Richtlinien zur Annotation *definitorischer Textsegmente*

Irene Cramer
cramer@hytex.info
Institut für deutsche Sprache und Literatur,
Technische Universität Dortmund
Stand: 2008-08-04

1 Aufgabenstellung 2

1.1 Anwendungsszenario 2

1.2 Beschreibungsebenen von definitorischen Textsegmenten 5

2 Grundlegende Bestandteile 5

2.1 Das Definiendum (Plural: Definienda) 6

2.2 Das Definiens (Plural: Definientia) 7

3 Typische Muster von Definitionen 10

3.1 Muster typischer Informationsstrukturen 10

3.2 Muster typischer Formulierungen 12

4 Begrenzung, Kontext und schwer entscheidbare Fälle 17

4.1 Abgrenzung und Kontext 17

4.2 Anaphorische Strukturen 18

4.3 Eigennamen 19

4.4 Überlappungen und verschränkte Definitionen 19

4.5 Vollständigkeit & Korrektheit 20

4.6 Unauffällige Formen und Kurzformen der Definition 21

4.7 Grenzfälle und Negativbeispiele 21

5 FAQ 22

6 Quellen 25

1 Aufgabenstellung

Die vorliegenden Richtlinien sind eine Zusammenstellung der Merkmale und Kriterien, die bei der Annotation definitorischer Textsegmente berücksichtigt werden sollen.

Definitorische Textsegmente (DEF) im Sinn dieser Richtlinien sind Sätze, die als Ganzes oder in Teilen die Bedeutung von Begriffen erklären.

Jedes Textsegment, das heißt also jeder Satz, wird entsprechend bei der Annotation mit einer der beiden Klassen markiert:

DEF der Satz oder ein Teil des Satzes stellt eine Definition dar;

O weder der Satz noch einer seiner Teile stellt eine Definition dar.

Die Zuordnung ist dabei immer eindeutig: Ein Satz kann nur mit **DEF oder O** markiert werden. Ein Beispiel für diese Art der Annotation ist in Abbildung 1 dargestellt.

Abbildung 1: Beispiel für Art und Format der Annotation definitorischer Textsegmente

```
<s>       DEF     <CASTOR>
In        APPR    in
Deutschland       NE      Deutschland
heißen    VVFIN   heißen
sie       PPER    sie
CASTOR    NN      CASTOR
(         $(      (
das       ART     d
Akronym   NN      Akronym
steht     VVFIN   stehen
für       APPR    für
"         $(      "
CA        NE      CA
sk        ADJA    sk
for       FM      for
Storage   NN      Storage
and       FM      and
```

Außerdem wird hinter die Angabe, ob ein DEF vorliegt oder nicht, notiert, welche Wörter bzw. Satzteile in dem Textsegment erläutert werden
(vgl. Abbildung 1: `<s>  DEF  <CASTOR>`).

1.1 Anwendungsszenario

Die vorliegenden Richtlinien sind im Rahmen der Annotation von Daten im Zusammenhang mit einem bestimmten Anwendungsszenario entstanden, das im Folgenden näher beschrieben werden soll.

In den letzten Jahrzehnten haben Computerlinguisten und Sprachwissenschaftler eine große Menge Sprachdaten als Korpora zusammengestellt. Um Experten, die eine bestimmte wissenschaftliche Fragestellung verfolgen, oder Laien, die sich z.B. für den Wortgebrauch oder die Bedeutung eines Wortes interessieren, den Umgang

mit diesen Daten zu erleichtern (bzw. zu ermöglichen), wurden außerdem in den letzten Jahren einige interessante Abfragemöglichkeiten entwickelt.

Abbildung 2: Screenshot des online-Angebots des DWDS (www.dwds.de) zum Suchwort *Nachhaltigkeit*

Ein Beispiel hierfür sind die Arbeiten, die im Rahmen des Projekts "Digitales Wörterbuch" an der Berlin-Brandenburgischen Akademie der Wissenschaften durchgeführt werden (vgl. online Ressource DWDS www.dwds.de). Abbildung 2 zeigt das Suchergebnis der DWDS Webschnittstelle für das Wort *Nachhaltigkeit*. Wie die Abbildung verdeutlicht, werden zum Suchwort verschiedene Typen von Informationen angeboten, das sind

- links oben der Wörterbucheintrag (falls vorhanden)[1],
- links unten Korpusbelege (chronologisch oder zufällig sortiert),
- rechts oben semantische Relationen (in diesem Fall keine im Korpus auffindbar)[2],
- rechts unten Kollokationen (in diesem Fall die Kombination *Prinzip der Nachhaltigkeit*).

[1] In diesem Fall enthält das Wörterbuch keinen Eintrag zum Wort Nachhaltigkeit, wohl aber zu nachhaltig: […] */Adj./ lange nachwirkend, andauernd, stark* […]. Dieser umfasst allerdings wie man sieht nicht die Bedeutung von *Nachhaltigkeit/nachhaltig* im Sinn des ökologischen Prinzips. Insofern liefern die Korpusbelege, die in Abbildung 3 dargestellt sind zusätzliche, bisher im Wörterbuch nicht erfasste Aspekte.

[2] Semantische Relationen, die hier berücksichtigt werden, sind Hyperonymie und Hyponymie. Für das Wort *Nachhaltigkeit* konnten allerdings offenbar Hyperonyme/Hyponyme berechnet werden, weswegen in diesem Beispiel im rechten oberen Feld keine Einträge enthalten sind.

Die Abbildung zeigt, dass u. a. auch (halb-)automatisch Informationen aus Korpora gewonnen und ergänzend (vgl. rechts oben und unten) zu den manuell erstellten (vgl. links oben) dem Nutzer angeboten werden[3]. Ähnlich der in Abbildung 2 dargestellten Informationen wie z.B. semantische Relationen können natürlich auch weitere Aspekte (halb-)automatisch aus den DWDS-Korpora extrahiert werden. Im Rahmen unserer Forschungsarbeit untersuchen wir daher Verfahren zur automatischen Extraktion definitorischer Textsegmente.

Dabei geht es wie bei den in Abbildung 2 dargestellten semantischen Relationen und Kollokationen auch bei der Extraktion von Definitionen in erster Linie darum, die in den Korpora versteckten Informationen dem Nutzer besser zugänglich zu machen. Dazu sollen automatisch Korpusbelege ermittelt werden, die für den gesuchten Begriff (im Hinblick z.B. auf den Wörterbucheintrag) ergänzende oder komplementäre Informationen liefern. Ein Beispiel hierfür ist das Suchwort *Nachhaltigkeit* (vgl. Abbildung 3). Bisher werden die Korpusbelege des DWDS entweder chronologisch oder nach dem Zufallsprinzip sortiert angezeigt. Es ist offensichtlich, dass das nicht die nutzerfreundlichste Anordnung sein kann. Alternativ also zu einer chronologischen bzw. nach Zufallsprinzip sortierten Liste von Korpustreffern könnte die Liste etwa umsortiert werden, so dass die definitorischen Treffer auf die ersten Positionen rücken. Abbildung 3 zeigt drei solcher Korpusbelege für das Wort *Nachhaltigkeit*, die definitorischen Eigenschaften aufweisen. Die drei Treffer würden im Sinn des Anwendungsszenarios daher an die ersten drei Positionen der Trefferliste rücken und vom Nutzer somit als Ergänzung (oder hier sogar als Ersatz) zum Wörterbucheintrag verwendet werden können.

Abbildung 3: Definitorische Textsegmente aus den DWDS-Korpora zum Suchwort *Nachhaltigkeit*

Wie schon die Waldvernichtung im Altertum führte eine nicht "nachhaltige" Forstwirtschaft zu Erosionen, Wasserarmut und Versteppung. In Zukunft sollte nicht mehr Holz eingeschlagen werden als nachwächst ("Prinzip der **Nachhaltigkeit** "). Wiederaufforstungen, insbesondere in der Form von schützenden Waldstreifen, sichern in Zukunft die Landwirtschaft und halten die Versteppung fern. Damit ist die Holznutzung teilweise zu einer sekundären Frage geworden, und die Wohlfahrtswirkungen können im Vordergrund stehen.
In: Sandermann, Wilhelm, Grundlagen der Chemie und chemischen Technologie des Holzes, Leipzig: Geest & Portig 1956, S. 72

Doch der Begriff der "Nachhaltigkeit" drohte im Streit der Experten unterzugehen - die einen leiten aus ihm die Forderung ab, mit dem Abbau von Rohstoffen ganz aufzuhören, die anderen glauben, daß Rohstoffe im Prinzip durch Kapital ersetzbar sind. Gleichzeitig wurde **Nachhaltigkeit** zum Allerweltswort auf politischen Konferenzen und Jahreshauptversammlungen von Großunternehmen - die ökologischen Ökonomen drohten zu Stichwortgebern für Werbebroschüren zu werden. Um dagegenzuhalten, arbeitet Nutzinger bis heute vor allem an der präzisen Formulierung von Regeln für die Nachhaltigkeit.
In: Thomas Fischermann, Der Umweltökonom, in: DIE ZEIT 08.11.1996, S. 29

Der Umweltökonom Hans-Christoph Binswanger sprach ihn 1980 an, ob er nicht an einem Buchprojekt über Umwelt und Ökonomie mitarbeiten wolle - für Nutzinger damals "völliges Neuland". Als er sich intensiver damit auseinandersetzte, fühlte sich Nutzinger vor allem von der Schule der "ökologischen Ökonomie" angezogen, die der Amerikaner Herman E. Daly mitbegründet hat. Diese Schule hatte den heute geläufigen Grundsatz der " **Nachhaltigkeit** " vorgedacht. Nach diesem Prinzip soll die Ausbeutung der Natur so begrenzt werden, daß sie sich immer wieder regenerieren kann.
In: Thomas Fischermann, Der Umweltökonom, in: DIE ZEIT 08.11.1996, S. 29

Das oben skizzierte Anwendungsszenario bestimmt damit den Rahmen für die Annotation definitorischer Textsegmente, wie sie in den vorliegenden Annotationsrichtlinien beschrieben werden. Die Bedingungen, die sich daraus ergeben, werden in Abschnitt 2 erklärt.

[3] Mit zunehmenden Datenmengen lässt sich zwar einerseits ein genaueres bzw. empirisch fundiertes Bild über bestimmte sprachliche Phänomene zeichnen, andererseits bedeutet die Arbeit mit großen Datenmengen einen erheblichen Zeit-, Kosten- und Arbeitsaufwand. Daher wurde in den letzten Jahren intensiv an Verfahren geforscht, mit deren Hilfe auch große Datenmengen zugänglich gemacht und analysiert werden können.

1.2 Beschreibungsebenen von definitorischen Textsegmenten

Unsere bisherige Forschung hat gezeigt, dass sich mindestens zwei verschiedene Ansatzpunkte bzw. Ebenen zur Beschreibung von Definitionen finden lassen. Beide Blickwinkel ergänzen sich und sind daher hilfreich, um ein gutes Verständnis des Konzepts definitorisches Textsegment zu entwickeln.

Der eine Blickwinkel leuchtet die **grundlegenden Bestandteile** (Abschnitt 2) definitorischer Textsegmente aus. Während der andere den Fokus auf bestimmte **Muster** (Abschnitt 3) legt, die wir in unserer Arbeit mit Korpora bisher besonders häufig im Zusammenhang mit der Suche und Analyse definitorischer Textsegmente beobachtet haben.

2 Grundlegende Bestandteile

Definitorische Textsegmente setzen sich aus zwei grundlegenden Bestandteilen zusammen: Definiendum und Definiens.

1. Das **Definiendum** ist ein Wort oder lexikalisiertes[4] Syntagma (~ eine Wortgruppe, ein Teilbereich eines Satzes), dessen Bedeutung das betrachtete Textsegment erläutert. Das Definiendum ist also der zu erläuternde bzw. zu definierende Bestandteil.
2. Das **Definiens** ist ein Syntagma, das die Bedeutung des Definiendums angibt bzw. erklärt. Damit ist das Definiens also der definierende Bestandteil.

Im Hinblick auf das oben skizzierte Anwendungsszenario ist über Definiendum und Definiens außerdem folgendes zu sagen:

1. Das Definiendum entspricht dem vom Nutzer eingegebenen Suchwort. In Abbildung 2 also der Begriff *Nachhaltigkeit*. **Daher können natürlich nur solche Wörter bzw. Syntagmen als zulässige Definienda im Sinn der vorliegenden Annotationsrichtlinien betrachtet werden, die als Stichwort in einem Wörterbuch[5] denkbar wären**.
2. Das Definiens ist der Bestandteil eines Textsegments, der dem Nutzer **Informationen über das von ihm eingegebene Suchwort (= Definiendum) liefert, wie man sie etwa in einem Wörterbuch finden würde**. Das heißt, dass das Definiens dem Nutzer beim Verorten des Suchbegriffs in seinem Wortschatz helfen können sollte.

Die Abbildung 4 verdeutlicht die zwei grundlegenden Bestandteile eines definitorischen Textsegments. **Ein Textsegment kann nur dann als definitorisch angesehen und entsprechend mit DEF annotiert werden, wenn Definiendum und Definiens vorhanden, eindeutig identifizierbar sind und den in diesen Annotationsrichtlinien angegebenen Kriterien genügen.**

[4] Lexikalisiert: Die Bedeutung lässt sich nicht mehr anhand der Bedeutung der Bestandteile vorhersagen und/oder die Sprecher nehmen den lexikalisierten Ausdruck nicht mehr als kompositional aufgebaut wahr, sondern als angemessenen Ausdruck für ein Konzept.
[5] Wörterbuch (wie Duden, oder Wiktionary) nicht Enzyklopädie (wie Brockhaus oder Wikipedia)!

Zu beachten ist, dass ein Textsegment nicht automatisch definitorisch ist, nur weil es etwas für den Annotator Neues oder Interessantes benennt!

Abbildung 4: Zweiteilung von definitorischen Textsegmenten mit Bestandteilen Definiendum und Definiens

Eine Dienstreise ist gegeben, wenn jemand andernorts, also außerhalb seiner Wohnung oder außerhalb des regelmäßigen Arbeitsortes, aus dienstlichen Gründen vorübergehend tätig sein muss.

Definiendum: Dienstreise

Definiens: Erkennungsmerkmal(e) einer Dienstreise (Wann ist etwas eine Dienstreise?)

2.1 Das Definiendum (Plural: Definienda)

Syntaktische Kriterien: Die sprachlichen Einheiten, die als Definiendum in Frage kommen, sind

- einzelne Wörter, wobei nur **Substantive, Adjektive und Verben** berücksichtigt werden, sowie
- kurze Syntagmen, wenn sie als Einträge in einem Wörterbuch denkbar wären.

Andere Wortarten und größere Wortgruppen sind als Defininedum nicht zulässig. In Tabelle 1 sind einige Beispiele für definitorische Textsegmente mit verschiedenen Typen von Definienda dargestellt.

Tabelle 1: Übersicht mögl. Definienda

Typ des Definiendums	definitorisches Textsegment
Substantiv à ***Gesellschaft***	In der Regel bezeichnet der Begriff "Gesellschaft" die Gesamtheit der Menschen eines Landes bzw. einer Nation und ihre Beziehungen zueinander …
Adjektiv à ***weise***	Erfahren, besonnen, gescheit, verständig, wissend, gedankenversunken - all dies fassen wir unter dem Begriff "weise" zusammen.
Verb à ***putzen***	… putzen, auch reinigen, sauber machen, bedeutet in seiner allgemeinen Verwendung Schmutz entfernen … Anmerkung: bei Verben wird häufig der Umweg über die substantivierte Form in definitorischen Textsegmenten gewählt, z.B. das Putzen anstelle von putzen
Fremdwort, Lehnwort etc. à ***confidential disclosure***	Im Englischen nennt man das "confidential disclosure": Wenn ich Ihnen etwas verrate und Sie bitte, mit niemandem darüber zu sprechen, …
Fachwortschatz à ***Kolon- bzw. Rektumkarzinom***	… die Entartung der Schleimhautzellen, die das Innere des Dickdarms (Kolon) und Enddarms (Rektum) auskleiden, bezeichnet man als Kolon- bzw. Rektumkarzinom …
Abkürzung, Akronym etc. à ***CASTOR***	In Deutschland heißen sie CASTOR (das Akronym steht für "CA sk for S torage and T ransport O f R adioactive material", was so viel bedeutet wie …

Semantische Kriterien: Wie die Beispiele in Tabelle 1 zeigen, ist zu ganz verschiedenen semantischen Konzepten eine Definition sinnvoll. Dazu zählen in jedem Fall alle konkreten und sehr viele abstrakte Substantive. Zu den meisten

Handlungen, Prozessen und Eigenschaften ist natürlich auch eine Definition denkbar. Wörter mit wenig semantischem Gehalt, bei denen evtl. die grammatische Funktion im Vordergrund steht, werden nicht berücksichtigt.

2.2 Das Definiens (Plural: Definientia)

Syntaktische Kriterien: Die sprachlichen Einheiten, die als Definiens in Frage kommen, umfassen meistens größere Syntagmen bzw. Teilsätze. In manchen Fällen können auch einzelne Wörter die Funktion eines Definiens übernehmen. Im Gegensatz zum Definiendum, das eher über syntaktische Kriterien bestimmt werden kann, **sind die relevanten Kriterien beim Definiens vorwiegend semantischer Natur**.

Semantische Kriterien: Inhaltlich kann das Definiens eine Vielzahl von verschiedenen Aspekten aufweisen, denn natürlich können ganz verschiedene Informationen zur Definition herangezogen werden. Tabelle 2 zeigt eine Liste verschiedener Informationseinheiten, die für Definitionen relevant sein können. Diese Übersicht ist in verschiedenen Aspekten angelehnt an Mönke, 1978 und Herbst / Klotz, 2003. Der Großteil der aufgeführten Angaben und alle Beispiele wurden allerdings bei unseren eigenen Arbeiten zu definitorischen Textsegmenten in Korpora zusammengetragen.

Tabelle 2: Übersicht mögl. Informationseinheiten im Definiens

Informationseinheiten	Beispiele
1 Klassifikation bzw. taxonomische[6] Einordnung, extensionale[7] Bestimmung	
1a Nennung von Synonymen, Antonymen oder Oberbegriff	(1a/1b) Tripper ist eine Geschlechtskrankheit, die durch ... (1a/1b) Unter Radioaktivität versteht man die Eigenschaft von bestimmten Stoffen …
1b Fachliche oder hierarchische Einordnung in ein Begriffs-system	
1c Liste verwandter Begriffe (à Wortfeld)	(1c) … Im Laufe der Zeit änderte sich die Bedeutung von burschikos in Richtung "jungenhaft", "knabenhaft", "formlos", "unkompliziert im Verhalten" …
1d Liste von Unterbegriffen oder Beispielen	(1d) … Speicher , der die Daten auch ohne Strom behält, also eine Festplatte, Diskette, CD-ROM, Magnetkarte, aber auch die gute alte Lochkarte …

[6] Taxonomisch (laut: http://services.langenscheidt.de/fremdwb/fremdwb.html): "die → Taxonomie betreffend, auf ihr beruhend, systematisch" und Taxonomie: … "Teilgebiet der Linguistik, das sich mit der Segmentierung und Klassifikation sprachlicher Einheiten befasst" …

[7] Extensional (laut http://services.langenscheidt.de/fremdwb/fremdwb.html): "die → Extension […] betreffend, auf ihr beruhend" und Extension: "Umfang eines Begriffs, Gesamtheit der Unterbegriffe"

2 genetisch[8]-kausale Bestimmung, prozedurale Angabe	
2a Entstehung, Ursprung, Herkunft	(2a) … auch durch bestimmte Schmerzmittel selbst können unter Umständen Schmerzen ausgelöst werden (sog. medikamentinduzierter Kopfschmerz) … (2a) …Sedimentgesteine […] bilden sich aus den Ablagerungen von Verwitterungsprodukten ursprünglicher Gesteine oder von Überbleibseln von …
2b Herstellung	(2a/2b) … Bulgur entsteht, wenn Weizenkörner gekocht, geschält und getrocknet, schließlich zu Granulaten (grob, mittel, fein) zerkleinert werden … (2b/2c) … bei normaler Panade wird erst Mehl, dann gequirlte Eier und dann Brösel auf das Fleisch getan …
2c Angabe zu typischen Abläufen, Prozessen, Handlungen	(2c) … beim Baggern wird der Ball mit den parallel gehaltenen, durchgestreckten Unterarmen von unten gespielt…
3 Beschaffenheit, Kenndaten, Erkennungsmerkmale	
3a Bestandteile	(3a) … Haut besteht aus mehreren Schichten, die oberste wird Epidermis genannt, das darunter liegende Bindegewebe mit Muskeln, Nerven und Gefäßen heißt Lederhaut oder Korium, und das Unterhautfettgewebe wird Subkutis genannt …
3b Extension (= Umfang eines Begriffs)	(3b) … als Edelgase bezeichnet man Helium, Neon, Argon, Krypton, Xenon, Radon und Ununocticum …
3c (materielle) Beschaffenheit, Aussehen bzw. sonstige besonders relevante Erkennungsmerkmale …	(3c) … einen Weisheitszahn erkennt man daran, dass er bei vollständig bezahntem Gebiss, jeweils der 8. Zahn ist, von der Mitte ab gezählt … (3c) ... ausrasten tut man, wenn man einen völlig unkontrollierten Wutanfall hat …
3d (mathematische) Formeln, technische, wissenschaftliche Kennwerte	(3d) … der dann zu hörende charakteristische Brummton mit 50 Hz (der sog. "Netzbrumm") entspricht genau dem …

[8] Genetisch bedeutet (laut http://services.langenscheidt.de/fremdwb/fremdwb.html): "die → Genese betreffend, auf ihr beruhend" und Genese: "Entwicklung, Entstehung"

4 Funktion	
4a Funktion, funktioneller Zusammenhang	(4a) … die Sporen (Samen) der Pilze dienen zur Fortpflanzung und …
4b Zweck, Anwendungsmöglichkeit, Einsatzgebiet	(4a/4b) … Ein Router wird meistens dann eingesetzt, wenn man in einem lokalen Heimnetzwerk mit mehreren PCs einen gemeinsamen Internetzugang benutzen will …
4c Arbeitsweise	(4c) … Bluetooth arbeitet mit kurzwelligen Funksignalen zum Übertragen von Daten auf Netzwerkgeräten …
5 Angaben zu Verwendung und linguistische Aspekte	
5a Wortgebrauch, Situierung in Verwendungskontext auch Zuordnung zu Fach / Disziplin	(5a) … in der Semantik spricht man davon, daß bestimmte Unterscheidungen lexikalisiert, d.h. durch ein eigenes Lexem (=Wort) repräsentiert sind …
5b grammatische Angaben	(5b) … Das Verb "bewegen" ist transitiv, das heißt, es ist aktiv wie passiv zu verwenden: ich bewege jemanden, ich werde von jemandem bewegt …
5c Etymologie, Übersetzung (Äquivalente)	(5c) … die Italiener übernahmen aber nicht die Begriffe Football oder Soccer als Bezeichnung für ihren Fußball, sondern verwendeten das Wort Calcio… (5c) … der Begriff Qualität kommt seinem Ursprung nach aus dem lateinischen "qualitas", das mit "Beschaffenheit" ...
5d Angaben zu Stilmarkierung	(5d) … Das Wort Neger empfinde ich als Beleidigung …* * Textsegmente, die eine Meinung oder Empfindung beschreiben, dürfen nur dann als definitorisch annotiert werden, wenn sie gleichzeitig Informationen über Stilmarkierung oder Konnotation eines Wortes benennen. In diesem Fall beinhaltet der Satz die Information, dass das Wort Neger als Beleidigung verwendet/empfunden werden kann und daher nicht neutral ist.
6 Quelle	
6a Angabe der Quelle, Verweis auf Autor, Ort der Veröffentlichung	(6a) … unter einem Satz verstehe ich mit Zifonun, G. et al. (1997) und www.ids.mannheim.de/grammis (2001) eine syntaktische Wortgruppe mit einem finiten Verb …

Wie Tabelle 2 verdeutlicht, können die Typen von Informationseinheiten auch beliebig kombiniert im Definiens auftreten. Natürlich müssen nicht alle Angaben der Tabelle gleichzeitig vorhanden sein, damit es sich um ein definitorisches

Textsegment handelt. Normalerweise wird eine kleine Auswahl der möglichen Informationseinheiten oder auch nur ein einzelner Aspekt zur Definition herangezogen. **Wichtig ist allerdings, dass die Informationseinheit(en), die im Definiens genannt werden, relevante bzw. zentrale Aspekte der Bedeutung des Definiendums darstellen**.

3 Typische Muster von Definitionen

Bei der Suche und Analyse von definitorischen Textsegmenten in Korpora konnten wir beobachten, dass es einerseits bestimmte Muster gibt, nach denen die Informationseinheiten im Definiens angeordnet sind und dass andererseits bestimmte Formulierungen immer wieder auftreten. In diesem Abschnitt stellen wir daher besonders **typische Muster von Informationsstrukturen und Formulierungen definitorischer Textsegmente** zusammen. Natürlich darf die Darstellung in diesem Abschnitt nicht als umfassend angesehen werden – es können in den zu annotierenden Daten durchaus immer wieder ganz andere Formulierungen oder Informationsstrukturen auftreten, die trotzdem als Definition einzustufen und entsprechend zu annotieren sind.

3.1 Muster typischer Informationsstrukturen

Ein besonders bekanntes Muster stellt die **Einordnung über "genus proximum et differentia specifica"** dar. Dabei wird die Definition mit Hilfe von zwei sich ergänzenden Informationseinheiten vorgenommen, nämlich:

- Der sog. *nächst höheren Art* (à genus), das heißt einem direkten oder indirekten Oberbegriff (Hyperonym) und
- Den sog. *spezifischen Unterschieden* (à differentiae), das heißt den Merkmalen, die das Definiendum von allen seinen Geschwistern (Kohyponyme) abgrenzen. (Beispiel: Definiendum = Stuhl, genus = Möbelstück, Geschwister/Kohyponyme = Tisch, Schrank etc.)

Abbildung 5: Definition mit genus et differentia

Als Gonorrhoe (Tripper) bezeichnet man eine Geschlechtskrankheit, die durch bestimmte Bakterien, sog. Gonokokken, verursacht wird.

genus à Oberbegriff, hier: Geschlechtskrankheit

differentiae à Merkmale, die Gonorrhoe von anderen Geschlechtskrankheiten abgrenzen

Abbildung 6: Definition mit genus et differentia

… ein Histogramm ist ein Flächendiagramm, in dem die Flächen den (absoluten oder relativen) Häufigkeiten proportional sind …

genus à Oberbegriff, hier: Flächendiagramm

differentiae à Merkmale, die das Histrogramm von anderen Flächendiagrammen abgrenzen

Ein anderes relevantes und häufig verwendetes Muster stellt die **operationale Definition** dar. Hierbei wird im Definiens beschrieben, woran das Definiendum zu erkennen ist. Meist werden dazu **konditionale** Formulierungsmuster eingesetzt – im Sinn von: *falls X, Y und Z vorliegt, handelt es sich um Definiendum*.

Abbildung 7: Zwei Beispiele für operationale Definition

Ein Telearbeitsplatz liegt vor, falls ein Arbeitnehmer an seinem Wohnort dezentral für ein räumlich entferntes Unternehmen über elektronische Medien Tätigkeiten in Erfüllung seines Arbeitsvertrages ausübt …

Definiens: Woran erkennt man, ob ein Telearbeitsplatz vorliegt?
Idee: Überprüfe im Definiens genannte Erkennungsmerkmale, um festzustellen, ob etwas ein Telearbeitsplatz ist.

Zahlungsunfähigkeit liegt vor, wenn der Schuldner nicht in der Lage ist, seine derzeit fälligen Zahlungsverpflichtungen zu erfüllen …

Definiens: Woran erkennt man, ob Zahlungsunfähigkeit vorliegt?

Weitere häufig eingesetzte Muster, die leider besonders gerne bei der Annotation übersehen werden, stellen die verschiedenen **Kurzformen der Definition** dar. Im Gegensatz zu den beiden oben gezeigten Mustern (also genus et differentia sowie der operationalen Definition) zielt der Produzent eines Textes in diesen Fällen häufig darauf ab, dem Rezipienten eine (mehr oder weniger umfassende) **Verstehenshilfe** (für das Textverständnis des aktuell betrachteten Textes) anzubieten. Typischerweise wird hierbei ein Synonym oder Hyperonym zum Definiendum genannt. Alternativ werden häufig auch einige wenige typische Bestandteile oder Beispiele für das Definiendum angeben. Beispiele für diese Kurzformen sind in Abbildung 8 dargestellt.

Abbildung 8: Vier Beispiele für Kurzformen der Definition

… der Spieler verpflichtet sich des Weiteren, die Privatsphäre Dritter zu beachten, keine unaufgeforderten Massensendungen (so genannte "Spam") sowie keine ungebetene Werbung zu verbreiten …

Definiendum: Spam

Definiens: unaufgeforderte Massensendungen

… ... ein Webtagebuch, auch Weblog genannt, gehört für Firmen, Vereine, Verbände und Privatpersonen mittlerweile zum Pflichtprogramm .....

Definiendum: Weblog

Definiens: Webtagebuch

Kurzformen der Definition hier mit 2 Lesarten:

… bringen Sie bitte Ausweisdokumente, wie z.B. einen Reisepass, mit und …

Definiendum: Ausweisdokumente

Definiens: Reisepass (betrachtet als ein Beispiel für Ausweisdokumente)

… bringen Sie bitte Ausweisdokumente, wie z.B. einen Reisepass, mit und …

Definiendum: Reisepass

Definiens: Ausweisdokumente (betrachtet als Oberbegriff zu Reisepass)

3.2 Muster typischer Formulierungen

Häufig, aber natürlich nicht immer, kann man bestimmte Formulierungsmuster bei definitorischen Textsegmenten beobachten. In der Regel handelt es sich um auf

bestimmte Weise lexikalisierte syntaktische Strukturen. Meistens bilden bestimmte Verbkomplexe den Kern dieser Formulierungsmuster. Dabei kann man grundsätzlich unterscheiden zwischen Formulierungen, die eine metasprachliche[9] Aussage anzeigen (z.B. `X nennt man Y`), und solchen, die eher auf das Vorliegen bestimmter Informationseinheiten im Definiens hindeuten (z.B. Funktionsangabe: `X dient Y`). Häufig werden aber auch bestimmte Satzzeichen und typographische Markierungen für die Strukturierung definitorischer Textsegmente verwendet. Beispiele für diese Art der Strukturierung und die oben angesprochenen Formulierungsmuster sind in Tabelle 3 und Tabelle 4 dargestellt.

Tabelle 3: Auswahl wichtiger Formulierungsmuster (mit eher metasprachlichem Charakter)

Forumlierungsmuster	Beispiel
… DEFINIENDUM ist / stellt dar / gilt als / kann man sich vorstellen als … DEFINIENS …	… Rückbau ist im Fachjargon die etwas vornehmere Formulierungsvariante für Abriss … … eine Zeitschrift ist eine Sammelpublikation, die regelmässig (täglich, wöchentlich, monatlich, zweimonatlich, halbjährlich, jährlich) erscheint … … ein Histogramm ist ein Flächendiagramm, in dem die Flächen den (absoluten oder relativen) Häufigkeiten proportional sind … … Forschungsresultate belegen, dass Vitamin E ein wirksames Mittel im Kampf gegen die Arteriosklerose darstellt … … ein Urlaubssemester gilt als unterrichtsfreie Zeit … … anschaulich kann man sich die Delta-Distribution als eine beliebig hohe und beliebig schmale Funktion vorstellen, deren Fläche den Grenzwert 1 besitzt und …
… DEFINIENDUM liegt vor, wenn / falls … DEFINIENS … (falls / wenn) … DEFINIENS (vorliegt,) spricht man von DEFINIENDUM …	… Arbeitslosigkeit liegt vor, wenn ein Teil der arbeitsfähigen und arbeitswilligen … … Zahlungsunfähigkeit liegt vor, wenn der Schuldner nicht in der Lage ist, seine derzeit fälligen Zahlungsverpflichtungen zu erfüllen … … wenn ein Arbeitnehmer an seinem Wohnort dezentral für ein räumlich entferntes Unternehmen über elektronische Medien Tätigkeiten in Erfüllung seines Arbeitsvertrages ausübt, spricht man von einem Telearbeitsplatz ... … die Hefe wird zuerst, nach getaner Arbeit, abgezogen und man spricht dann vom "Jungbier" …
DEFINIENDUM nennt man DEFINIENS … … als/mit DEFINIENDUM bezeichnet man DEFINIENS…	… blasig nennt man ein Gestein, wenn es von regelmässig oder unregelmässig gestalteten, gerundeten Hohlräumen durchzogen ist, ... … als "cache" (gesprochen käsch) bezeichnet man einen Speicherbereich (Zwischenspeicher), der …

[9] Metasprachlich bedeutet (laut: http://services.langenscheidt.de/fremdwb/fremdwb.html): die Metasprache betreffend und Metasprache bedeutet: Sprache über Sprache.

DEFINIENDUM benannt nach … DEFINIENS	… mit Mimik und Gestik bezeichnet man in der Regel das nonverbale (nicht-sprachliche) Verhalten … … März (Lenzing, Lenzmond), benannt nach Mars, dem Gott des Krieges und der Vegetation; im alt-römischen Kalender begann das Jahr …
… DEFINIENDUM heißt, dass … DEFINIENS … DEFINIENDUM bedeutet wörtlich / übersetzt / soviel wie … DEFINIENS … … DEFINIENDUM bedeutet in diesem Zusammenhang / hier / hierbei …, dass … DEFINIENS	… wenn ich sage: etwas wegnehmen, dann heisst das, dass sie etwas beanspruchen, das ihnen zusteht … … Desertifikation bedeutet wörtlich 'Wüstenbildung' ... … Mehrbenutzersystem bedeutet in diesem Zusammenhang, dass mehrere Anwender einen Computer gleichzeitig benutzen können … … operativ bedeutet hierbei, dass … … Ganzheitlich bedeutet hier, daß vorwiegend bildhaft, gefühlsorientiert und vernetzend erlebt wird, nicht aber begrifflich, logisch und systematisierend … ACHTUNG: Das *hier* bzw. *hierbei* kann auf eingeschränkte Gültigkeit der Aussage hinweisen – ist die Gültigkeit auf den Text beschränkt, sollte es mit O und nicht mit BETS annotiert werden.
… DEFINIENDUM heißt auf … DEFINIENS … DEFINIENDUM kann man übersetzen mit … / kann man so übersetzen: … / übersetzt man mit … DEFINIENS	… "Màiá" heisst auf deutsch "Kranz" … … in Deutschland heißen sie CASTOR … … Shinkendo kann man so übersetzen: Shin = ernsthaft, ken = Schwert, do = Weg, somit also *Weg des ernsthaften Schwertes* …
… DEFINIENDUM lässt sich definieren als / definiert man als … DEFINIENS	… Wortbildung lässt sich definieren als Prozess, der … … Schulverweigerung definiert man als eine vom Kind ausgehende Weigerung, die Schule zu besuchen oder sein Unvermögen, den Schulalltag durchzustehen …
… bei / beim DEFINIENDUM handelt es sich um … DEFINIENS … bei / beim DEFINIENDUM geht es um / wird … DEFINIENS	… bei einem Axiom handelt es sich um eine Aussage, die keines Beweises durch andere Aussagen bedarf und als unbeweisbare Aussage die Basis aller Beweise in ... … bei der traditionellen Semantik geht es um die Bedeutung des Einzelwortes … … beim Baggern wird der Ball mit den parallel gehaltenen, durchgestreckten Unterarmen von unten gespielt … … bei der Photosynthese wird Sonnenlicht von Pflanzen oder Bakterien absorbiert und über eine komplexe Reaktionskette in chemische Energie umgewandelt …

… bei / beim DEFINIENDUM geschieht folgendes: … DEFINIENS	… beim Riechen geschieht folgendes: Die Geruchsmoleküle werden in die Nasenschleimhaut weitergeleitet und dort aufgelöst, wo sie an Rezeptoren andocken und das Riechgehirn weitergeleitet werden…
… DEFINIENDUM tut man, wenn … DEFINIENS	... abnehmen tut man, wenn man mehr Kalorien verbraucht als man zu sich nimmt, ganz einfach …
… unter / unter dem Begriff DEFINIENDUM versteht man … DEFINIENS	… unter Aufklärung versteht man einen sowohl individuellen wie gesellschaftlichen, geistigen Emanzipationsprozess …
… der Begriff DEFINIENDUM wird verstanden als / benutzt für / verwendet für … DEFINIENS	… wird der Begriff "Aborigines" heute häufig als Sammelbegriff für alle schwarz-australischen Völker benutzt …
… unter dem Begriff / den Begriff DEFINIENDUM versteht man / gebraucht man für… DEFINIENS	… den Begriff gebraucht man für eine zeitweilig heftigere vulkanische Tätigkeit, wie z.B. einen Vulkanausbruch …
… der Begriff DEFINIENDUM umfasst … DEFINIENS	… umfasst der Begriff "besondere Beziehungen" auch verwandtschaftliche Beziehungen und ganz allgemein jede Interessengemeinschaft , die neben dem Rechtsverhältnis …
… DEFINIENDUM kann man be-/umschreiben als … DEFINIENS	… Mobbing kann man umschreiben als systematisch betriebene Niederträchtigkeit von Mitarbeitern gegen andere Mitarbeiter, Untergebene oder Vorgesetzte …

Neben diesen eher metasprachlichen Mustern kann die Formulierung eines definitorischen Textsegments auch durch die Informationseinheit bestimmt sein, die im Definiens genannt wird. Beispiele hierfür sind in Tabelle 4 dargestellt.

Tabelle 4: Auswahl wichtiger (durch im Definiens genannte Informationseinheiten bestimmte) Formulierungsmuster

Typ der Informationseinheit & Forumlierungsmuster	Beispiel
Typ der Informationseinheit = Bestandteile	
… DEFINIENDUM besteht aus / setzt sich zusammen aus … DEFINIENS …	… ein soziales System besteht definitionsgemäß aus einer Mehrheit von Akteuren, die in einer bestimmten physischen Umwelt miteinander interagieren und durch ...
	… ein Geräusch setzt sich in der Regel aus Tönen verschiedener Frequenzen zusammen …
… die Bestandteile von DEFINIENDUM sind … DEFINIENS (	… die Bestandteile eines Prozessors sind das Rechenwerk, auch ALU genannt, das Steuerwerk, Speichermanager und das Register …

Typ der Informationseinheit = Funktion / Zweck … DEFINIENDUM dient / wird eingesetzt / verwendet … DEFINIENS Hauptzweck / Funktion von DEFINIENDUM ist … DEFINIENS	… eine Notstromversorgung dient der Überbrückung und der Aufrechterhaltung zentraler Arbeitsbereiche, beispielsweise in Krankenhäusern, Leitwarten und ... … ein Router wird meistens dann eingesetzt, wenn man in einem lokalen Heimnetzwerk mit mehreren PCs einen gemeinsamen Internetzugang benutzen will. … der Hauptzweck eines Repetitoriums ist es, den gesamten Stoff zu wiederholen, zu festigen und durch ständige Falllösung die Gutachtentechnik ... … die Hauptfunktion eines Messbechers ist das exakte Ablesen des Inhalts …
Typ der Informationseinheit = zentrale Merkmale …die Merkmale von DEFINIENDUM sind … DEFINIENS … … DEFINIENDUM ist gekennzeichnet durch / charakterisiert als (durch) … DEFINIENS	… die wesentlichen Merkmale eines Passivhauses sind: sehr kompakte Bauform; Minimierung aller Wärmebrücken; über das Niedrigenergiehaus hinausgehender ... … ein Déjà-vu-Erlebnis ist gekennzeichnet durch: Gefühl, etwas schon einmal gesehen zu haben Ablaufen des eigenen "Lebensfilms" in ganz kurzer Zeit ... … ILM (Information Lifecycle Management) kann man charakterisieren als die Zusammenführung verschiedener Storage, Archiv, Content und Document-Management-Technologien …
Typ der Informationseinheit = Entstehung / Ursprung … DEFINIENDUM entsteht durch / bei … DEFINIENS …	… Husten entsteht durch Reizung der Bronchialschleimhaut, wenn die sich entzündet, dann nennt man das Bronchitis … … ein Kuboktaeder entsteht, wenn man die Ecken eines Würfels abschneidet …
Typ der Informationseinheit = Bespiel(e) …ein Beispiel für / als Beispiel für DEFINIENDUM ist / kann sein / kann gelten … DEFINIENS	… ein Beispiel für ein Lerndefizit ist, wenn ein Schüler besondere Schwierigkeiten hat, Lesen und Schreiben zu erlernen, wobei er in anderen Bereichen …

VORSICHT: Besonders im Bereich des politischen Diskurses werden manchmal Formulierungen eingesetzt, die strukturell, lexikalisch und besonders im Hinblick auf bestimmte Verben einer Definition sehr ähnlich sehen. Solche Textsegmente sind bei der Annotation mit äußerster Vorsicht zu bearbeiten. Beispiel: … die Kandidatenwahl des Bundestagswahlkreises 195 Anfang Dezember bezeichnete er als "Lehrstück innerparteilicher Demokratie" … (http://www.ju-weimarerland.de/bilanz.htm Stand: 2008-02-04)

Tabelle 5 zeigt eine Auswahl Satzzeichen-basierter Strukturierungsmuster, wie sie in definitorischen Textsegmenten auftreten können. Manchmal werden diese allerdings mit Formulierungen wie *z.B.* oder *wie* kombiniert.

Tabelle 5: Auswahl von Satzzeichen zur Strukturierung von Definitionen

Zeichen	Beispiel
()	… sie fahren von Ihrer nächstgelegenen Haltestelle mit Zügen des Nahverkehrs (RegionalBahn, StadtExpress, RegionalExpress, S-Bahn) oder mit U-Bahnen, Bussen oder Straßenbahnen direkt zu Ihrem Flughafen …
:	Bewusstsein: die Gesamtsumme der kooperierenden Aktivitäten der Gehirnstrukturen, die körperliche Prozesse vermitteln …
–	… Feuerwerkskörper der Klasse I (P I) – etwa Knallerbsen, Knallbonbons, Zünd- oder Knallplättchen – dürfen dagegen ganzjährig auch von Personen unter 18 Jahren benutzt werden …
, … ,	… für bronchienerweiternde Medikamente, die sogenannten Beta-Sympathomimetika, reicht diese Art der Inhalation meistens aus …

4 Begrenzung, Kontext und schwer entscheidbare Fälle

In den folgenden Abschnitten werden verschiedene Grenzfälle, schwerentscheidbare Fälle und möglicherweise in der praktischen Annotationsarbeit problematische Bereiche diskutiert.

4.1 Abgrenzung und Kontext

Meisten ist es nicht ganz einfach, ein definitorisches Textsegment ohne Kontext zu verstehen, denn selten kann man es aus seinem Kontext herausreißen, ohne das dadurch Information verloren geht. Daher kann eine, dem Gefühl nach ohne Kontext schwache Definition mit Kontext stärker wirken. Beispiele hierfür sind in Abbildung 9 zum Definiendum *Unwetter* dargestellt. Wie ist nun aber mit dem Einfluss des Kontextes auf die Annotationsentscheidung umzugehen? Zunächst einmal gilt, dass **jeder Satz separat** im Hinblick auf die Frage analysiert wird, ob eine Definition vorliegt oder nicht. Dann stellt sich die Frage, ob das Textsegment auch ohne Kontext zentrale Informationen über ein Definiendum enthält, dass heißt, **ob die Information zum Definiendum lokal ungebunden ist. Lokal ungebunden bedeutet hier, die Information über das Definiendum muss unabhängig von dem Text in dem es vorkommt gelten können, also eine vom Text unabhängige Information darstellen.** Ist das der Fall, muss das Segment als Definition annotiert werden.

Abbildung 9: Bedeutung des Kontextes für das Verständnis von Definitionen
(http://www.unwetter.de/pages/unwetter.php Stand: 2008-02-04)

Unwetter-Definition

Was verstehen wir unter einem Unwetter?

Meteorologisch lässt sich der Begriff Unwetter anhand objektiver physikalischer Kriterien nicht definieren. Ein Unwetter lässt sich daher nur anhand der Auswirkungen bestimmten Wetters definieren, wie dies z.B. im Zeit-Lexikon (2004, Band 16: S. 267) geschieht.

Unwetter sind Wetterereignisse, die „stärkste Auswirkungen haben" und einen "Notstand" hervorrufen können.

Ein Notstand ist ein Zustand der unmittelbaren Gefahr. Im Notstandgebiet sind die „Lebensbedingungen" und das „Wirtschaftsgefüge" erheblich beeinträchtigt.

Der Notstandsbegriff lässt sich damit auf die öffentliche Ordnung und das öffentliche Leben anwenden, welches durch Wettereinflüsse erheblich gestört werden kann.

Ein Unwetter ist also ein Ereignis welches die öffentliche Ordnung und das öffentliche Leben gefährdet.

Schwere Regenfälle, extreme Dürren, Stürme, Glättesituationen werden heute bei den meisten Wetterdiensten weltweit als Unwetter gewertet. Andere nicht so sichtbaren Wetterauswirkungen, mit Gefahrenpotential für Leib und Leben, werden nur vereinzelt bisher mit in die Unwetterwarnungen einbezogen. Dies sind Gefahren durch thermische Belastungen, starke Sonneneinstrahlung und eine schlechte Luftqualität.

Vielfach wird daher auf die Formulierung „Wettergefahr" ausgewichen, der aber so allgemein ist, dass fast jeder Wetterzustand eine Gefahr für bestimmte Bevölkerungsgruppen darstellen kann.

Letztendlich hilft man sich in der Meteorologie mit Schwellenwerten aus. Man legt also mehr oder weniger eindeutig fest, z.B. bei welcher Windgeschwindigkeit gewarnt wird, oder nicht.

Übrigens haben Unwetter in der Natur eine wichtige Funktion. Sie sorgen z.B. für eine „natürliche" Veränderung des Flussbettes, nach Erosion für eine neue Vegetation, für das „umfallen" morscher Bäume usw...

Wie Abbildung 9 zeigt, kommen definitorische Textsegmente häufig als Ketten oder Cluster in Korpora vor. Das heißt, man findet in bestimmten Bereichen gleich mehrere Definitionen nacheinander. Dabei muss es sich aber nicht immer um eine Kette handeln, wie das in Tabelle 6 dargestellt ist.

Tabelle 6: Beispiel für eine Kette von Definitionen

Als **pflegebedürftig** gilt, wer mindestens 90 Minuten täglich der Hilfe bedarf. Dabei muss die Hilfebedürftigkeit im Bereich der Grundpflege überwiegen. **Grundpflege** bedeutet …

Manchmal clustern (~ *ballen sich*) Definitionen in Texten z.B. um ein bestimmtes Thema – das kann (muss aber nicht) z.B. das Textthema (oder das Thema eines Textteils) sein. In diesen Fällen ist mit besonderer Vorsicht zu annotieren, weil das Textthema (Thema eines Textteils) leicht von Definitionen ablenken kann. **Außerdem dürfen Textsegmente nicht alleine deswegen als definitorisch annotiert werden, weil sie eine relevante Information über das Textthema enthalten.** Solche Textsegmente dürfen nur dann als Definition annotiert werden, wenn sie das Definiendum (evtl. = Textthema / Thema des Textteils) explizit nennen, das Definiens eindeutig zu identifizieren ist und beide außerdem den Bedingungen entsprechen, die in den entsprechenden Abschnitt dieser Richtlinien zu Definiens / Definiendum genannt sind.

4.2 Anaphorische Strukturen

Manchmal findet man in definitorischen Textsegmenten sog. anaphorische Strukturen. Definiens oder Definiendum können dabei z.B. durch ein Pronomen oder andere Proformen ersetzt werden. Beispiele dafür sind in Tabelle 7 dargestellt.

Tabelle 7: Anaphorische Beispiele

Definiens à ***diese*** (nämlich: ***diese Behälter zur …***)	… für den Transport von den Kernkraftwerken in entsprechende Zwischenlager (für die spätere direkte Endlagerung) oder zur Wiederaufbereitung benötigt man ganz

	besondere Behälter. In Deutschland heißen diese CASTOR …
Definiens à ***diese*** (nämlich: ***die computerverwalteten Zuordnungen zwischen Ankern***)	… Ein konstitutives Element von Hypertexten sind computerverwaltete Zuordnungen zwischen Ankern. Diese bezeichne ich als *Links* …
Definiendum à ***diese Vertriebsarten*** (nämlich: ***der sog. Fernabsatz***)	… seit Anfang der 80er hat der sog. Fernabsatz stark zugenommen. Diese Vertriebsarten sind insbesondere dadurch gekennzeichnet, daß Anbieter und Verbraucher sich nicht physisch begegnen und der Verbraucher die Ware oder Dienstleistung in der Regel nicht vor Vertragsschluss in Augenschein nehmen kann ...

Textsegmente dieser Art können zwar als Definition betrachtet werden und stellen sicherlich interessante Fälle dar, **sie werden aber der Einfachheit halber im Sinn dieser Annotationsrichtlinien nicht als Definition annotiert.**

4.3 Eigennamen

Eigennamen gelten grundsätzlich nicht als zulässige Definienda. Textsegmente mit Informationen über Personen, Organisationen etc. werden daher im Sinn der vorliegenden Annotationsrichtlinien **auf keinen Fall als Definition annotiert**. Zu beachten ist an dieser Stelle die Unterscheidung von der **Bezeichnung eines Individuums**, zum Beispiel Frankreich (= Ländername und somit Eigenname), die **nicht annotiert** wird und der **Bezeichnung einer Klasse**, wie „Beagle“, die **annotiert** werden müssen.

4.4 Überlappungen und verschränkte Definitionen

Wie bereits in Abschnitt 4.1 skizziert, tendieren definitorische Textsegmente zur Kettenbildung bzw. zum Clustern. In diesen Fällen findet man in kurzer Abfolge bzw. in direkt aufeinander folgenden Sätzen Definitionen. Daneben kann es sogar passieren, dass ein Textsegment gleich mehrere Definitionen zu verschiedenen Definienda enthält. Einige Beispiele hierfür sind in Tabelle 8 dargestellt.

Tabelle 8: Beispiel für verschränkte und überlappende Definitionen

Beispiel definitorisches Textsegment	**Anmerkung**
Weltwirtschaftsgipfel, (G8-Gruppe), jährliche Konferenz der sieben größten westlichen Industrienationen (USA, Kanada, Japan, Großbritannien, Italien, Frankreich und Dtld.) und Russlands.	Definiendum 1: **Weltwirtschaftsgipfel** Definiendum 2: **G8-Gruppe** Die beiden Definitionen sind ineinander verschränkt.
… chronisch-entzündliche Darmerkrankungen mit entzündlichen Veränderungen der Schleimhaut (z.B. Colitis ulcerosa oder Morbus Crohn) können …*	Definiendum 1: **Colitis ulcerosa** Definiendum 2: **Morbus Crohn** Die beiden Definienda sind Kohyponyme zum gemeinsamen Hyperonym chronisch-entzündliche Darmerkrankungen. Das Definiens bezieht sich also auf beide Definienda gleichermaßen.

* Natürlich stellt gerade dieses Beispiel aus der Sicht des Autors keine Definition dar. Für einen Leser aber, der mit dem Begriff *Colitis ulcerosa* nichts anfangen kann, beinhaltet das Segment eine – wenn auch knappe – Information zur Einordnung in seinen Wortschatz.

Wie die Beispiele in Tabelle 8 zeigen, lässt sich in den meisten Fällen eine bestimmte semantische Beziehung zwischen den Definienda feststellen, die in ein

und demselben Textsegment eingeführt werden. In der Regel handelt es sich um Kohyponyme oder (Teil-)Synonyme.

4.5 Vollständigkeit & Korrektheit

Philosophisch betrachtet muss eine Definition genau die (und nur die) Unterscheidungsmerkmale beinhalten, die es ermöglichen, das Definiendum zweifelsfrei von seinen Geschwistern oder Kohyponymen abzugrenzen – theoretisch. Praktisch setzt das allerdings eine eindeutige Klassifizierung aller dem aktuellen Definiendum benachbarten Konzepte voraus. Nur auf dieser Grundlage kann nämlich überhaupt entschieden werden, was die Unterscheidungsmerkmale und der direkte Oberbegriff in einem konkreten Fall sein könnten. Innerhalb einer bestimmten wissenschaftlichen Disziplin kann die Entscheidung über eine solche Klassifikation genau so wie die Beurteilung ihrer Qualität nur ein Experte vornehmen – und meistens kann dieser Experte das sogar nur für einen Ausschnitt, nämlich seinen speziellen wissenschaftlichen Teilbereich.

Tabelle 9 zeigt ein (allgemeinsprachliches) Beispiel einer aus streng philosophischer Sicht "*falschen"* bzw. *"unzulänglichen"* Definition. Jedem – Laie oder Experte – fällt an diesem Beispiel sofort die fragwürdige Auswahl der Unterscheidungsmerkmale auf, die in Kombination mit dem sehr allgemein gewählten Oberbegriff nämlich z.B. auch auf den *Löwen* zutreffen würden. Hätte der Autor hingegen anstelle von *Tier* den (direkten) Oberbegriff *Unpaarhufer* gewählt, müsste man das Beispiel völlig anders bewerten. In diesem Fall würde das Merkmal *lange Mähne* das Pferd von seinen Geschwistern (*Esel*, *Zebra, Nashorn, Tapir*) unterscheiden, da diese nur kurze oder keine Mähnen haben.

Tabelle 9: Beispiel für eine offensichtlich problematische Auswahl von Oberbegriff und Unterscheidungsmerkmalen in definitorischen Textsegmenten

… ein Pferd ist ein Tier mit vier Beinen, einem Schwanz und einer langen Mähne…

Ein weniger offensichtlicher Fehler ist im Beispiel in Tabelle 10 enthalten. Dem Laien würde hier wahrscheinlich nicht auffallen, dass Titan chemisch betrachtet kein Edelmetall ist und demzufolge die (zwar für Edelmetalle richtig beschriebenen) Merkmale der Reaktion und des Vorkommens in der Natur auch für Titan nicht zutreffen können.

Tabelle 10: Beispiel für eine versteckt problematische Zuordnung zwischen Oberbegriff und Definiendum in definitorischen Textsegmenten

… Die Edelmetalle, wie z.B. Kupfer, Titan und Silber, zeichnen sich dadurch aus, nicht unter Wasserstoffbildung mit Wasser oder wässrigen Säurelösungen zu reagieren, aus diesem Grund kommen sie in der Natur oft gediegen vor …

In der Praxis ist es daher häufig schwer möglich, den Übergang zwischen einer "*unzureichenden", "unvollständigen"* oder gar "*falschen"* Definition genau zu bestimmen. **Bei der Annotation definitorischer Textsegmente sollte daher die Bewertung nach *richtig* und *falsch* keine große Rolle spielen und die Annotationsentscheidung nicht zu stark beeinflussen!** Trotz der offensichtlich (Tabelle 9) und versteckt (Tabelle 10) problematischen Auswahl von Oberbegriff und

Merkmalen müssten beide dargestellten Beispiele im Sinn der vorliegenden Annotationsrichtlinien als definitorische Textsegmente bewertet werden.

Darüber hinaus sollte man sich bei der Annotation von Texten die **Frage nach der Vollständigkeit einer Definition NICHT stellen**. Definitionen in Texten unterscheiden sich von solchen in Wörterbüchern (bzw. Enzyklopädien) u. a. dadurch, dass sie **nicht unbedingt alle möglichen Bedeutungen, Lesarten bzw. Bedeutungsaspekte** eines Definiendums erfassen müssen. Wichtig ist aber, dass es sich bei definitorischen Textsegmenten immer um Sätze handelt, die über ein Definiendum relevante bzw. zentrale Informationen (im Bezug auf die Bedeutung des Definiendums) benennen.

4.6 Unauffällige Formen und Kurzformen der Definition

Bei der bisherigen Annotationsarbeit haben wir immer wieder festgestellt, dass man bestimmte Typen von Definitionen leicht übersieht. Das gilt besonders, wenn das Definiendum am Ende des Textsegments oder hinter dem Definiens steht, auch zwischen den Bestandteilen des Definiens eingeschobene Definienda sind problematisch. Daneben werden die meisten Kurzformen der Definition leicht übersehen. Außerdem werden solche Textsegmente häufig falsch annotiert, bei denen nur ein sehr kleiner Teil des Segments die eigentliche Definition ausmacht – natürlich besonders, wenn dieser Teil auch noch am Segmentende steht. Für einige dieser unauffälligen Typen sind in Tabelle 11 Beispiele dargestellt.

Tabelle 11: Beispiele für mögliche Formulierungen unauffälliger Definitionen

… Radon und Ununocticum bezeichnet man als Edelgase genauso wie Helium, Neon, Argon, Krypton und Xenon …
… die interessanteste Wirkstoffgruppe des grünen Tees sind die sekundären Pflanzenstoffe, u.a. die Catechine, und in dieser Gruppe besonders das schon genannte "Epigallo-Cathechin-3-O-gallat ", abgekürzt EGCG …
… einige der während des Eingriffs eingesetzten Maßnahmen zur Schmerzunterdrückung (z. B. Epiduralanästhesie) können für die Behandlung postoperativer Schmerzen weiter verwendet werden …
… Die Italiener übernahmen aber nicht die Begriffe Football oder Soccer als Bezeichnung für ihren Fußball, sondern verwendeten das Wort Calcio …
… Gipfel wird auch synonym für Spitzentreffen verwendet …
… Die edlen Metalle Kupfer, Silber und Gold kommen in der Natur bereits in einem hohen Reinheitsgrad gediegen (elementar) vor, während die unedlen Metalle in Form ihrer *Erze* auftreten, d.h., sie sind als Verbindungen (meistens als Sulfide und Oxide) in Gesteinen der Erdkruste enthalten …

4.7 Grenzfälle und Negativbeispiele

Wie bereits weiter oben angerissen, gibt es einige Textsegmente, die Definitionen auf den ersten Blick sehr ähnlich sehen. Das hängt meistens damit zusammen, dass einzelne Merkmale oder sogar ganze Merkmalsbündel (z.B. Informationsstrukturen oder Formulierungsmuster) darauf hindeuten, dass es sich um eine Definition handeln könnte. In Tabelle 12 sind einige solcher Beispiele aufgeführt. Dabei sind die verführerischen Bereiche blau markiert. Trotz dieser Merkmale handelt sich bei keinem der Beispiele in Tabelle 12 um eine Definition.

Tabelle 12: Beispiele für Textsegmente mit irreführenden Formulierungsmustern / Informationsstrukturen

… den Funktionen des Hubs entsprechend, bestehen die neuen Anlagen aus drei Hauptelementen: erstens dem als Hochregallager ausgebildeten Sendungslager für Euro-Paletten, zweitens den Umladestationen und drittens dem luftseitig ausgerichteten sogenannten ULD-Stacker, der leere und beladene Frachtbehälter für Flugzeuge aufnimmt. Sendungslager und ULD-Stacker sind vollständig computerisiert und mechanisiert, während im Umschlagbereich mechanisch unterstützte Handarbeit geleistet wird …
… man setzte ihn täglich an das Tor des Tempels, das man die Schöne Pforte nennt; dort sollte er bei denen, die in den Tempel gingen, um Almosen betteln…
… die Agiv, ein Konzern, bestehend aus drei in verschiedenen Feldern tätigen Maschinenbau- und einem Meßtechnikunternehmen, wird als Ganzes aber nur schwer zu verkaufen sein …
… eine Provinzialregierung wird vor dem 30. Juni gebildet und aus einem Präsidenten und fünf bis zehn Mitgliedern bestehen, die vom Provinzialrat bestimmt werden …
… der derzeit heiß diskutierte Sachs-Plan, benannt nach dem US-Wirtschaftsberater, der ihn für Polen aufgestellt hat, schlägt die Abschaffung aller Subventionen - mit Ausnahme der für den Wohnungsbau - und zugleich die Einführung der Zloty-Konvertibilität vor …

Außerdem ist besondere Vorsicht geboten bei Textsegmenten, die ein Gefühl oder eine Stimmung ausdrücken. Hier handelt es sich i. d. R. **nicht um definitorische Textsegmente** im Sinn dieser Richtlinien. Zu dieser Regel gibt es nur eine einzige Ausnahme: Das Gefühl bzw. die Stimmung bezieht sich auf eine Stilmarkierung bzw. Konnotation, die der Sprecher / Schreiber dem Definiendum zuordnet (vgl. das Beispiel zum Begriff "Neger" in Tabelle 2). In diesen Fall und nur dann, darf das Segment als Definition annotiert werden.

Bei der praktischen Annotation können Textsegmente vorkommen, die einige Annotatoren klar als Definition ansehen und andere Annotatoren bewusst mit O markieren. Solche Textsegmente sehen wir als Grenzfälle an. Ob ein Textsegment als definitorisch angesehen wird oder nicht, hängt nämlich nicht nur vom Vorhandensein der oben beschriebenen Merkmale ab, sondern auch davon, ob die in dem Textsegment genannten Informationen als relevant für die Bedeutung des Definiendums angesehen werden. Und genau hierbei spielt die persönliche Erfahrung der Annotatoren eine große Rolle. Daher lassen sich Unterschiede bei der Annotation derselben Textsegmente durch verschiedene Annotatoren nicht ausschließen. Wichtig ist jedoch, dass diese Unterschiede so klein wie möglich bleiben und so wenig wie möglich Individualismus in die Annotation einfließt. Daher sollte man sich bei der Annotation an die folgenden zwei Grundsätze halten:

1. Lieber ein Textsegment zu viel als eins zu wenig mit BETS annotieren.
2. Muss man das Textsegment 5x drehen und wende, um ihm doch noch ein paar definitorische Eigenschaften abzupressen, dann ist es sehr wahrscheinlich keine Definition und sollte mit O annotiert werden.

5 FAQ

Abschließend haben wir noch einige Fragen zusammengetragen, die man sich während der Annotation z.B. zu ganz konkreten Beispielen stellen könnte. Natürlich haben wir bisher nur die wichtigsten Aspekte aufgenommen. Dieser Bereich soll aber kontinuierlich erweitert werden und daher kann es sinnvoll sein, eigene Fragen zu notieren und an die Autorin der vorliegenden Richtlinien weiterzugeben.

Was war noch mal ein Textsegment?

Ein Textsegment ist in der Regel ein Satz. Es kann aber auch z.B. ein längerer Teilsatz als Textsegment behandelt werden. Einzelne Wörter gelten allerdings nicht als Textsegmente. Die Annotation bezieht sich nämlich immer auf größere Einheiten.

Kann eine Übersetzung als Definition zählen?

Ja, eine Übersetzung gilt als Definition, wenn denkbar ist (und das ist es fast immer), dass es irgendjemanden gibt, dem die Übersetzung beim Verständnis eines Begriffs hilft. Wenn man ein Textsegment mit Wort (= Definiendum) und Übersetzung findet, sollte man das immer als Definition annotieren. Ein Beispiel dazu ist auch in
Tabelle 11 gezeigt.

Ich habe ein Textsegment gefunden, das aussieht wie eine Definition aber auch wie ein Sprichwort oder eine Metapher. Was mache ich damit?

Kann das Textsegment jemandem beim Erschließen der Bedeutung eines Definiendums helfen? Wenn ja, dann ist es auf jeden Fall eine Definition.

Ich habe ein Textsegment gefunden, das wie ein spezieller Definitionstyp in einer bestimmten Wissenschaft aussieht (z.B. eine mathematische Definition). Was mache ich damit?

Ja, das kann vorkommen. Es gibt ganz unterschiedliche Traditionen des Definierens und der Definition in den verschiedenen wissenschaftlichen Disziplinen. Man sollte solche Textsegmente aber – auch wenn sie nicht leicht verständlich sind und syntaktisch / lexikalisch evtl. komisch aussehen – als definitorische Textsegmente betrachten und annotieren.

Ich habe ein Textsegment gefunden, das man mit einigen Schlussfolgerungen (logischen Folgerungen, Ableitungen, Analogiebildung etc.) als möglicherweise definitorisch ansehen könnte. Wie soll ich das annotieren?

Wenn man einem Textsegment über Schlussfolgerungen, Analogiebildungen oder Berechnungen etc. noch ein bisschen Bedeutung abzupressen muss, sollte man es in der Annotation nicht als Definition berücksichtigen.

Ich habe ein evtl. definitorisches Textsegment mit einem Eigennamen gefunden. Was mache ich?

Textsegmente mit Eigennamen als potentielle Definienda werden grundsätzlich nicht als Definition annotiert.

Ich habe ein Textsegment gefunden, das zu einem Begriff nur Beispiele nennt. Ist das eine Definition?

Ja, auf jeden Fall dann, wenn die Beispiele besonders typisch für den Begriff (das Definiendum sind). Im Prinzip kann man das als eine verkürzte Form der extensionalen Definition (Definition über alle zugehörigen Beispiele, Elemente eines Begriffs) ansehen.

Ich habe ein Textsegment gefunden, das eine negative Definition gibt. Ist das eine Definition?

Nein, das ist keine Definition. Die Aussage, *dass für X nicht Y oder Z gilt,* oder, dass *X nicht die Merkmale A und B hat*, ist ganz klar keine Definition. Das würde auch wieder Schlussfolgerungen erfordern und das hatten wir (vgl. die Frage oben) ja ausgeschlossen.

Ich habe ein Textsegment gefunden, in dem die Definition eines anderen Autors zitiert wird. Zählt das als definitorisches Textsegment?

Ja, auf jeden Fall.

Ich habe ein Textsegment gefunden, das aussieht wie eine Definition, aber im Konjunktiv formuliert ist. Wie gehe ich damit um?

Das ist egal, sind alle Bedingungen für das Vorliegen einer Definition erfüllt, so wir das Textsegment als definitorisch annotiert. Selbst wenn der Autor sich nicht sicher ist, ob seine Aussage richtig/falsch ist und das mit dem Konjunktiv oder einer Unsicherheitsfloskel markiert, spielt das – wie in Abschnitt 4.5 erläutert – für die Annotation keine Rolle. Sogar ein Experte kann sich schließlich unsicher sein.

Hier ist ein Textsegment, das aussagt, dass eine Bedeutung nur unter einer bestimmten Bedingung oder in einem bestimmten Fachgebiet / Kontext zutrifft. Gilt das trotzdem als Definition?

Ja, das ist trotzdem ein definitorisches Textsegment.

Mein Textsegment erläutert die Bedeutung von einem Adverb, Pronomen etc. Was mache ich denn damit?

Das kann nicht als Definition gewertet werden. In Abschnitt 2.1 wurden ja die Wörter und Wortgruppen eingeführt, die in diesen Richtlinien als Definiendum gewertet werden sollen. Unsere bisherige Arbeit hat nämlich gezeigt, dass die Definition von anderen Wortarten wenig sinnvoll bzw. hilfreich ist, deswegen berücksichtigen wir das auch nicht.

Ich habe ein Textsegment gefunden, das eine Meinung äußert (Formulierungen etwa mit: 'ich empfinde ... ', 'ich verstehe unter ...', 'ich bin der Meinung, dass ...'), wie soll ich damit umgehen?

Textsegmente in denen Meinungen geäußert werden, sind i. d. R. keine Definition. Zu dieser Regel gibt es nur einen Ausnahmefall, nämlich wenn in dem Textsegment darüber etwas ausgesagt wird, ob ein Begriff als Beleidigung etc. empfunden wird - dann wird nämlich etwas über die Stilmarkierung bzw. Konnotation eines Begriffs ausgesagt. Nur in einem solchen Fall darf das Textsegment als definitorisch annotiert werden. In jedem anderen Fall handelt es sich nicht um eine Definition.

Ich habe ein Textsegment gefunden, das die Struktur oder bestimmte Formulierungsmuster von Definitionen aufweist. Ich bin mir aber nicht sicher über Definiendum / Definiens. Wie gehe ich damit um?

Es gibt Textsegmente, die einzelne Merkmale oder sogar Bündel von Merkmalen von Definitionen aufweisen und trotzdem keine sind (vgl. hierzu Abschnitt 4.7).

6 Quellen

Alle in diesen Richtlinien aufgeführten Beispiele sind, wenn nichts anderes angegeben ist, den folgenden Quellen entnommen:

- Im Hinblick auf Explikatorische / definitorische Textsegmente vorannotierte Stichprobe des deutschen Wacky-Corpus (vgl. http://wacky.sslmit.unibo.it/doku.php Stand: 2008-02-09);
- Wellinghoff, S. (2006). Dokumentation: Manuelle Annotation Explikatorischer Textsegmente incl. Guidelines Phase I und II. Projektdokumentationen Hytex. Institut für deutsche Sprache und Literatur, TU Dortmund;
- Mönke, H. (1978): Definitionstypen und Definitionsmatrix. In: Nachrichten für Dokumentation, 29, S. 51-60;
- Herbst, T. / Klotz, M. (2003): Lexikografie. Eine Einführung. UTB.
- sowie den Unterlagen und Referaten des Seminars *Bedeutung und Bedeutungsexplikation*, Leitung Prof. A. Storrer, Wintersemester 2007/2008, TU Dortmund.