem”). Außerdem muss nur das Triggerwort mit dem SNOMED CT Konzept annotiert werden. In Abschnitt 5 wird erläutert, dass eine Relation ausgehend von dem Triggerwort gezogen werden muss.

Zeitigkeit	Trigger	SNOMED CT	Beispiel
<b>Vorzeitigkeit</b>	vor, bevor, seit, ...	SCTID: 272113006   Before values (qualifier value)	<b>[Vor]</b> der OP wurde ein EKG durchgeführt.
<b>Gleichzeitigkeit</b>	während, währenddessen, als, ...	SCTID: 272114000   During values (qualifier value)	<b>[Während]</b> der Untersuchung sprang der Patient auf.
<b>Nachzeitigkeit</b>	nachdem, danach, ...	SCTID: 288563008   After values (qualifier value)	<b>[Nachdem]</b> der Katheter entfernt wurde, führten wir weitere Untersuchungen durch.

Tabelle 3: Caption

Sollte eine Behandlung geplant sein, so ist zusätzlich das Konzept “SCTID: 397943006 |

Planned (qualifier value)” zu annotieren. Wie in Beispiel (51) illustriert kann das Konzept über Triggerwörter, wie “in” annotiert werden.

- (51) Die [Dialyse](a.) findet [in](b.) [4](c.) [Wochen](d.) statt.
- a. SCTID: 108241001 | Dialysis procedure (procedure)
  - b. SCTID: 397943006 | Planned (qualifier value)
  - c. (4.0)
  - d. SCTID: 258705008 | week (qualifier value)

### 3.3.3.2 Quantitäten und Numerischer Basistyp

**Quantitäten** umfassen alle *puren* Zahlenwerte, meist im Zusammenhang mit Beobachtungen und Messungen, in den Texten. Grundsätzlich werden numerische Werte und sprachliche Ausdrücke dieser in INCEption annotiert ohne dabei ein spezielles SNOMED CT Konzept auszuwählen. Stattdessen wird ihr Wert als Dezimalzahl in das **literal**-Feld eingetragen (s. (52)-(55)).

- (52) Wir geben [1](a.) mg.
- a. (1.0)
- (53) Laut Patient bestehen die Sehstörungen seit [drei](a.) Wochen und [6](b.) Tagen.
- a. (3.0)
  - b. (6.0)
- (54) Der [Durchschnitt](a.) der [Herzfrequenz](b.) betrug [50](c.) bpm.
- a. SCTID: 255586005 | Mean (qualifier value)
  - b. SCTID: 364075005 | Heart rate (observable entity)
  - c. (50.0)
- (55) [[RR](a.)](b.) in Ruhe [140](c.)/[80](d.) mmHg.
- a. SCTID: 271649006 | Systolic blood pressure (observable entity)
  - b. SCTID: 271650006 | Diastolic blood pressure (observable entity)
  - c. 140.0
  - d. 80.0

Eine Ausnahme bilden ordinalskalierte Zahlenwerte (z.B. “erster” oder “1.”). Ordinalwerte sollen mit spezifischen SNOMED CT Codes der Hierarchie “SCTID: 272070003 | Ordinal number (qualifier value)” annotiert werden (s. (56)).

- (56) Krankheitsstadium im [1.](a.) Zyklus.
- a. SCTID: 255216001 | First (qualifier value)

### 3.3.3.3 Units

Unter diese Kategorie fallen alle möglichen Angaben zu Einheiten, welche grundsätzlich als **Unit of measure** bzw. dessen Kindkonzepte annotiert werden sollen (s. (57)).

Gemeinsam auftretende Einheiten sollen auch als solche annotiert werden (s. (58)).

(57) RR in Ruhe 140/80 [**mmHg**](**a.**).

a. SCTID: 259018001 | Millimeter of mercury (qualifier value)

(58) Wir geben 10 [**mmol/L**](**a.**).

a. SCTID: 258813002 | Millimole/liter (qualifier value)

### 3.3.3.4 Komparative

Es kann vorkommen, dass Lateralitäten oder Werte in Bezug gesetzt und *verglichen* werden. In solchen Fällen können auch die Ungleichheitssymbole verwendet werden. Hierbei kann es zu Verwechslungen zwischen den *Mathematischen Symbolen* unter dem Konzept “SCTID: 276135000 | Mathematical sign (qualifier value)” und den *Topografischen Qualifiern* unter dem Konzept “SCTID: 106233006 | Topographical modifier (qualifier value)” kommen (z.B. “SCTID: 276139006 | Less-than symbol < (qualifier value)” vs. “SCTID: 351726001 | Below (qualifier value)”). Bei den Ungleichheitsymbolen zwischen Werten sind stets die Konzepte der *Mathematischen Symbole* und nicht der *Topografischen Qualifier* zu wählen (s. Bsp. (59)). Letztere sind nur dann wählen, wenn es sich tatsächlich um räumlich-topologische Aussagen handelt (s. Bsp. (60), das axiomatisch mit ‘SCTID: 351726001 | Below (qualifier value)’ verbunden ist).

(59) [KÖF](a.) [<](b.) [1.3](c.).

- a. SCTID: 251010005 | Cardiac valve area (observable entity)
- b. SCTID: 276139006 | Less-than symbol < (qualifier value)
- c. (1.3)

(60) Pt. kommt mit [amputierter linker Extremität unterhalb des Knies](a.).

- a. SCTID: 816116009 | Amputated left lower limb below knee (finding)

Sollte in einem Text jedoch nicht nur eine Ungleichheitsrelation, sondern auch ein expliziter Bezug zu einem *Referenzwert* hergestellt werden, so ist ein Konzept unter den *Referenzwertinterpretationen* (“SCTID: 442705008 | Reference range interpretation value (qualifier value)”) zu wählen (s. Bsp. (61), das axiomatisch mit “SCTID: 281300000 | Below reference range (qualifier value)” verbunden ist).

(61) [Laktosegehalt unter Referenzwert](a.).

- a. SCTID: 124001001 | Lactose below reference range (finding)

### 3.3.3.5 Negation und Faktualität

**Negation:** Jede Art der Negation wird durch das Konzept “SCTID: 410516002 | Known absent (qualifier value)” annotiert. Dies gilt für alle Vorkommnisse von Negationspartikeln (s. (62)). Adjazente Negations- und Faktualitätsmarker sollen gemeinsam mit dem Konzept annotiert werden, das sie modifizieren. Natürlich nur, sofern ein solches präkoordiniertes Konzept in SNOMED CT existiert. Konzepte, die bereits negiert sind, bedürfen keiner gesonderten Negationsannotation (vgl. (63)).

(62) Es ist [keine](a.) weitere Diagnose möglich.

- a. SCTID: 410516002 | Known absent (qualifier value)

(63) Wir können [keine Lähmung](a.) feststellen.

- a. SCTID: 274811005 | Paralysis absent (finding)

**Faktualität: Differentialdiagnosen:** Eine Differentialdiagnose nennt die Erkrankungen mit nahezu identischer Symptomatik zur Verdachtsdiagnose. Die Differentialdiagnosen selbst sind mit ihren jeweiligen SNOMED CT Konzepten zu annotieren. Ein *Trigger*-Wort, das auf die Existenz einer Differentialdiagnose hinweist (z.B. “DD” in (64)) ist mit “SCTID: 47965005 | Differential diagnosis (contextual qualifier) (qualifier value)” zu annotieren.

- (64) Der Patient hat [**Rheuma(a.)**], [**DD(b.)**]: [**Psoriasisarthritis(c.)**].
- a. SCTID: 396332003 | Rheumatism (disorder)
  - b. SCTID: 47965005 | Differential diagnosis (contextual qualifier) (qualifier value)
  - c. SCTID: 156370009 | Psoriatic arthritis (disorder)

**Faktualität: Assertionssicherheit:** Grundsätzlich wird bei der Annotation angenommen, dass Konzepte *weder negiert noch anderweitig in ihrer Faktualität modifiziert sind*, sofern der Text keine expliziten Gegenschlüsse zulässt. Tab. 4 illustriert die zu annotierenden SNOMED CT Konzepte für Faktualitätstrigger zur Assertionssicherheit (= (Un)Sicherheit von Aussagen) im Text. Die Tabelle ist wie eine Skala geordnet: Die Extrema sind mit höchster Sicherheit belegt, einmal im positiven (z.B. eine Erkrankung ist *bestätigt*) und einmal im negativen Sinne (z.B. eine Erkrankung ist *ausgeschlossen*). Dazwischen findet sich eine *positive* (z.B. “wahrscheinlich”) und eine *negative* Stufe (z.B. “unwahrscheinlich”) von Unsicherheit. Im Default-Fall, also der Annahme, dass etwas vorliegt, soll kein Triggerwort annotiert werden. Beispiel (65) zeigt, wie Faktualitäten noch ausgedrückt werden können.

- (65) [**Laut Patient(a.)**] bestand [**Divertikulitis(b.)**].
- a. SCTID: 410592001 | Probably present (qualifier value)
  - b. SCTID: 307496006 | Diverticulitis (disorder)

Sollte ein Negationspartikel direkter Nachbar eines Faktualitätsmarkers sein, so kann ein gemeinsames Konzept verwendet werden, s. (66). Komposita, die gerne im Deutschen vorkommen, erzwingen oft eine gestackte Annotation, s. (67).

- (66) Ein Tonsillenstein ist [**nicht wahrscheinlich(a.)**].
- a. SCTID: 410593006 | Probably not present (qualifier value)
- (67) Es besteht [[**Leukämieverdacht(a.)**](**b.**)].
- a. SCTID: 93143009 | Leukemia, disease (disorder)
  - b. SCTID: 410592001 | Probably present (qualifier value)

**Empfehlungen und Indikationen:** Eine weitere Faktualität ist die *Empfehlung*. Sollte eine Prozedur im Text empfohlen worden sein, so kann sie entweder als *recommended* oder als *indicated* annotiert werden. Während ersteres sich auf Guideline-Empfehlungen bezieht, meint das letztere medizinisch notwendige Empfehlungen (s. Tab. 5). Es gilt: Grundsätzlich sind Entscheidungen und Behandlungen als medizinisch notwendig einzustufen. Lediglich bei expliziten Nennungen von Guidelines o.ä. sollte auf das Konzept



Sicherheit einer Assertion	Faktualitätstrigger	SNOMED CT	Textbeispiel
<b>Sehr Sicher (+)</b>	Status post (St.p.), Zustand nach (Z.n.), bestätigt, eindeutig, zweifellos, sicher, bekannt, ...	SCTID: 410605003   Confirmed present (qualifier value)	Tonsillenstein konnte <b>[bestätigt](.)</b> werden.
<b>Sicher (+)</b> <i>DEFAULT</i>	vorliegen, bestehen, ...	∅	Tonsillenstein liegt vor.
<b>Weniger Sicher (+)</b>	Verdacht auf (V.a.), wahrscheinlich, sicherlich, mit hoher Wahrscheinlichkeit, tendenziell, vermutlich, möglicherweise, eventuell, gegebenenfalls, Konjunktiv, ...	SCTID: 410592001   Probably present (qualifier value)	Tonsillenstein liegt <b>[wahrscheinlich](.)</b> vor.
<b>Weniger Sicher (-)</b>	unwahrscheinlich, unerwartbar, Konjunktiv, ...	SCTID: 410593006   Probably not present (qualifier value)	Ein Tonsillenstein ist <b>[unwahrscheinlich](.)</b> .
<b>(Sehr) Sicher (-)</b>	Ausschluss von, ausgeschlossen, widerlegt, Kein Anhalt von, Kein Nachweis von, ...	SCTID: 410516002   Known absent (qualifier value)	Die Tonsillenstein-diagnose wurde <b>[widerlegt](.)</b> .

Tabelle 4: GeMTeX Faktualitätsskala (Sicherheit) nach SNOMED CT

“SCTID: 897015005 | Recommended (qualifier value)” zurückgegriffen werden.

**Vergangenheit von klinischen Zuständen ode Prozeduren:** Krankheiten oder Prozeduren können in klinischen Dokumenten als *vergangen* oder *vorüber* gekennzeichnet sein oder sie liegen einfach schon etwas in der Vergangenheit. Ein klassisches Triggerwort ist hierfür “Zustand nach”. In einem solchen Fall wird das Triggerwort mit “SCTID: 410513005 | In the past (qualifier value)” annotiert (s. (68)).

- (68) **[[Z.n.](a.)](b.) [Vasektomie](c.): [Unauffällig](d.).**
- SCTID: 410605003 | Confirmed present (qualifier value)
  - SCTID: 410513005 | In the past (qualifier value)
  - SCTID: 22523008 | Vasectomy (procedure)
  - SCTID: 17621005 | Normal (qualifier value)

Konzept	Bedeutung/ Verwendung	Textbeispiel
SCTID: 410535002   Indicated (qualifier value)	<b>DEFAULT:</b> Eine Behandlung ist medizinisch geboten	Die Behandlung wird [ <b>aufgrund der klin. Verf. empfohlen</b> ](.).
SCTID: 897015005   Recommended (qualifier value)	Eine Behandlung ist von Best Practices oder Guidelines empfohlen, aber nicht obligatorisch ( <i>Nur verwenden, wenn explizit Guidelines oder Best-Practices o.ä. genannt werden.</i> )	Eine Grippe-Impfung wird [ <b>von StIKo empfohlen</b> ](.).

Tabelle 5: Faktualitätsmarker zur Empfehlung & Planung.

#### Hinweis

Kindkonzepte von “SCTID: 243796009 | Situation with explicit context (situation)” sind von der GeMTeX Annotation ausgeschlossen und sollen nicht benutzt werden.

### 3.3.3.6 Weitere Qualifier-Konzepte

Es gibt in SNOMED CT noch zahlreiche weitere Qualifiers, die bei der Annotation Anwendung finden können. Ein Beispiel für eine solche Kategorie ist “SCTID: 272099008 | Descriptor (qualifier value)”, die sich in zahlreiche Tochterkonzepte aufgliedert. So findet sich unter dieser Kategorie z.B. das Konzept “SCTID: 14803004 | Transitory (qualifier value)”, das ein zeitlich begrenztes Auftreten eines Findings beschreibt, wie in (69); Oder das Konzept “SCTID: 281302008 | Above reference range (qualifier value)”, das i.d.R. einen Messwert beschreibt, der explizit über einem Referenzwert liegt (s. (70)). Generell soll den Qualifier-Konzepten der Vorzug gegeben werden, die selbst innerhalb von SNOMED CT Konzeptdefinitionen vorkommen. Veranschaulichen lässt sich das mittels des Tabs “References” im SNOMED CT International SNOMED CT Browser.

- (69) Es war ein [**passagerer**](a.) Diabetes Mellitus.  
a. SCTID: 14803004 | Transitory (qualifier value)
- (70) Messung [**erhöht**](a.).  
a. SCTID: 281302008 | Above reference range (qualifier value)

### 3.4 Familiäre Erkrankungen

Es kann vorkommen, dass in einem klinischen Dokument familiäre (Vor-)Erkrankungen aufgezählt werden. Um diese klar von den Erkrankungen des Patienten zu trennen, müssen sie in den Texten gesondert markiert werden.

In GeMTeX sollen grundsätzlich keine Personen annotiert werden. Bei der Familienanamnese machen wir hierbei eine Ausnahme und annotieren Familienmitglieder, sofern sie im Zusammenhang mit einer Erkrankung oder Behandlung genannt werden. Dafür sollen SNOMED CT Konzepte unter “SCTID: 303071001 | Person in the family (person)” verwendet werden (s. (71)). Sollte keine Person genannt, aber dennoch klar sein, dass es sich um familiäre Erkrankungen handelt, so ist das eben genannte generische Konzept “SCTID: 303071001 | Person in the family (person)” zu verwenden (s. (72)). In Kapitel 5 wird verdeutlicht, dass Erkrankung und ihre Zugehörigkeit zur Familie durch eine Relation miteinander verbunden werden sollen.

- (71) Der **[Vater](a.)** des Betroffenen litt unter einem **[Katarakt li.](b.)**.
- a. SCTID: 9947008 | Natural father (person)
  - b. SCTID: 816119002 | Cataract of left eye (disorder)
- (72) **[In der Familie](a.)** der Betroffenen gab es einen Fall von **[Knochenmarkskrebs.](b.)**.
- a. SCTID: 303071001 | Person in the family (person)
  - b. SCTID: 445738007 | Myelodysplastic/myeloproliferative disease (disorder)

#### Hinweis

Auch bei Familienanamnese gilt, dass Kindkonzepte von “SCTID: 243796009 | Situation with explicit context (situation)” nicht für die Annotation zu verwenden sind.

## 4 Das No-Human Layer: Markierung von ausschließlich maschinell annotierten Bereichen

Um den zeitlichen Aufwand der manuellen Annotationsarbeiten gering zu halten, sind vor allem nicht-narrative Bereiche eines klinischen Texts von der Annotation ausgenommen. Folgende Bereiche sollen nicht annotiert werden:

- Lange Aufzählungen von Laborwerten
- Dosierungs- oder Vergabeangaben von Medikamenten
- Lange Aufzählungen von Medikamenten
- Tabellen
- ICD-Codes
- Dokumentstrukturen / Abschnittsüberschriften (z.B. die Überschrift “Diagnose”)

Die Annotationskräfte haben die Wahl zwischen zwei Vorgehensweisen: (A) Sie können existierende Vorannotation löschen oder (B) das sog. **No-Human** Layer über die betreffenden Bereiche ziehen. Beide Varianten sind äquivalent.

**Variante A: Löschung von Vorannotation:** Sollten die oben genannten Teile eines Dokuments vorannotiert sein, so kann die betreffende Vorannotation gelöscht werden. Hierbei ist sicherzustellen, dass alle Konzepte von der jeweiligen Textstelle entfernt wurden. **Variante B: Das No-Human Layer:** Eine Alternative zur Löschung bietet das Ziehen des No-Human Layers, um eindeutig kennzuzeichnen, dass ein Bereich nicht von Menschen überprüft wurde. In INCEPTION muss dazu zunächst oben rechts im Drop-Down Menü von dem Layer **Concept** zum Layer **No\_Human** gewechselt werden (s. Abb. 4).

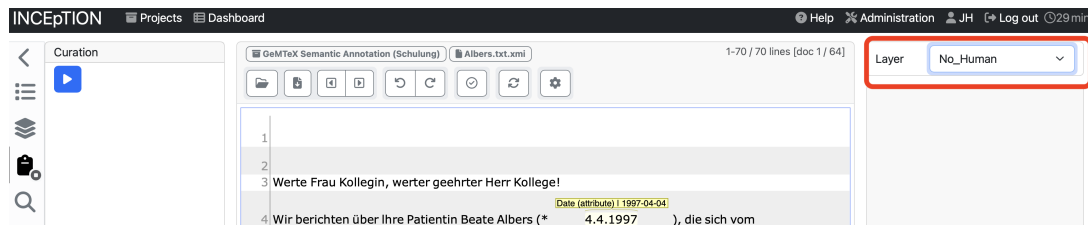


Abbildung 4: Auswahl des No\_Human Layers in INCEPTION.

Es wird geraten, die Varianten situativ effizient einzusetzen: So kann es zum Beispiel sinnvoll sein, eine Vorannotation eines ICD-10 Codes einfach zu löschen, da das in wenigen Klicks erledigt ist. Eine lange Liste von Laborwerten kann andererseits ein guter Fall für das No-Human Layer sein.

## 5 Das Hinzufügen von Relationen

Neben den eingeführten **Konzepten**, die über Spannen im Text annotiert und nach SNO-MED CT gegroundet werden, sollen diese in der semantischen Annotation von GeMTeX auch durch **Relationen** miteinander verbunden werden. Es gibt nur eine generische Art der Relation, hier dargestellt als  $\overrightarrow{r}_{x \in D_e, y \in D_e}$ , die keine Labels besitzt und grundsätzlich nur zwei Annotate  $x$  und  $y$  miteinander verbindet. Im folgenden werden die Grundregeln der semantischen Relationsannotation für GeMTeX erläutert.

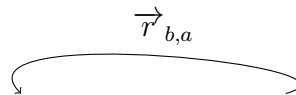
### Hinweis

In den folgenden Abschnitten wird eine jeweils *präferierte Relationsrichtung* angegeben. Dieser sollte grundsätzlich gefolgt werden. Sollte dies zu komplex werden, können Relationen per Default auch von links nach rechts gezogen werden.

### 5.1 Inhaltliche Relationen

Die grundlegendste Bedingung für das Ziehen einer Relation zwischen zwei Annotaten im Text ist die inhaltliche Kohärenz. Zwei Annotate müssen inhaltlich zusammenhängen, sodass eine Relation gezogen werden soll. Inhaltliche Kohärenz meint hier, dass zwei oder mehr Konzepte unbedingt gemeinsam interpretiert werden müssen, um den Sinn hinter der tatsächlichen Aussage im Dokument begreifen zu können.

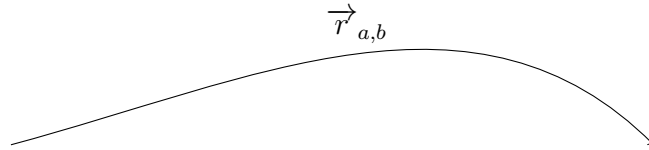
Beispiel (73) illustriert einen solchen Fall: Da es kein gemeinsames Konzept für eine transitorische ischämische Attacke der linken Hemisphäre gibt, müssen die Konzepte getrennt voneinander annotiert werden. Um die Interpretation des Befunds komplett zu ermöglichen, werden die beiden Konzepte durch eine Relation miteinander verbunden. Die präferierte Richtung der Relation geht von den Core Konzepten auf Modifier oder Qualifier Konzepte und von Modifier Konzepten auf Qualifier Konzepte. Sollten zwei gleichrangige Konzepte miteinander verbunden werden, so ist die Relationsrichtung von links nach rechts zu wählen.



- (73) Wir berichten über eine [linkshemisphärielle](a.) [TIA](b.).
- a. SCTID: 72792008 | Left cerebral hemisphere structure (body structure)
  - b. SCTID: 266257000 | Transient ischemic attack (disorder)

## 5.2 Relationen bei familiären Erkrankungen

Sollten Erkrankungen nicht beim Patienten, sondern bei Familienangehörigen vorkommen, so ist die Krankheit mit dem zugehörigen Familienangehörigen zu verbinden. Wie in (74) dargestellt, ist die präferierte Relationsrichtung ausgehend von der Person zur Krankheit.



(74) Der **[Vater](a.)** des Betroffenen litt unter einem **[Katarakt li.](b.)**.

- a. SCTID: 9947008 | Natural father (person)
- b. SCTID: 816119002 | Cataract of left eye (disorder)

## 5.3 Kausalrelationen

Stehen zwei Konzepte in einer offensichtlichen kausalen Beziehung zueinander, muss eine Relation zwischen den beiden gezogen werden. Eine offensichtliche kausale Beziehung liegt dann vor, wenn die Kausalität explizit ausgedrückt wird, z.B. durch eine kausale Konjunktion wie “weil” oder “wegen”. Darüberhinaus muss eines der Konzepte die Ursache für das andere Konzept sein. Die präferierte Relationsrichtung beginnt an der Ursache und endet beim Resultat.



(75) Pt. aufgenommen mit **[akutem Myokardinfarkt](a.)** wg. **[arteriosklerot. Gefäßerkr.](b.)**.

- a. SCTID: 57054005 | Acute myocardial infarction (disorder)
- b. SCTID: 72092001 | Arteriosclerotic vascular disease (disorder)

## 5.4 Anaphorische Relationen

Anaphern sollen stets mit dem Konzept verbunden werden, auf das sie sich beziehen. Das Beispiel (76) zeigt, dass beide Spannen mit demselben SNOMED CT Konzept annotiert wurden. Zusätzlich sollen sie durch eine Relation miteinander verbunden werden. Die präferierte Relationsrichtung geht von der konkreten Nennung auf die anaphorische Referenz.



(76) Pt. wurde **[Voltaren](a.)** verschrieben. Pt. hat **[das Medikament](a.)** nicht genommen.

- a. SCTID: 775563008 | Product containing only diclofenac (medicinal product)

## 5.5 Messwertrelationen

Im Kontext von *Observable Entities* kommen oft Messwerte und Einheiten vor. Diese sind stets in eine Relation miteinander zu bringen. Dabei sollte die Relation vom Observable Entity zu dem Messwert und vom Messwert zur Einheit gerichtet sein (vgl. (77)). Sollte eines der drei Merkmale im Text nicht vorhanden sein, so ist dennoch eine Relation zwischen den verbliebenen zwei Merkmalen zu ziehen.

- (77) Die **[Herzfrequenz](a.)** liegt bei **[55](b.) [bpm](c.)**.
- a. SCTID: 364075005 | Heart rate (observable entity)
  - b. **(55.0)**
  - c. SCTID: 258983007 | Beats/minute (qualifier value)

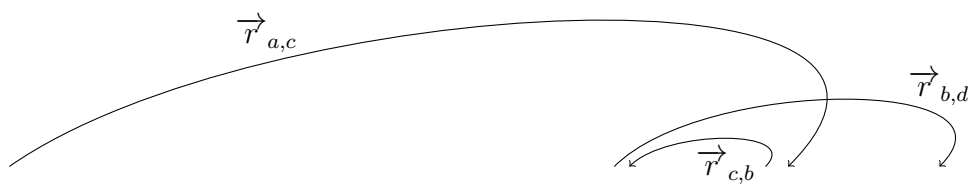
## 5.6 Schrankenrelationen

Sobald ein Wertebereich bzw. eine Spanne zwischen zwei Messwerten, Daten oder auch Graden angegeben ist, muss eine Relation ausgehend von der unteren Schranke zur oberen Schranke gezogen werden (s. Bsp. (78) und (79)).

- (78) Der Patient war von **[23.10.](a.)-[31.10.21](b.)** in stationärer Behandlung.
- a. SCTID: 410672004 | Date property (qualifier value) **(2021-10-23)**
  - b. SCTID: 410672004 | Date property (qualifier value) **(2021-10-31)**

- (79) Sie erlitt **[[Verbrennungen 2.](a.) - 3. Grades](b.)**.
- a. SCTID: 403191005 | Partial thickness burn (disorder)
  - b. SCTID: 403192003 | Full thickness burn (disorder)

Sollte eine Messwertrelation und die Schrankenrelation zusammenfallen, ist die Komplexität wie folgt zu reduzieren: Die erste Relation geht von der Observable Entity zur unteren Schranke. Die zweite Relation zeigt von der unteren zur oberen Schranke. Die letzte Relation zeigt von der oberen Schranke zur Einheit (vgl. (80)).

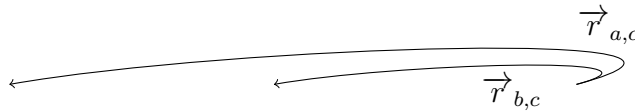


- (80) [Syst. Blutdr.](a.) höchstens zwischen [146](b.) und [120](c.) [mmHg](d.).
- a. SCTID: 271649006 | Systolic blood pressure (observable entity)
  - b. 146.0
  - c. 120.0
  - d. SCTID: 259018001 | Millimeter of mercury (qualifier value)

## 5.7 Koordinationsrelationen

In Kapitel 2 wird im Rahmen von Maxime II auf (syntaktische) Koordination eingegangen. Die Regel definiert, dass koordinierte Einheiten getrennt voneinander und vom Koordinationsprädikat, auf das sie sich beziehen, annotiert werden müssen. Dies ist auch in Beispiel (21) dargestellt, hier in vollständiger Annotation als (81) widergegeben.

Es wird deutlich, dass die koordinierten Einheiten “Kopf” und “Ohren” getrennt voneinander annotiert werden, ebenso wie das Koordinationsprädikat “drittgradig verbrannt”, welches den Sachverhalt der koordinierten Einheiten näher bestimmt. Eine Relation soll nun ausgehend vom Koordinationsprädikat auf die beiden koordinierten Einheiten gezogen werden - es ist nicht notwendig, eine Relation zwischen den koordinierten Einheiten zu ziehen.



- (81) [Kopf](a.) und [Ohren](b.) [drittgradig verbrannt](c.).
- a. SCTID: 69536005 | Head structure (body structure)
  - b. SCTID: 117590005 | Ear structure (body structure)
  - c. SCTID: 1403192003 | Full thickness burn (disorder)

## 5.8 Faktualitätsrelationen

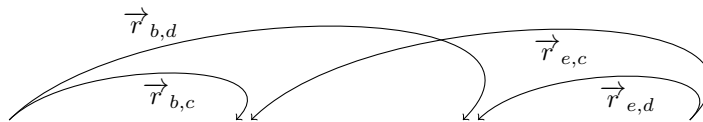
Faktualitätsanzeiger und Negationspartikel modifizieren oder invertieren die Semantik bestimmter Ausdrücke und haben daher massive Auswirkungen auf die korrekte Interpretation von Aussagen. Um diesem Beitrag gerecht zu werden, muss von Faktualitäts- oder Negationspartikeln eine Relation auf jedes Annotat gezogen werden, das von ihnen modifiziert wird. Sollten mehrere Annotate von der Modifizierung betroffen sein, so ist eine Relation zu jedem dieser Annotate zu ziehen. Das Beispiel (82) illustriert die Annotation von Faktualitäten.



- (82) [Wahrscheinlich](a.) liegt eine [Läsion im Auge](b.) vor.
- a. SCTID: 410592001 | Probably present (qualifier value)
  - b. SCTID: 301905003 | Lesion of eye (disorder)



In (83) wird ein Fall illustriert, wo eine Negation mehrere Konzepte modifiziert: Der Satz sagt aus, dass weder intrakranielle noch intracerebrale Traumata vorliegen, ohne zu spezifizieren, ob weitere Traumata vorliegen. Folglich modifiziert die Negation hier die beiden Lokalisationsangaben, die wiederum die Traumata näher spezifizieren. Es ist also von “keine” ausgehend je eine Relation zu den beiden Lokalisationsangaben zu ziehen. Da die beiden Lokalisationsangaben mit dem Konzept der Traumata koordiniert sind, ist weiterhin je eine Relation von “Traumata” zu den beiden Lokalitten zu ziehen.



(83) [Kopf-CT](a.) ergab [keine](b.) [intrakr.](c.) oder [intracer.](d.) [Traumata](e.).

- a. SCTID: 303653007 | Computed tomography of head (procedure)
- b. SCTID: 410516002 | Known absent (qualifier value)
- c. SCTID: 128319008 | Intracranial structure (body structure)
- d. SCTID: 12738006 | Brain structure (body structure)
- e. SCTID: 417746004 | Traumatic injury (disorder)

## 5.9 Temporalrelationen

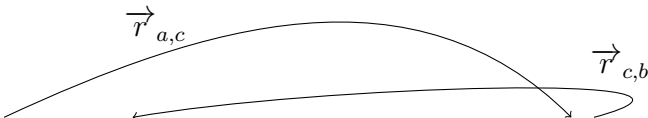
Temporalrelationen knnen ausschlielich durch ein Triggerwort ausgelst werden, das Vor-, Gleich- oder Nachzeitigkeit ausdrckt (vgl. Abschnitt 3.3.3.1). Kommt ein solches Triggerwort vor (z.B. “vor” in Bsp. (84)), so ist eine Relationskette nach dem folgenden Prinzip zu bilden:

- VORZEITIG(Konzept A, Konzept B) und Konzept A findet vor Konzept B statt.
- GLEICHZEITIG(Konzept A, Konzept B) und Konzept A steht vor Konzept B im Text.
- NACHZEITIG(Konzept A, Konzept B) und Konzept A findet nach Konzept B statt.

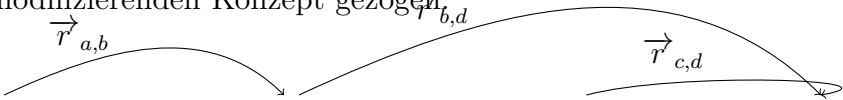
Im Falle der Vorzeitigkeit ist also zunchst eine Relation vom Triggerwort zum frher stattfindenden Konzept zu ziehen. Anschließend ist eine Relation vom frher zum spter stattfindenden Konzept zu ziehen. Bei der Nachzeitigkeit verhlt es sich analog (Also: Der Trigger zeigt auf das spter stattfindende Konzept, welches wiederum auf das frher stattfindende Konzept zeigt). Bei Gleichzeitigkeit entspricht die prferierte Relationsrichtung der Leserichtung (von links nach rechts). Jeweils ein Beispiel fr die Temporalrelation ist in den nachfolgenden Beispielen (84), (85) und (86) dargestellt.

(84) [Vor](a.) der [OP](b.) wird ein [ambulantes EKG](c.) durchgefhrt.

- a. SCTID: 272113006 | Before values (qualifier value)
- b. SCTID: 5387713003 | Surgical procedure (procedure)
- c. SCTID: 164850009 | Ambulatory electrocardiogram (procedure)

- 
- (85) [Nach](a.) der [OP](b.) wird ein [ambulant<sup>es</sup> EKG](c.) durchgeführt.
- a. SCTID: 288563008 | After values (qualifier value)
  - b. SCTID: 5387713003 | Surgical procedure (procedure)
  - c. SCTID: 164850009 | Ambulatory electrocardiogram (procedure)

Das Beispiel (86) zeigt zudem, wie Faktualitätstrigger in eine solche Relationskette einzubinden sind: Wie in Abschnitt 5.8 beschrieben, wird der Faktualitätstrigger nicht in die bestehende Kette eingebunden. Es wird lediglich eine Relation zwischen dem Triggerwort und dem zu modifizierenden Konzept gezogen.

- 
- (86) [Während](a.) der [OP](b.) kam es zu [keinen](c.) [Komplikationen](d.).
- a. SCTID: 288563008 | After values (qualifier value)
  - b. SCTID: 5387713003 | Surgical procedure (procedure)
  - c. SCTID: 410516002 | Known absent (qualifier value)
  - d. SCTID: 738778005 | Intraoperative complication (disorder)

## 6 Annotationspragmatik: Komplette Beispielannotationen und Grenzfälle

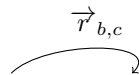
In diesem Abschnitt sollen Beispiele für vollständig annotierte Texte gegeben werden. Die Beispiele sind von unterschiedlicher Komplexität, Länge und Konzeptdichte. Das erste Beispiel (87) zeigt die Annotation einer längeren Textspanne, die jedoch *sinnhaft* ist und durch *adjazente sprachliche Einheiten* miteinander verbunden ist.

- (87) Die Patientin litt unter [**symptomatischem Anfallsleiden**](a.).  
a. SCTID: 230381009 | Focal epilepsy (disorder)

Die Beispielsätze (88) und (89) illustrieren die Regeln von Maxime II sehr deutlich: (88) zeigt einen Satz, der komplett adjazente Token enthält, die eine sinnhafte Einheit ergeben, und damit als eine Gesamtspanne mit einem präzisen SNOMED CT Konzept annotiert werden kann.

Der Satz in (89) enthält zwei Tücken: Die Spanne enthält einen Bereich mit zwei Schranken<sup>10</sup>. Daher müssen zunächst die beiden Schranken durch eine Relation miteinander verbunden werden. Als zweites stellen wir fest, dass das ‘Ohr’ durch ein Kopulaverb ‘war’ mit dem Rest des Satzes verbunden ist. Somit annotieren wir den gesamten Satz mit dem SNOMED CT Konzept “SCTID: 9065001 | Burn of ear (disorder)” und behandeln die Verbrennungsgradangaben als Einschübe (vgl. Maxime II, Kaptiel 2).

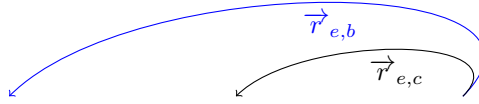
- (88) Das [**Ohr war im II. Grad verbrannt**](a.).  
a. 269233004 | Superficial partial thickness burn of ear (disorder)



- (89) Das [**Ohr war im [II.](b.) bis [III. Grad](c.) verbrannt**](a.).  
a. SCTID: 39065001 | Burn of ear (disorder)  
b. SCTID: 403191005 | Partial thickness burn (disorder)  
c. SCTID: 403192003 | Full thickness burn (disorder)

Das Beispiel in (90) beinhaltet eine Aufzählung von medizinischen Prozeduren. Das Datum “01/99” muss in keine Relation eingebunden werden. Die Lokalisationsangabe “li. Handrücken” ist das Koordinationsprädikat der beiden koordinierten Einheiten “VAC-Wechsel” und “Debridement”. Nach der Koordinationsregel müssen beide koordinierte Einheiten mit ihrem Koordinationsprädikat verbunden werden (präferiert: ausgehend vom Koordinationsprädikat).

<sup>10</sup>Anmerkung: Verbrennungsgrade sind kein richtiger Wertebereich, da sie lediglich kategoriell geordnet sind. Daher wird diese auch nicht mit Zahlenwerten, sondern mit ihren tatsächlichen Bedeutungen als SNOMED CT Konzept annotiert. Da sprachlich durch das Wort “bis” ein Bereich indiziert wird, greift hier die Schrankenregelung.



- (90) [1/99](a.): [VAC-Wechsel](b.) u. [Debridement](c.) [li. Handrücken](d.).
- a. SCTID: 410672004 | Date property (qualifier value) (**1999-01-xx**)
  - b. SCTID: 1217562000 | Replacement of negative pressure wound dressing (procedure)
  - c. SCTID: 36777000 | Debridement (procedure)
  - d. SCTID: 789218009 | Structure of dorsum of left hand (body structure)

Das nachstehende Beispiel illustriert, wie mit Differentialdiagnosen umzugehen ist. Das Triggerwort “DD” wird mit dem im Richtlinie vorgeschlagenen SNOMED CT Konzept annotiert. Die Originaldiagnose und die Differentialdiagnose, “Rheuma” und “Psoriasisarthritis”, werden mit den jeweilig passenden Konzepten annotiert. Das Triggerwort zeigt auf das Element, das es modifiziert, hier also “Psoriasisarthritis”.



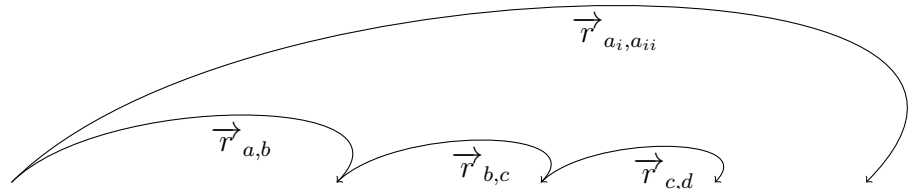
- (91) Der Patient hat [Rheuma(a.)], [DD](b.): [Psoriasisarthritis](c.).
- a. SCTID: 396332003 | Rheumatism (disorder)
  - b. SCTID: 47965005 | Differential diagnosis (contextual qualifier) (qualifier value)
  - c. SCTID: 156370009 | Psoriatic arthritis (disorder)

In (92) wird erneut ein Faktualitätsbeispiel, jedoch mit Relationsrichtung entgegen der Leserichtung veranschaulicht. Die Relation zeigt immer von einem Faktualitätsanzeiger auf dasjenige Annotat, das modifiziert wird.



- (92) Ein [erhöhtes Schlaganfallrisiko](b.) wird [vermutet](b.).
- a. SCTID: 866240007 | At increased risk of cerebrovascular accident (finding)
  - b. SCTID: 410592001 | Probably present (qualifier value)

Eine anaphorische Konstruktion kann in (93) betrachtet werden: “Blutdruckwerte” und “systolisch” beziehen sich auf dasselbe Konzept “SCTID: 271649006 | Systolic blood pressure (observable entity)” und werden daher durch eine Relation miteinander verbunden. Weiterhin geht von der ersten Instanz der Anapher “Blutdruckwerte” eine Relation zu der unteren Schranke der tatsächlich gemessenen Werten “170”. Aufgrund der Schrankenregelung zeigt letzters Konzept auf die obere Schranke “180”, welche wiederum auf die Einheit “mmHg” gerichtet ist.



(93) [Blutdruckwerte](a.) zwischen [170](b.) und [180](c.) [mmHg](d.) [systolisch](a.).

- a. SCTID: 271649006 | Systolic blood pressure (observable entity)
- b. **(170.0)**
- c. **(180.0)**
- d. SCTID: 259018001 | Millimeter of mercury (qualifier value)

## 7 Kurationshinweise

Dieses Kapitel ist für das lokal operierende Kurationspersonal bestimmt und listet kurze Hinweise zu Kuration der semantischen Annotation in GeMTeX auf.

Grundsätzlich erfolgt die semantische Annotation nach dem **2-Augen-Prinzip**. Das bedeutet, dass eine studentische Hilfskraft mit Hilfe der automatischen Vorannotation und des ID Recommenders pro Dokument annotiert. Das steht im Gegensatz zu dem in der De-Identifizierung praktizierten 4-Augen-Prinzip. Um die Alignierung der Annotationskräfte zu messen und die Korrektheit sowie Regelhaftigkeit der semantischen Annotationen regelmäßig zu evaluieren, sollten 1 % der Dokumente eines Batches *von allen Annotationskräften* annotiert werden. Diese sollten auf Inter-Annotator-Agreement Werte und die inhaltliche Korrektheit hin überprüft werden. Darüber hinaus empfiehlt es sich, Textbeispiele in der Runde der Annotationskräfte gemeinsam zu diskutieren. Zumeist bringen die Hilfskräfte medizinisches und die Kurationskräfte technisch-organisatorisches Wissen mit. Erfahrungsgemäß lassen sich Unklarheiten und Grauzonen so gut auflösen. Bei besonders kniffligen Fällen wird empfohlen, die Debatte auf das standortübergreifende Kurationsmeeting auszuweiten.

0