Eine umfassende Constraint-Dependenz-Grammatik des Deutschen

Kilian A. Foth

24. Mai 2004

Inhaltsverzeichnis

Dies ist eine umfassende Constraint-Dependenz-Grammatik des Deutschen. Sie soll im Prinzip beliebigen Text abdecken können, setzt allerdings aus mehreren Gründen geschriebenen Input voraus:

1. Einige Regeln verwenden die An- und Abwesenheit von Satzzeichen, um bestimmte Konstruktionen zu erlauben. Ohne Satzzeichen sind diese Regeln wirkungslos oder gar hinderlich.

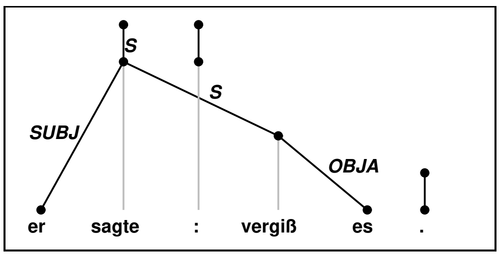
2. Eindeutige sprachliche Fehler werden insofern robust verarbeitet, daß es ungeachtet der Eingabe immer mindestens eine mögliche Struktur geben sollte. Jedoch ist dies nicht immer diejenige, die der Sprechintention entsprechen würde. Beispielsweise wird ein Kongruenzfehler gewöhnlich nicht zu einer veränderten syntaktischen Struktur führen, ein kategorieändernder Schreibfehler aber sehr wohl ("Die Firma will mehr Rechner *verkaufe*"). Daher wird Spontansprache gewöhnlich nicht optimal modelliert.

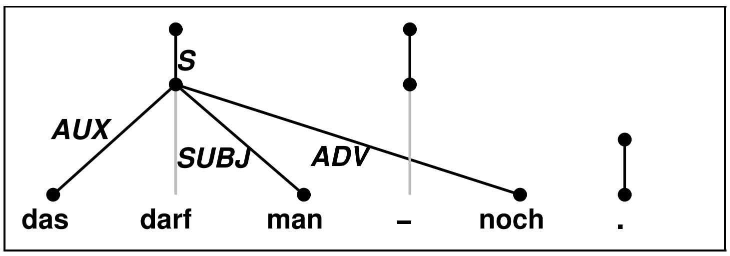
1 Elemente der Grammatik

1.1 Die Syntaxebene

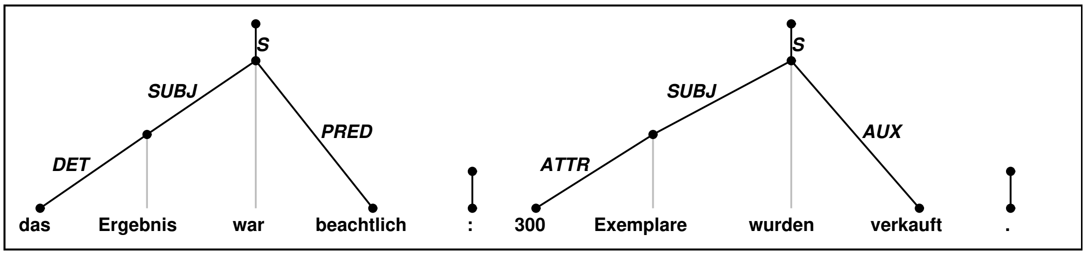
Nur die Ebene `SYN` stellt die eigentliche Modellierung der Satzstruktur dar. Die anderen Ebenen dienen dazu, Verbindungen zwischen solchen Worten zu markieren, die nicht direkt syntaktisch untergeordnet sind, aber dem Parser dabei helfen, zwischen sonst ununterscheidbaren Strukturen zu wählen. Zum Beispiel darf ein Relativsatz (label `REL`) nur auftreten, wenn er ein Relativpronomen enthält; das Relativpronomen ist aber nicht immer direkt mit dem Nebensatzverb verbunden, daher muß diese Beziehung außerhalb des Syntaxbaumes markiert werden.

Die syntaktische Struktur wird als eine Analyseebene `SYN` aufgebaut. Im allgemeinen sollte eine durch Punkt abgeschlossene Äußerung durch einen zusammenhängenden Baum dargestellt werden. Es gibt jedoch Ausnahmen; die Satzzeichen `:`, `-`, `()` und `;` werden manchmal zur Unterteilung eines vollständigen Satzes verwendet:





Manchmal aber dienen sie zur Abgrenzung zwischen zwei vollständigen Sätzen. In diesem Fall ist es angebracht, zwei Syntaxbäume aufzustellen:



Constraints erlauben daher ausdrücklich mehrere Hauptsätze ohne Bestrafung, wenn sie durch solche Satzzeichen getrennt sind.

1.2 Label der Syntaxebene

Alle Dependenzkanten der Syntaxebene tragen ein erklärendes Label. Es wird grob unterschieden zwischen

- Komplementen: Das sind alle Arten von Wörtern, die mehr oder weniger regelmäßig mit ihrem Regenten auftreten: Subjekte, Artikel, Objekte etc.
- Modifikatoren: Das sind alle anderen Beziehungen.

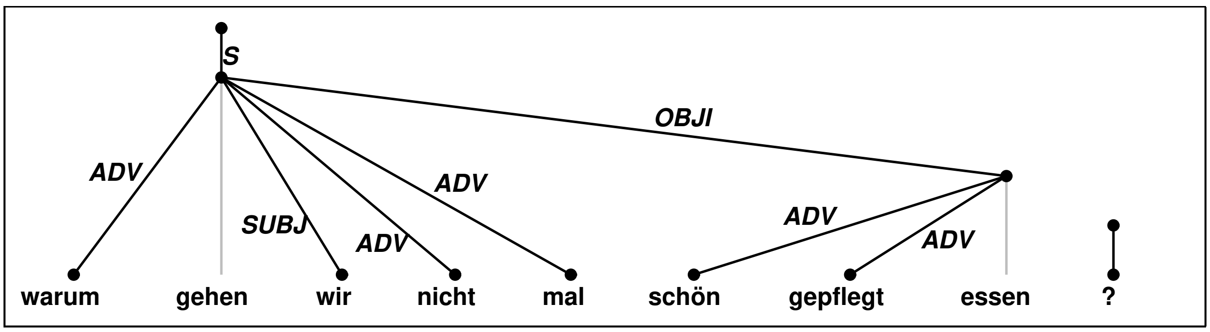
Komplemente werden durch folgende Label ausgedrückt:

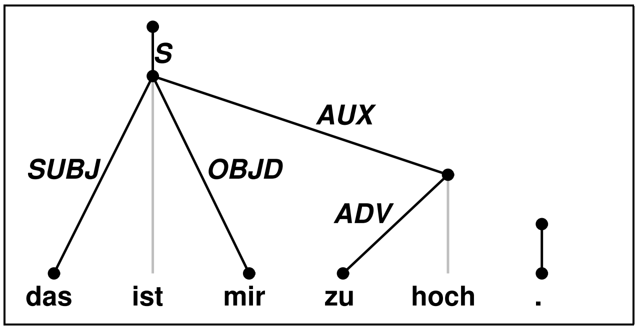
`AUX AVZ CJ DET PN PRED OBJA OBJA2 OBJC OBJD OBJG OBJI OBJP SUBJ`

Alle anderen Label sind Modifikatoren, d.h. niemals obligatorisch.

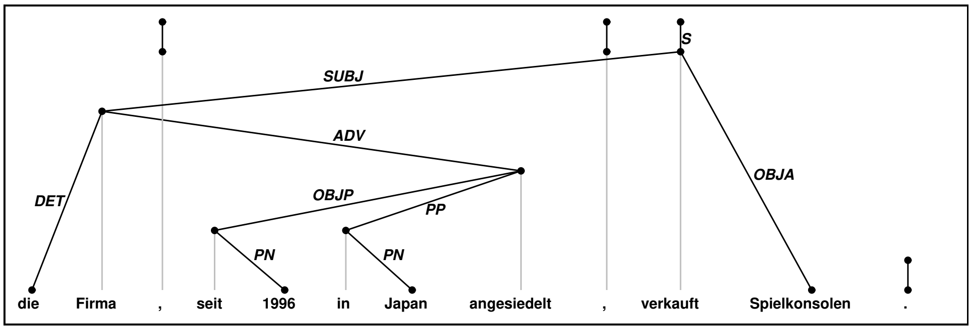
1.2.1 Das Label ADV

`ADV` bezeichnet adverbiale Modifikation, entweder durch wirkliche Adverben (ADV) oder durch verwandte Wortklassen wie ADJD, PTKNEG, VVPP, PWAV oder PTKA.





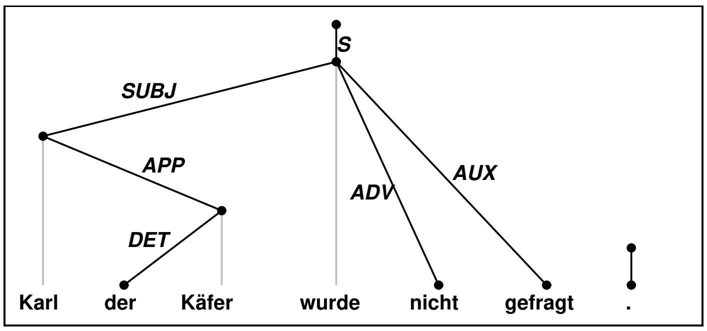
Als ADV wird auch eine nachgestellte Modifikation von Nomen durch prädikatives Adjektiv bezeichnet:

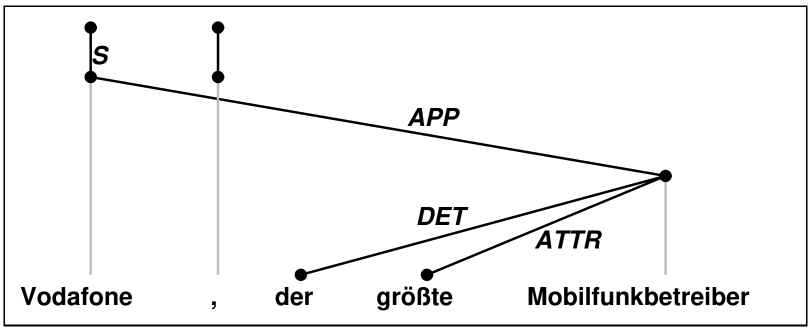


Inhaltlich könnte diese Konstruktion auch als elliptischer Relativsatz angesehen werden; da kein Relativpronomen auftritt, betrachten wir sie aber als Art der adverbialen Bestimmung.

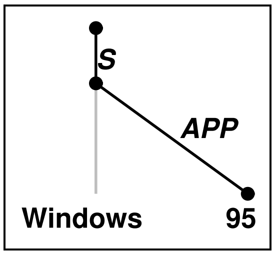
1.2.2 Das Label APP

Das Label APP verbindet aufeinanderfolgende Worte derselben NP, falls sie nicht Determiner oder Attribute sind (Artikel, attributive Pronomen, attributiv gebrauchte Zahlen etc.). Der Normalfall ist die Beziehung zwischen zwei Nomen (NN, NE oder FM), also eine echte Apposition:

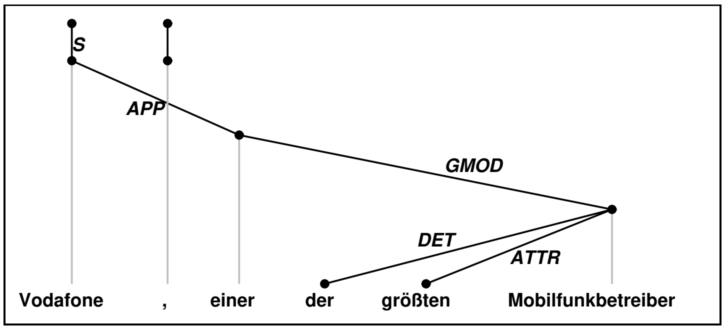




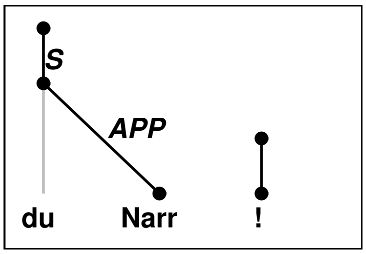
Auch andere aufeinanderfolgende Bestandteile einer Nominalphrase werden mit APP bezeichnet, zum Beispiel nachgestellte Zahlen:

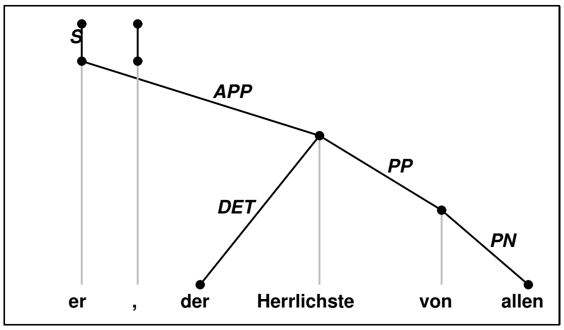


Schließlich können auch substituierende Pronomen als Appositionen auftreten:

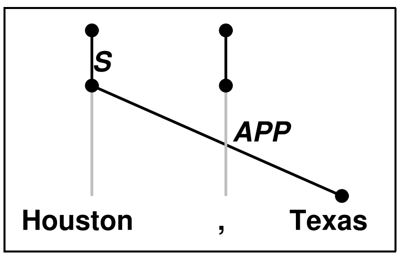


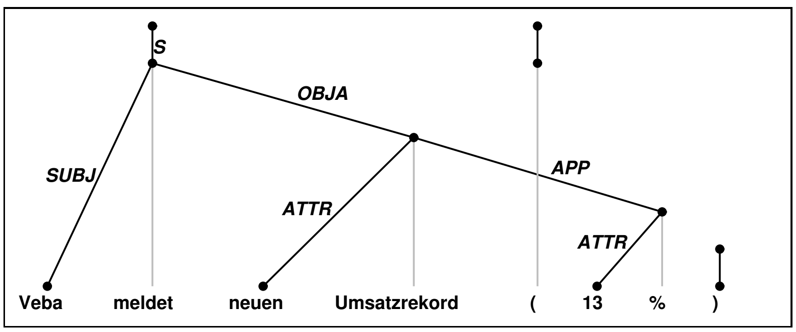
Das modifizierte Wort ist normalerweise ein Nomen. Nur ganz selten werden Personalpronomen modifiziert:





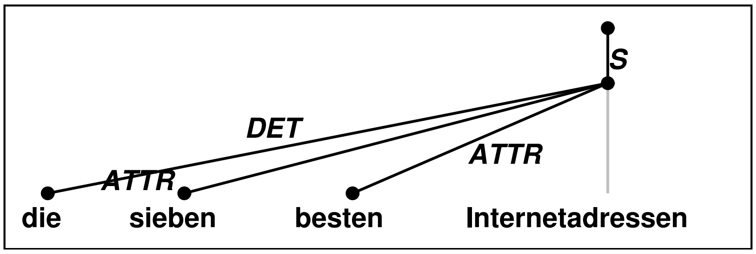
Es gibt eine Reihe von erläuternden Beziehungen zwischen Nominalphrasen, die nicht im strengen Sinne Appositionen (d.h. Gleichsetzungen) sind, sondern nur enge Beziehungen (z.B. ‘A hat B’, ‘A befindet sich am Ort B’). Solche Beziehungen dürfen ebenfalls als APP modelliert werden, wenn sonst keine Unterordnung möglich ist:



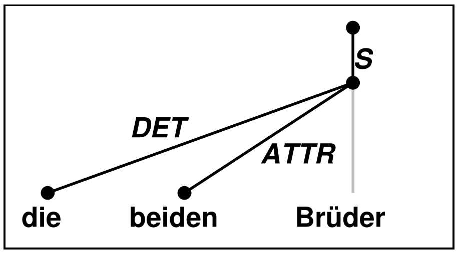


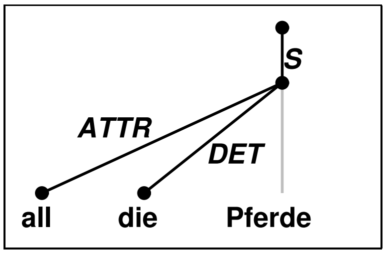
1.2.3 Das Label ATTR

Dies Label bezeichnet Attribute von Nomen. Meistens sind das attributive Adjektive oder Zahlen:



Wenn ein attributives Pronomen (`PIDAT`) zusammen mit einem Artikel o.ä. steht, dann ist dieser `DET` und das Pronomen `ATTR`:

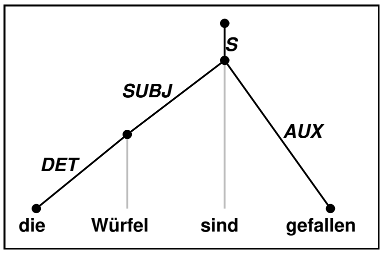


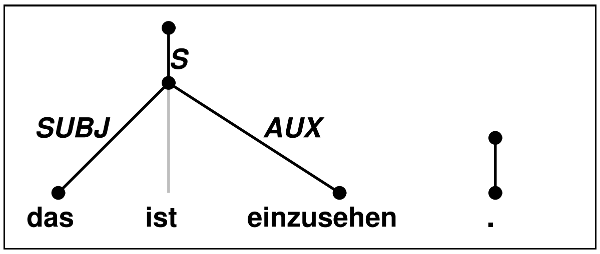


1.2.4 Das Label AUX

Mit `AUX` werden Verbgruppen zusammengefügt, die aus Hilfsverb und Vollverb bestehen. Grundsätzlich ist immer das finite Verb der Kopf der gesamten Verbgruppe; in nicht-finiten Verbphrasen ist das Vollverb dem Hilfsverb untergeordnet.

Das Verb 'sein' kann viele verschiedene Wortarten als Komplement nehmen: Nomen, Infinitive, Präpositionen, Adverben etc. Wenn es sich um ein Verb handelt, wird es mit `AUX` bezeichnet:



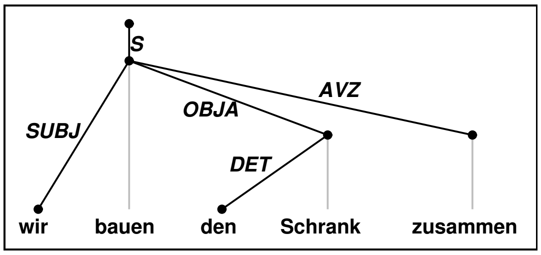


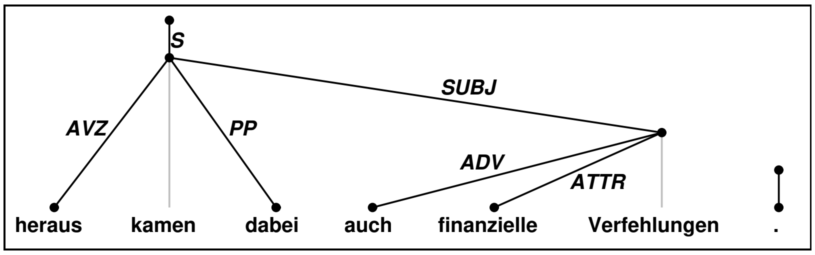
Anderenfalls handelt es sich um ein Prädikat, das mit `PRED` bezeichnet wird (vgl. unten).

Das Verb 'haben' kann einerseits als Auxiliarverb Verbphrasen bilden, andererseits ein normales Akkusativobjekt (`OBJA`) tragen.

1.2.5 Das Label AVZ

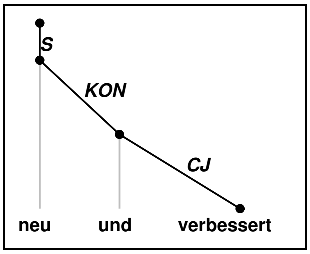
Diese Kante verbindet ein abgetrenntes Verb-Präfix mit seinem finiten Verb. Steht das Präfix alleine, so muß es dem Verb folgen oder unmittelbar vorausgehen.

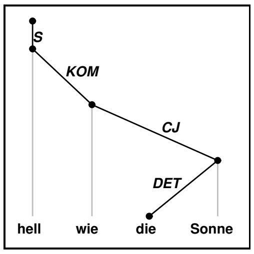


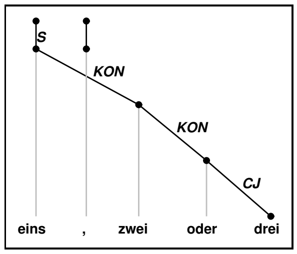


1.2.6 Das Label CJ

Hiermit wird das Komplement einer beiordnenden oder vergleichenden Konjunktion bezeichnet, also z.B. von “und”, “oder”, “wie” und “als”. Es kann vielen verschiedenen Wortklassen angehören.







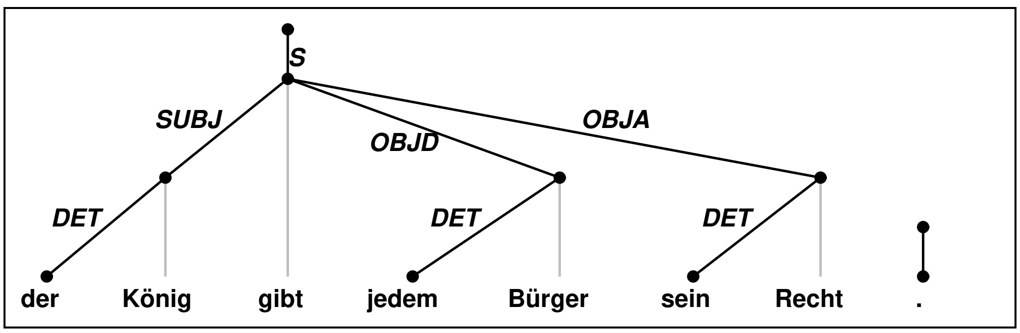
Das Komplement wird immer mit `CJ` bezeichnet, egal welche logische Funktion es erfüllt.

Einige Konjunktionen können mehr als zwei Elemente beiordnen. In diesem Fall wird nur das letzte mit `CJ` bezeichnet, alle anderen mit `KON`.

Zur Struktur von koordinierten Phrasen vgl. das Label `KON`.

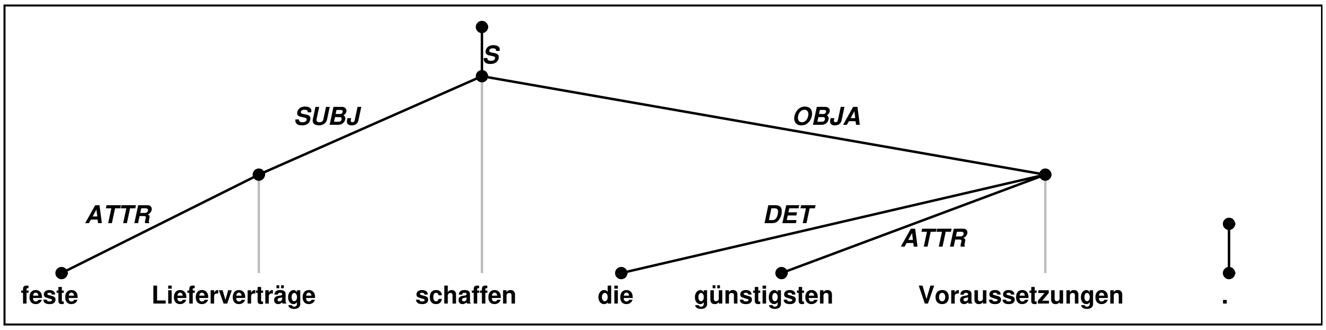
1.2.7 Das Label DET

Dieses Label bezeichnet den Determiner eines Nomen. Das sind gewöhnlich bestimmte oder unbestimmte Artikel, aber auch alle Arten von attributiven Pronomen:

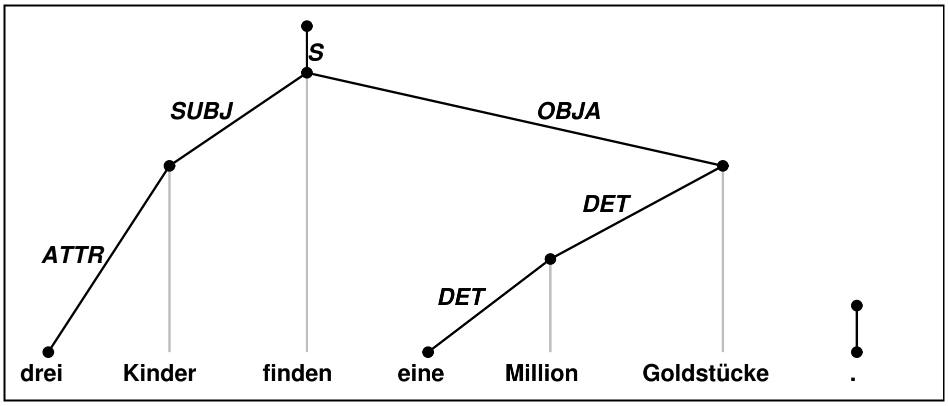


Artikel sind immer mit `DET` zu bezeichnen, außer sie sind beigeordnet (dann `CJ`) oder fragmentarisch (dann `S`).

Adjektive sind niemals `DET`, sondern immer `ATTR`:

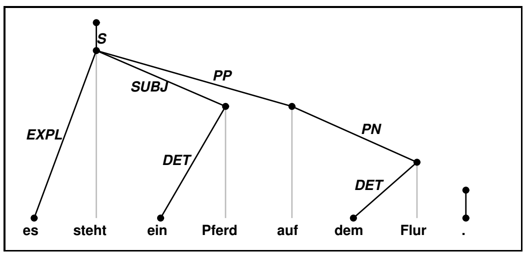


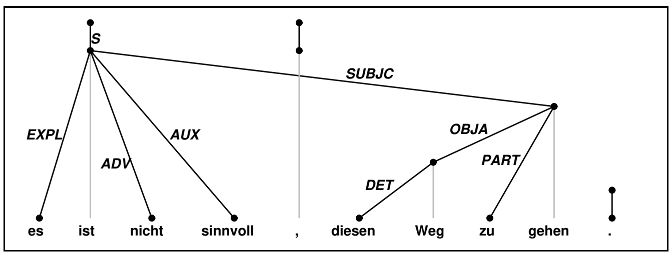
Auch der unbestimmte Artikel oder Zahlen, die Nomina sind, sind `DET`, aber normale Zahlen werden wie Adjektive mit `ATTR` bezeichnet.



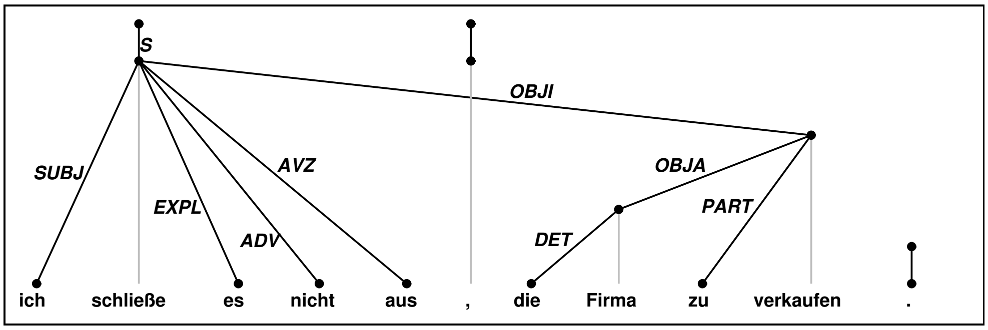
1.2.8 Das Label EXPL

Dieses Label bezeichnet das expletive 'es' an einem finiten Verb. Es tritt in zwei Varianten auf. Im Hauptsatz steht es vor dem Verb, wenn das Subjekt nachgestellt wird und keine andere Konstituente das Vorfeld füllen kann. (Das nachgestellte Subjekt kann auch ein Subjektsatz sein.)

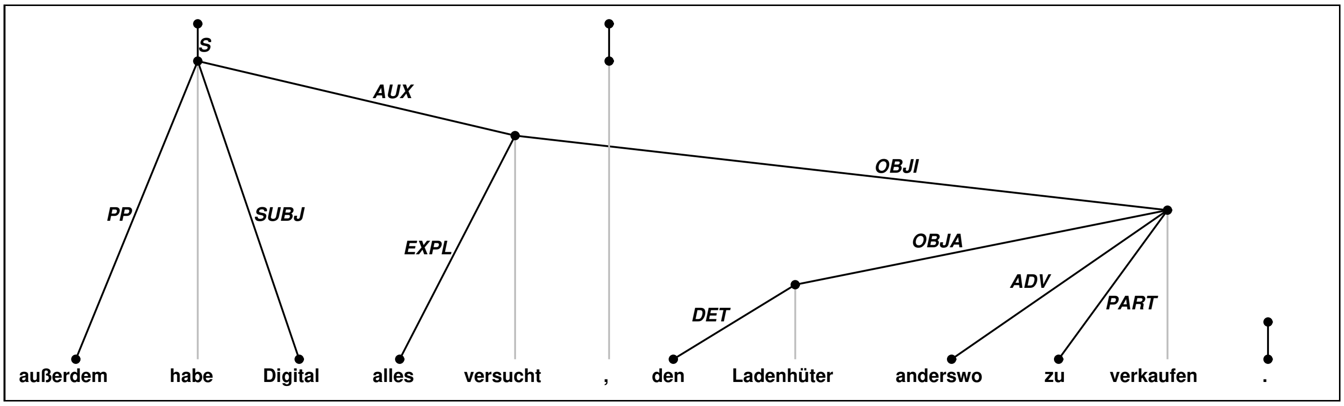




Nach dem Verb kann es als Vertreter eines Objektsatzes stehen:



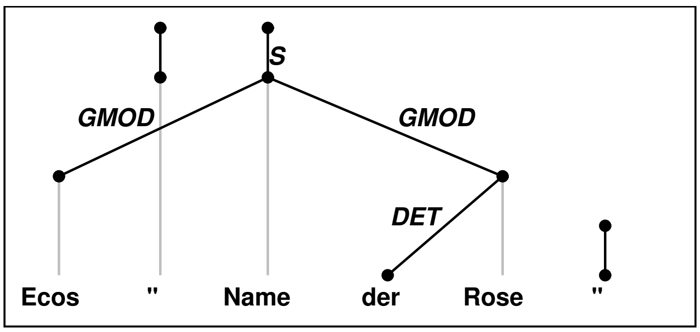
Bisweilen steht auch ein anderes Pronomen zusammen mit dem eigentlichen Objekt:

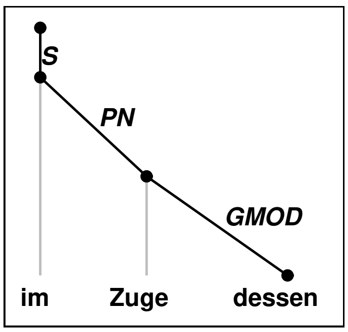


Möglicherweise ist dies auch ein expletiver Gebrauch. Derzeit erlaubt die Grammatik diese Struktur aber nicht.

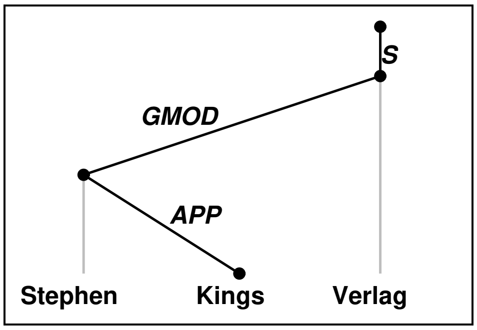
1.2.9 Das Label GMOD

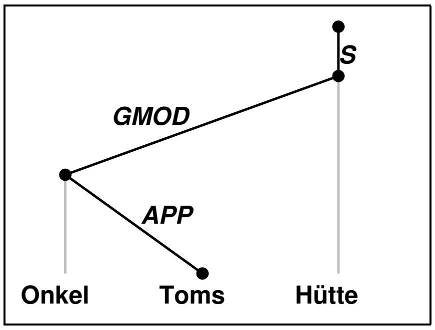
Dieses Label bezeichnet ein Genitivattribut. Sowohl das Attribut selbst als auch das modifizierte Wort ist ein Nomen oder ein substituierendes Pronomen.





Das Genitivattribut ist immer ein Wort, das erkennbar im Genitiv steht; entweder das Wort selbst oder sein Determiner muß eine Genitivendung besitzen. Bei Eigennamen kann die Genitivendung auch am Ende des Namens stehen:

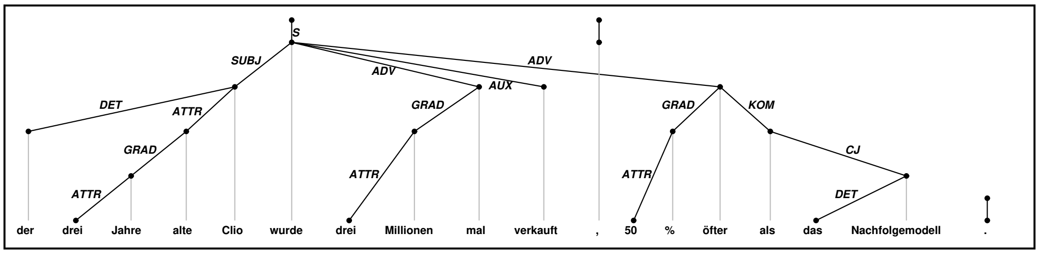




1.2.10 Das Label GRAD

Dieses Label bezeichnet eine im Akkusativ stehende NP, die als eine Art Maßangabe verwendet wird. Sie kann nur in ganz bestimmten Situationen auftreten:

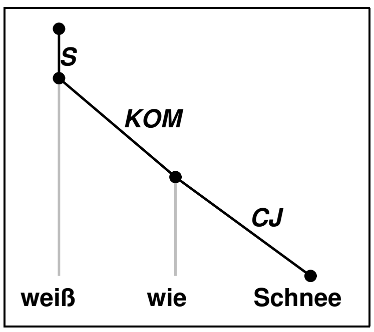
1. an einem Adjektiv, daß eine meßbare Eigenschaft ausdrückt (“hoch”, “schnell”). Alle diese Adjektive tragen das Feature `measurement`.
2. an einem Komparativ (“mehr”, “über”)
3. als `CARD` mit dem Adverb “mal”.

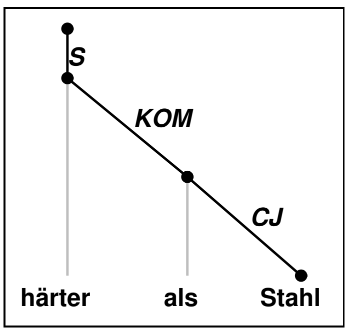


Die Maßangabe ist nur mit Zahlen und bestimmten Nomen möglich, nämlich Maßeinheiten (Jahr, Meter, Grad etc.).

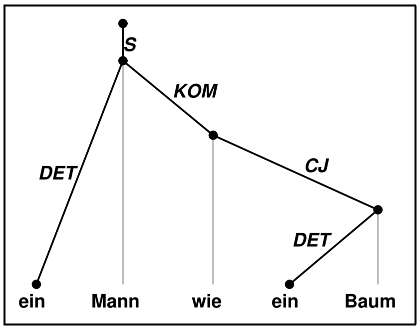
1.2.11 Das Label KOM

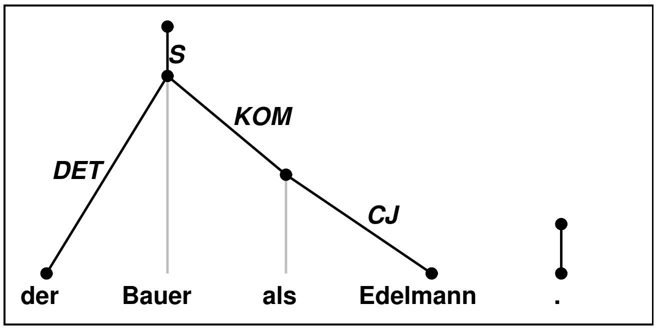
Mit `KOM` werden nur die Vergleichsworte “als” und “wie” untergeordnet. Sie können verschiedene Kategorien modifizieren. “als” modifiziert Adjektive und Pronomen im Komparativ, “wie” modifiziert Adjektive im Positiv.

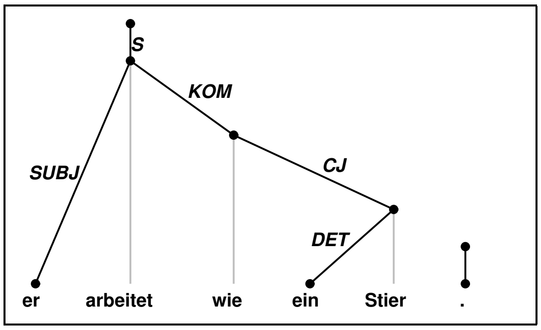


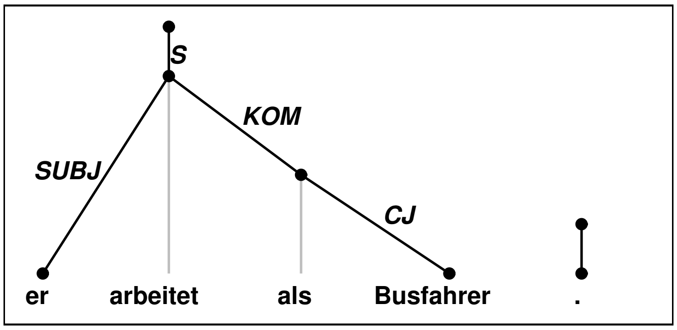


Beide Worte modifizieren Nomen und Verben:

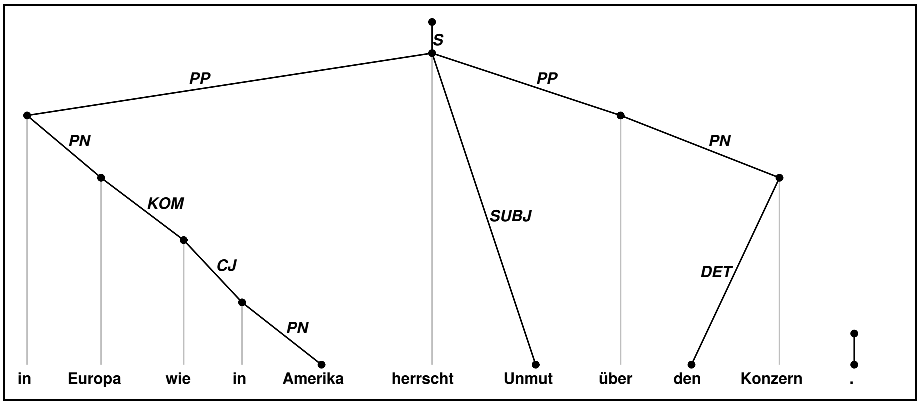




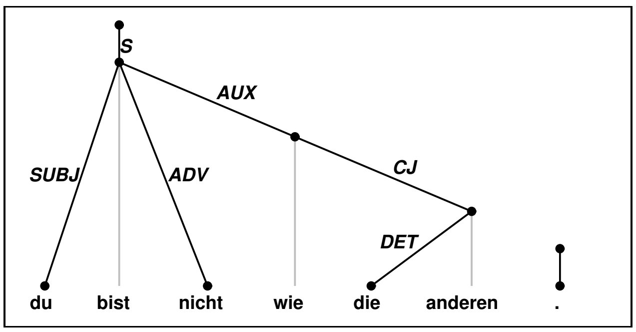


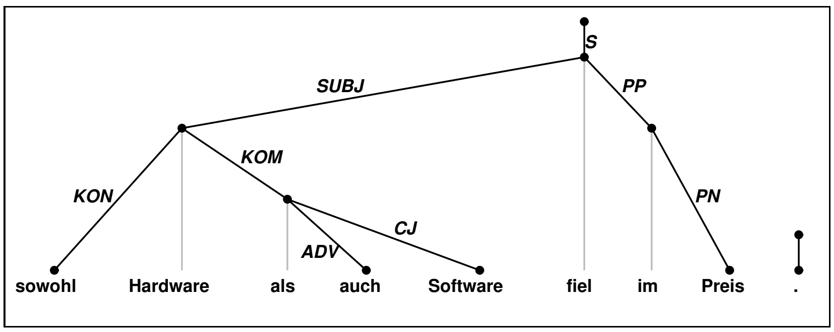


Das Wort “wie” kann auch Präpositionen modifizieren:



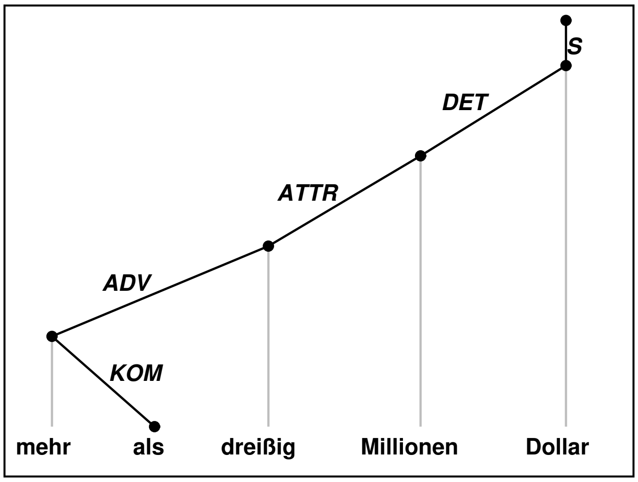
Beide Worte sind immer als `KOM` zu bezeichnen, wenn sie nicht Komplemente sind:





Nicht zu verwechseln mit den vergleichenden Konjunktionen “als” und “wie” sind die unterordnende Satzkonjunktion “als” (“als sieben Jahre vergangen waren”) und das Fragewort “wie” (“wie geht es dir?”).

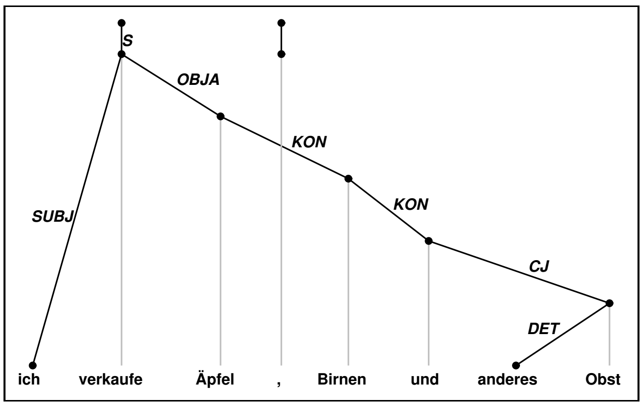
Die Konstruktion ‘mehr als’ bzw. ‘weniger als’ wird stets als ein adverbialer Modifikator behandelt,



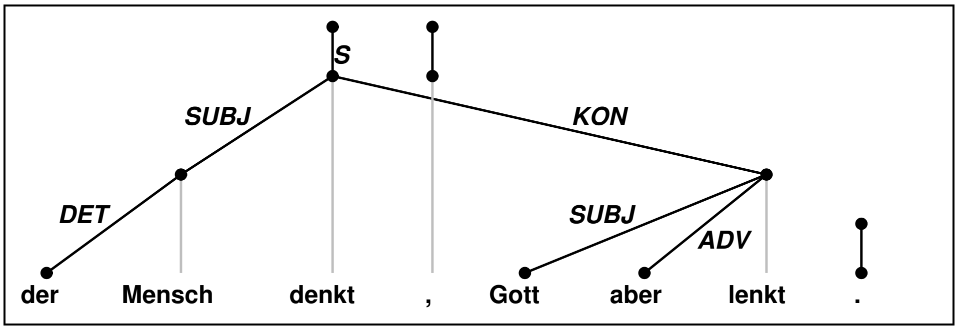
also tief angebunden, so wie in der Formulierung ‘über dreißig Millionen Dollar’.

1.2.12 Das Label KON

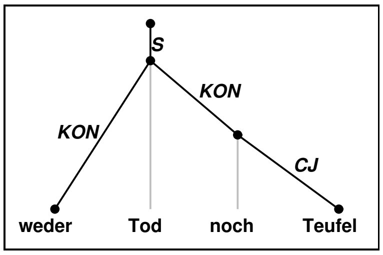
Beiordnungen werden als eine rechtsverzweigende Kette von Worten modelliert. Dabei trägt jedes Wort das Label `KON`, bis auf das letzte Wort, das unter der Konjunktion steht und mit `CJ` bezeichnet wird:



Einige Konjunktionen treten auch in adverbialer Form auf, d.h. nachgestellt. In diesem Fall sind sie als `ADV` zu bezeichnen, und die beigeordneten Worte modifizieren einander direkt als `KON`:



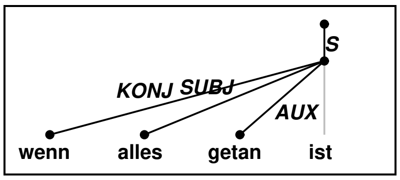
Einige Konjunktionen bestehen aus zwei Worten. Der linke Teil einer solchen Konjunktion wird als `KON` dem ersten Konjunkt untergeordnet:

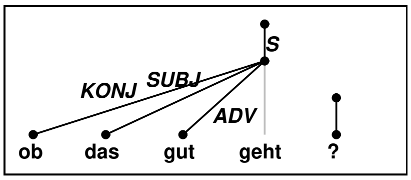


1.2.13 Das Label KONJ

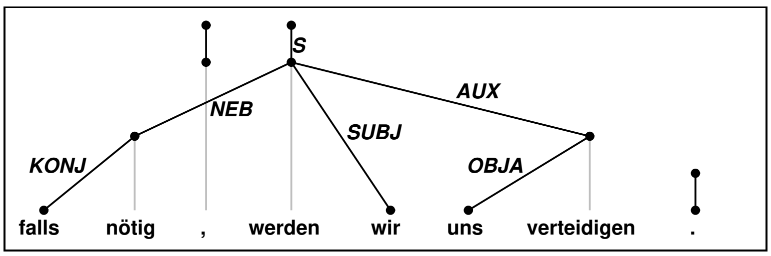
Das Label `KONJ` verbindet eine unterordnende Konjunktion mit ihrem Bezugswort. Gewöhnlich ist das das Nebensatzverb.

Satzkonjunktionen vom Typ `KOUS` modifizieren finite Verben:

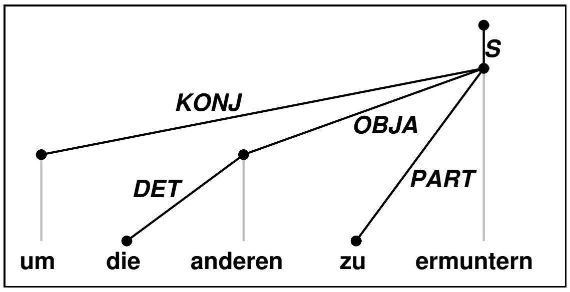


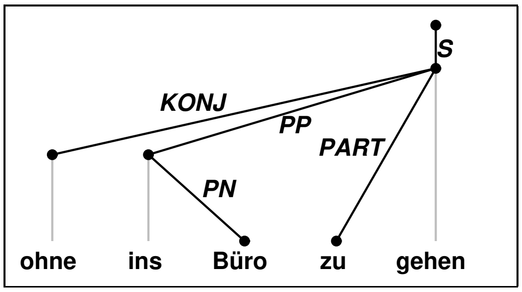


Bisweilen wird ein Nebensatz so weit verkürzt, daß nur noch ein Adjektiv verbleibt. In diesem Fall muß die Konjunktion das Adjektiv direkt modifizieren.

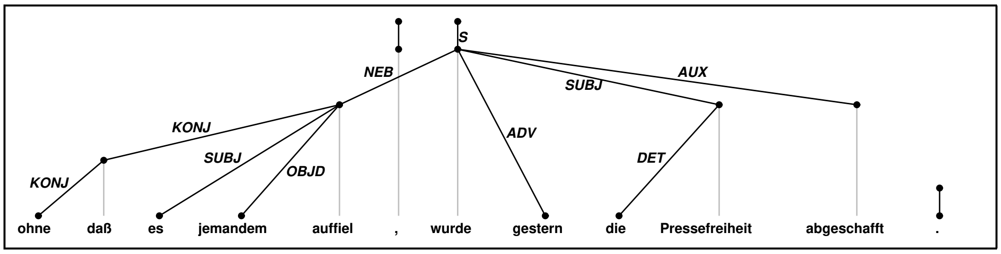


Konjunktionen vom Typ `KOUI` modifizieren erweiterte Infinitive:



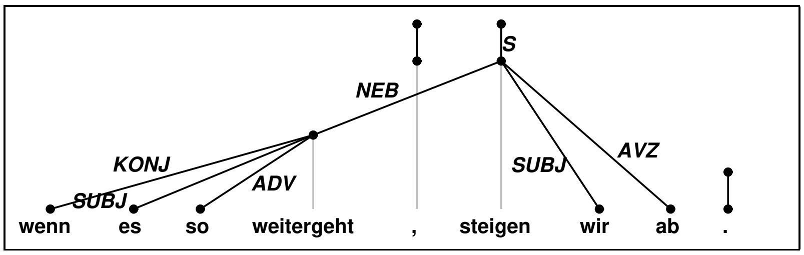


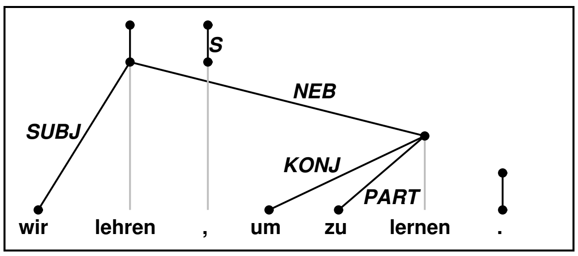
In den Konstruktionen “ohne daß” etc. soll die `KOUI` das “daß” modifizieren.

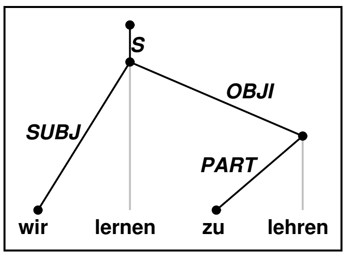


1.2.14 Das Label NEB

Das Label `NEB` verbindet das Verb eines Nebensatzes mit dem übergeordneten Wort.

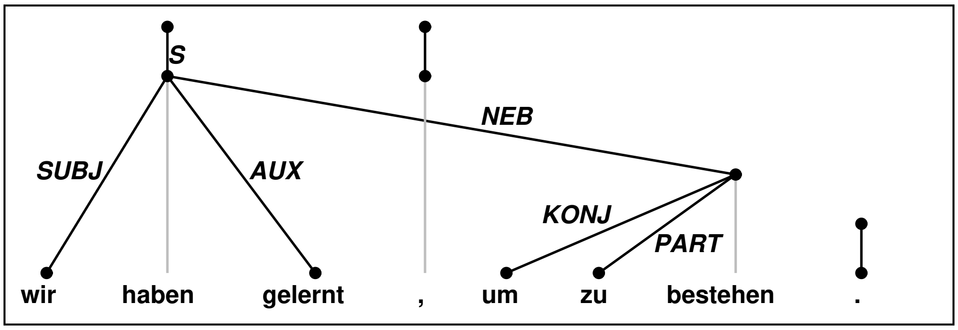






Das untergeordnete Wort muß ein finites Verb oder ein erweiterter Infinitiv mit einer Konjunktion vom Typ `KOUI` sein. Erweiterte Infinitive ohne eine solche Konjunktion sind immer Objektinfinitive (`OBJI`).

Wenn der Matrixsatz eine komplexe Verbgruppe besitzt, soll der Nebensatz das finite Verb modifizieren, nicht das Vollverb:

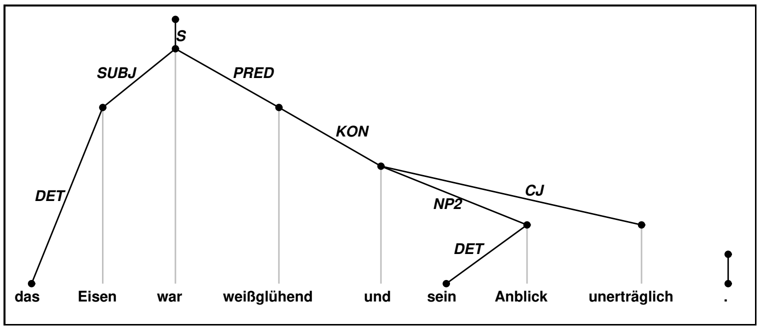


1.2.15 Das Label NP2

Das Label NP2 dient dazu, ein überzähliges logisches Subjekt in einer elliptischen Koordination einzubinden, damit es nicht als Fragment behandelt werden muß. Typischerweise geschieht dies in Sätzen wie dem folgenden:

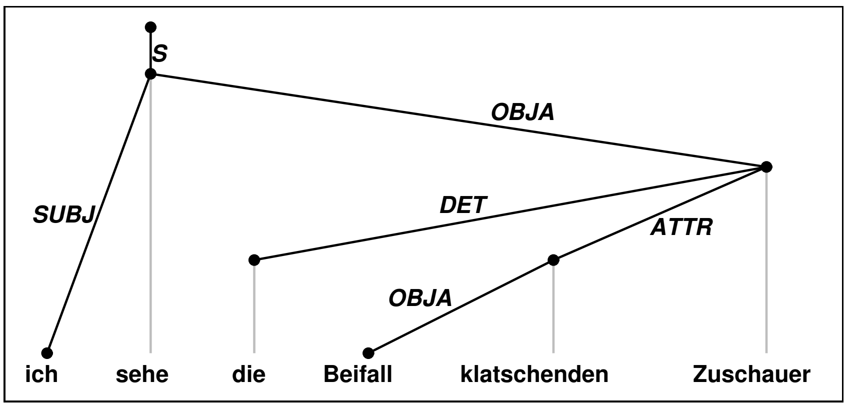
“Das Eisen war weißglühend und sein Anblick war unerträglich.”

“Das Eisen war weißglühend und sein Anblick unerträglich.”



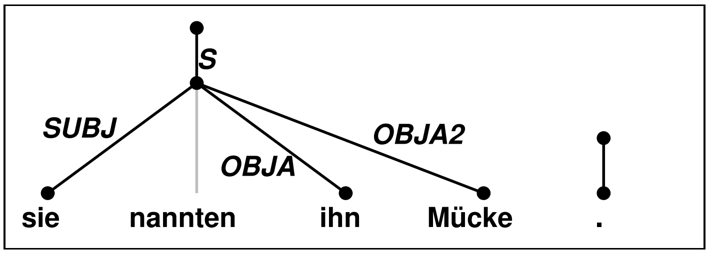
1.2.16 Das Label OBJA

Als Akkusativobjekt wird das zweite Komplement eines transitiven Verbs bezeichnet. Die Kante OBJA kann nur von Vollverben ausgehen (finit, infinit oder als Verbaladjektiv):



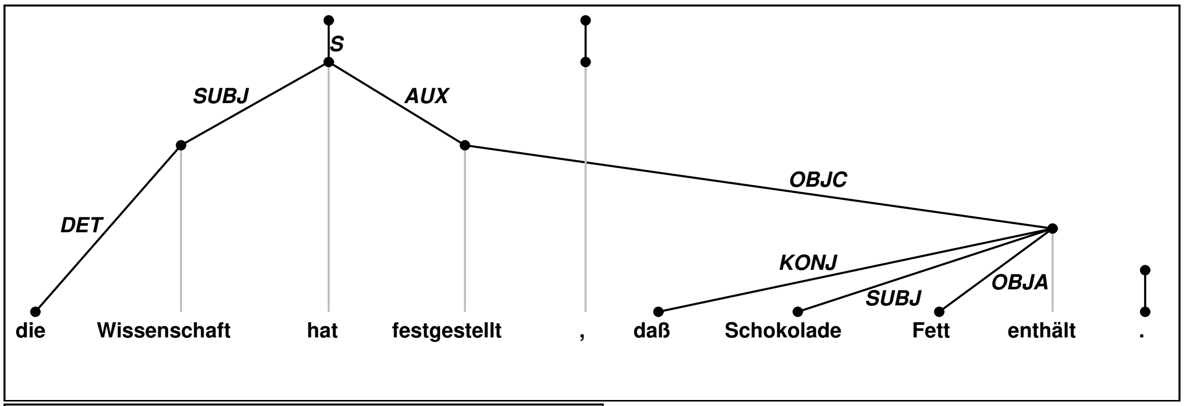
1.2.17 Das Label OBJA2

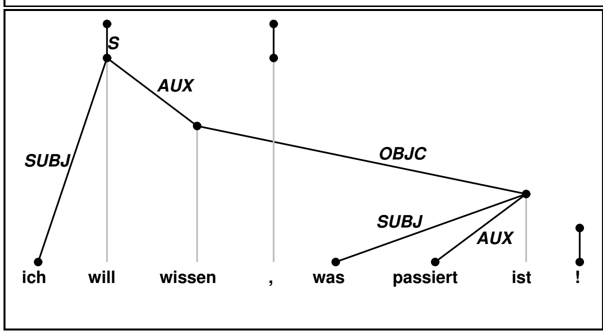
Dieses seltene Label kommt nur bei wenigen Verben vor, die zwei Akkusative verlangen (lehren, kosten, nennen).



1.2.18 Das Label OBJC

Ein Objektsatz ("object clause") ist ein finites Verb in Nebensatzstellung als Komplement eines anderen Verbs.

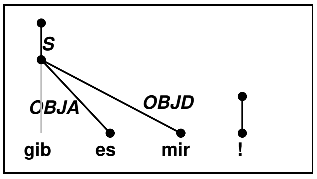


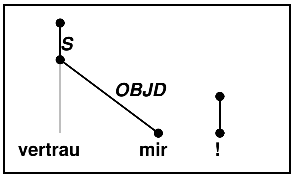


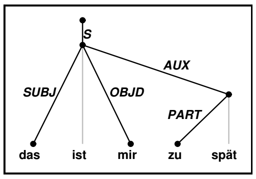
Solche Sätze treten nur mit 'daß' oder 'ob' oder Frageworten auf. Steht eine adverbiale Konjunktion am Verb, so handelt es sich um einen Nebensatz (`NEB`). Wenn der untergeordnete Satz in Hauptsatzstellung steht, ist er mit `S` zu bezeichnen.

1.2.19 Das Label OBJD

Das Dativobjekt kann entweder allein oder in Verbindung mit einem Akkusativobjekt auftreten. Was genau wo möglich ist, hängt vom Vollverb ab. Im Gegensatz zum Akkusativobjekt kann das Dativobjekt auch mit ‘sein’ auftreten.

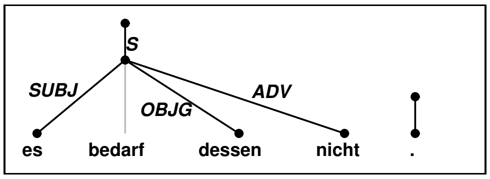






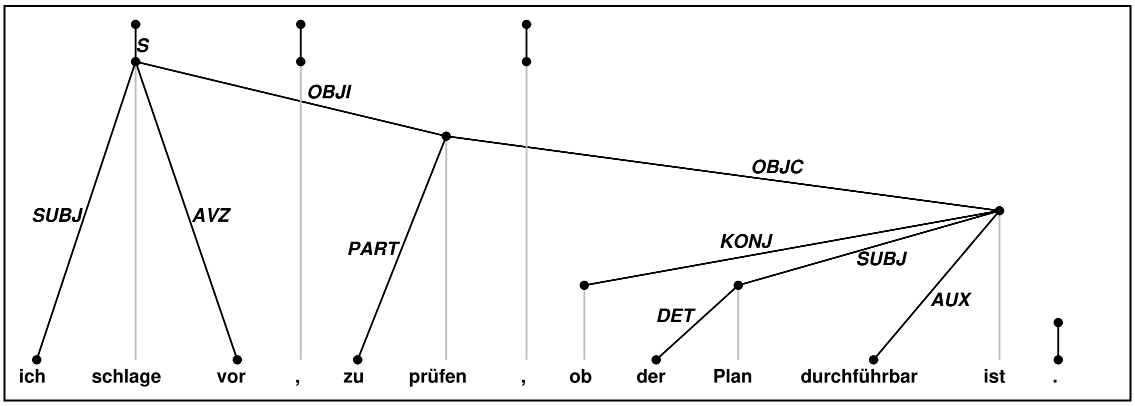
1.2.20 Das Label OBJG

Das Genitivobjekt ist das seltenste aller Objekte. Nur sehr wenige Verben verwenden es, und es ist immer obligatorisch.

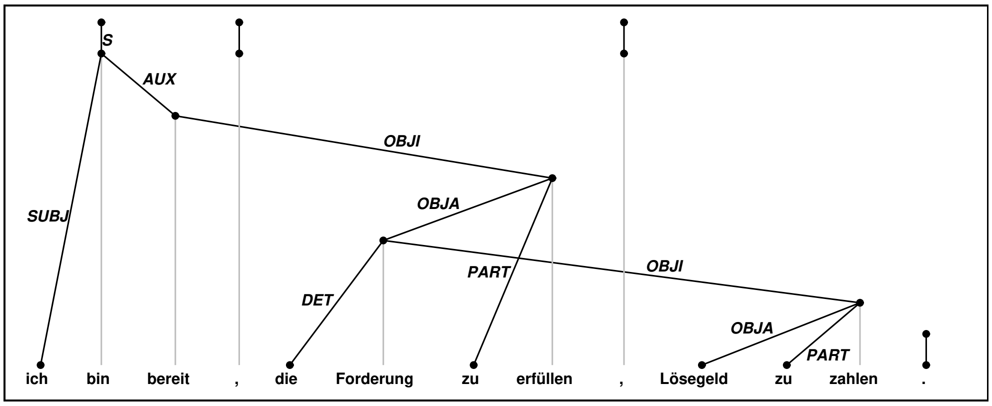


1.2.21 Das Label OBJI

Ein Objektinfinitiv ist ein Infinitiv als Komplement eines anderen Verbs. Nur bestimmte Verben können einen Objektinfinitiv tragen, und meistens kann er auch durch einen Akkusativ ersetzt werden.

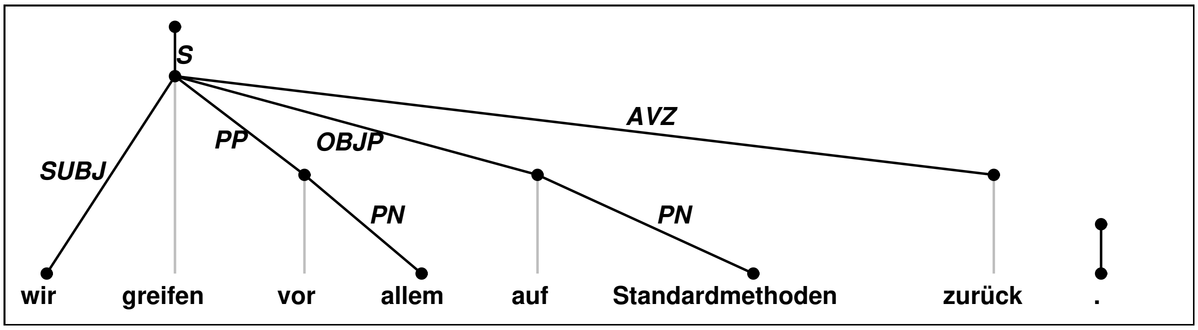


Bestimmte Nomen und Adjektive können ebenfalls Objektinfinitive nehmen. Meistens sind dies deverbale Wörter:



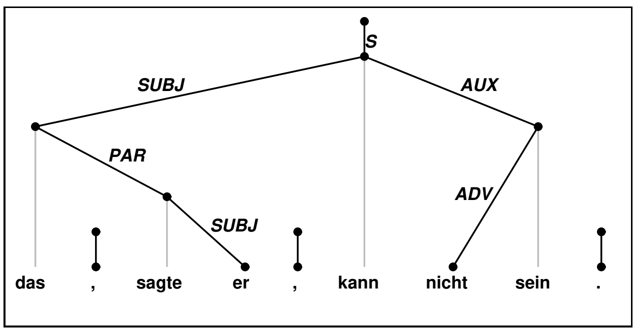
1.2.22 Das Label OBJP

Als `OBJP` wird eine Präpositionalobjekt bezeichnet, also eine Präpositionalphrase, die das Komplement eines Verbs ist. Im Unterschied zu einer normalen `PP` darf sie nicht fortgelassen werden.



1.2.23 Das Label PAR

Dies Label bezeichnet *Parenthesen*, also Einschübe, von Matrixsätzen in den logisch untergeordneten Objektsatz. Immer wenn die Unterordnung eines Objektsatzes einen nichtprojektiven Baum erzeugen würde, wird statt dessen der Matrixsatz dem Objektsatz untergeordnet. Das Label ist dann `PAR`, und der Regent ist das letzte Wort vor der Parenthese.

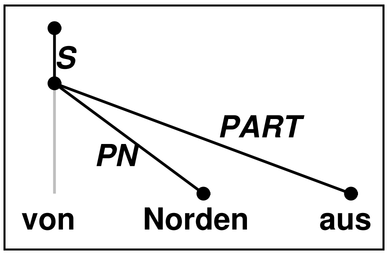


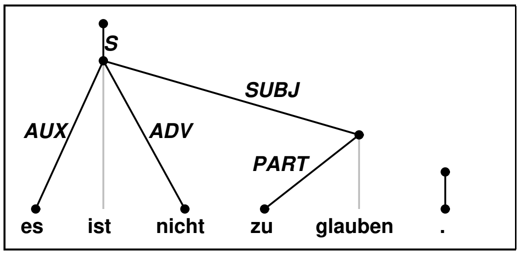
Es gibt verschiedene andere Konstruktionen, die manchmal Parenthese genannt werden; diese werden **nicht** als `PAR` bezeichnet, sondern so wie immer. Beispielsweise ist eine Präpositionalphrase `PP` und nicht `PAR`, selbst dann, wenn sie nicht in die normale Satzstruktur paßt, etwa im Vor-Vorfeld.

1.2.24 Das Label PART

Dieses Label bezeichnet die Unterordnungen von mehreren sehr stark eingeschränkten Partikeln:

1. eine Zirkumposition (`APZR`) modifiziert eine Präposition
2. das Wort "zu" (`PTKZU`) modifiziert einen Infinitiv

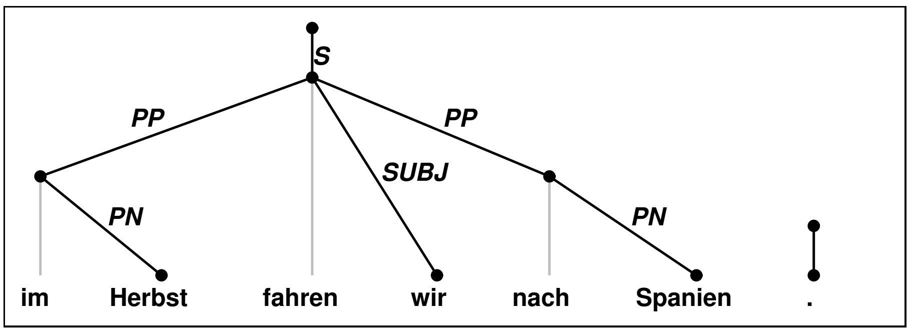




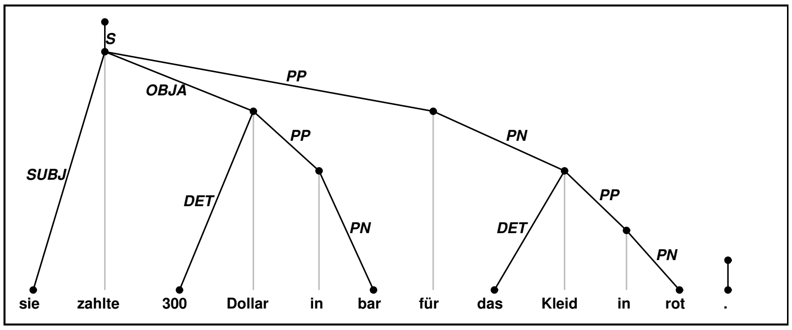
(Im ersten Fall ist die Partikel genau genommen ein Komplement und kein Modifikator; aber im Deutschen gibt es nur solche Zirkumpositionen, die auch als einfache Präpositionen dienen können, daher wird die Partikel immer als optional behandelt.)

1.2.25 Das Label PN

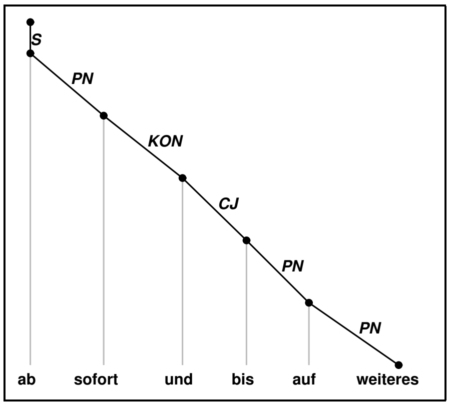
Dieses Label bezeichnet das Komplement einer Präposition. Gewöhnlich sind das Nomen und Pronomen:



Einige Präpositionen nehmen aber auch prädikative Adjektive:

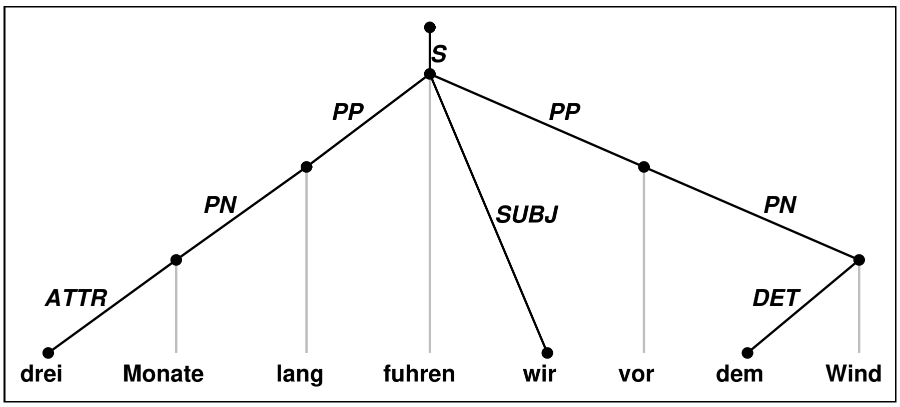


Andere Präpositionen nehmen Adverben oder wieder Präpositionen:



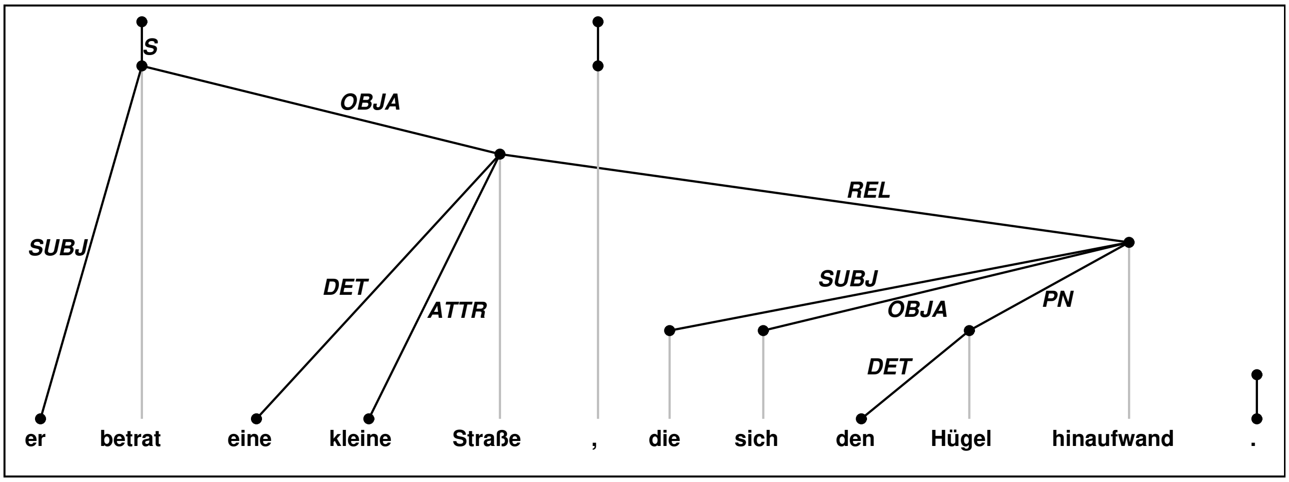
Hierbei handelt es sich fast immer um feststehende Fügungen.

Die wenigen Postpositionen, die das Deutsche aufzuweisen hat, nehmen ebenfalls Argumente mit dem Label `PN`:

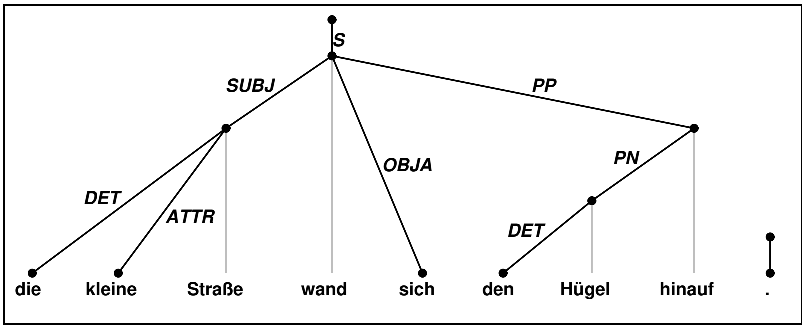


Die Vergleichsworte “wie” und “als” gelten nicht als Präpositionen; deshalb werden ihre Komplemente nicht mit `PN` bezeichnet, sondern mit `CJ`.

Einige Präpositionen und Postpositionen werden gelegentlich ohne Worttrennung an ein Vollverb angehängt. Es handelt sich hierbei um die Adpositionen ‘herunter’, ‘herauf’, ‘herab’, ‘hinunter’, ‘hinauf’, hinab’ und ‘entlang’. Obwohl sie sich also wie abtrennbare Präfixe verhalten, sind es doch immer noch Adpositionen, was daran erkenntlich ist, daß sie noch immer eine Valenz für eine weitere NP eröffnen. Bei diesen seltenen Fällen (und nur in der Nebensatzstellung) kann daher auch ein Vollverb durch `PN` modifiziert werden:

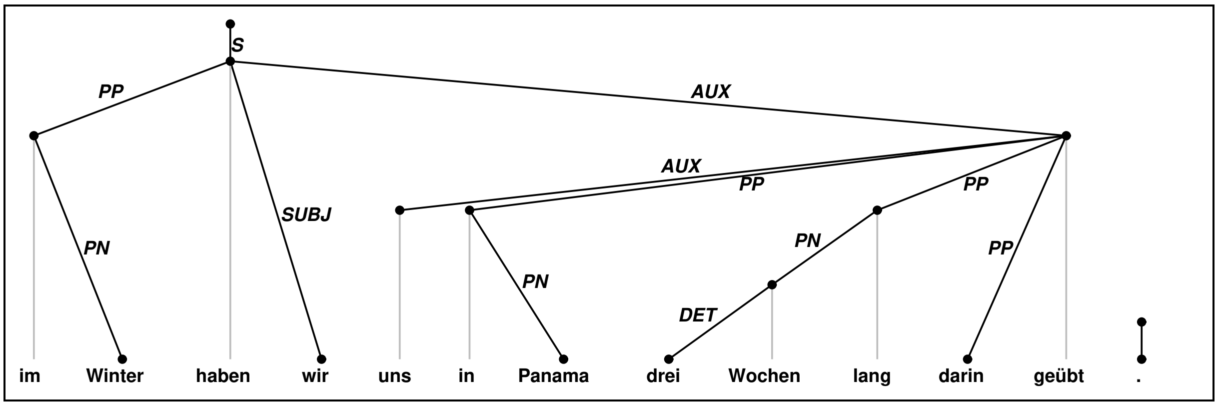


In der Hauptsatzstellung ist die Adposition als Adposition anzusehen und bildet daher eine normale PP:



1.2.26 Das Label PP

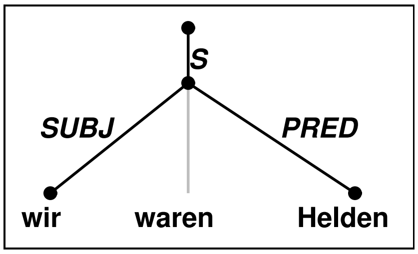
Mit diesem Label werden Präpositionalphrasen untergeordnet. Außer normalen Präpositionen (`APPR`) können verschmolzene Präpositionen (`APPRART`), Postpositionen (`APPO`) und Pronominaladverben (`PROAV`) eine Präpositionalphrase einleiten.

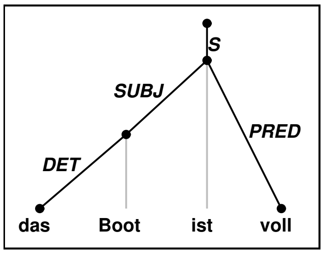


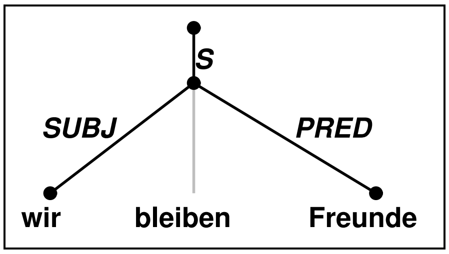
Präpositionalphrasen modifizieren vorwiegend Verben, weniger oft Nominalphrasen und Adjektive. Der Bezug ist oft nicht eindeutig zu klären.

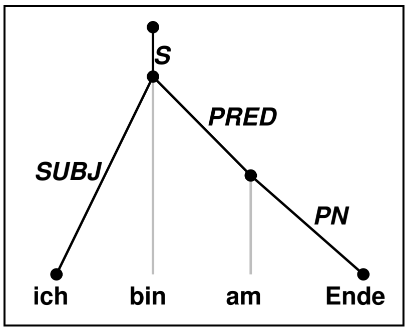
1.2.27 Das Label PRED

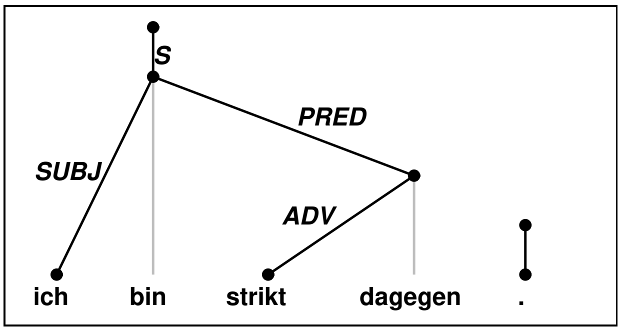
Dies Label bezeichnet prädikative Ergänzungen, typischerweise am Verb 'sein' oder ähnlichen Verben wie 'werden', 'wirken' oder 'scheinen'. Meistens ist das Prädikat ein prädikatives Adjektiv oder eine NP, es kann aber, insbesondere am Verb 'sein', auch Adverb oder Präposition sein:

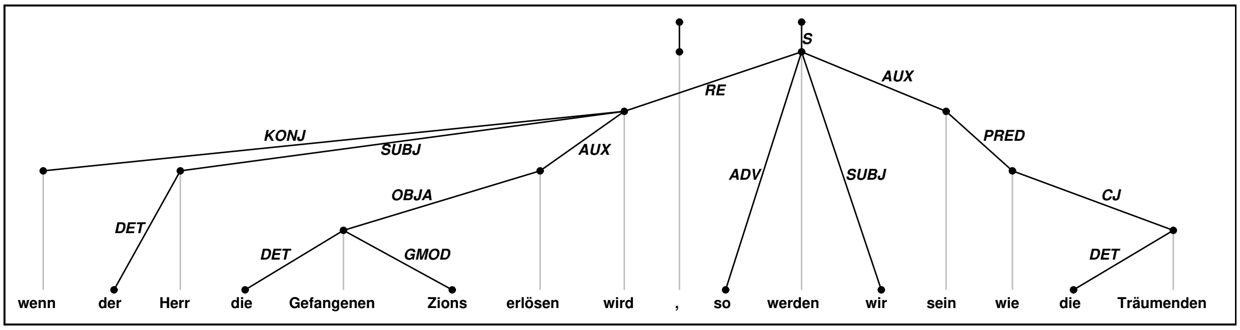






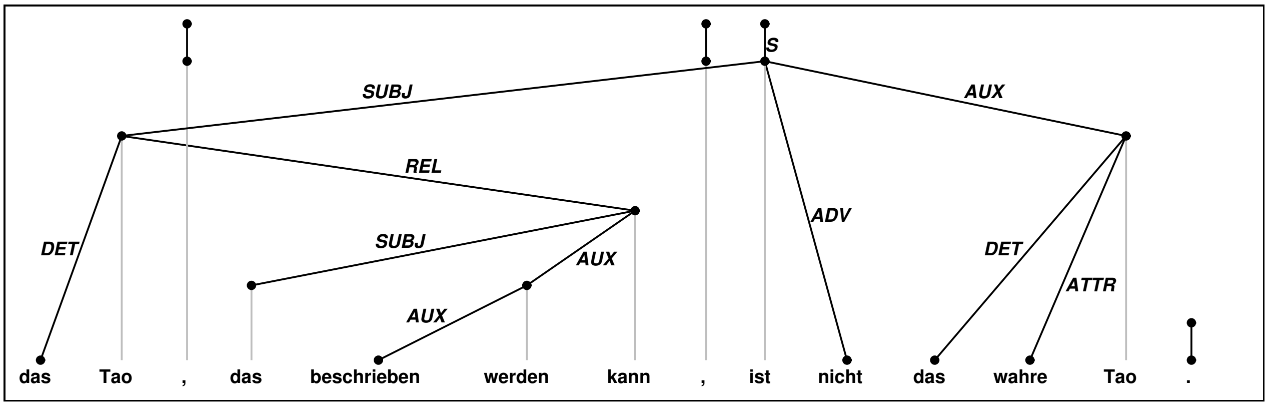


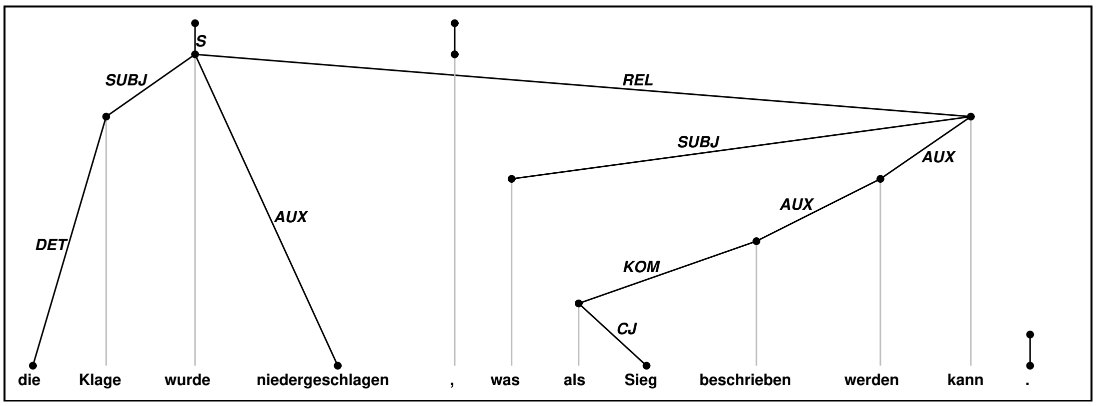




1.2.28 Das Label REL

Dieses Label bezeichnet die Beziehung zwischen dem Verb eines Relativsatzes und dem übergeordneten Wort. Das untergeordnete Wort ist immer ein finites Verb. Das übergeordnete Wort kann dagegen ein Verb oder ein Hauptwort sein.

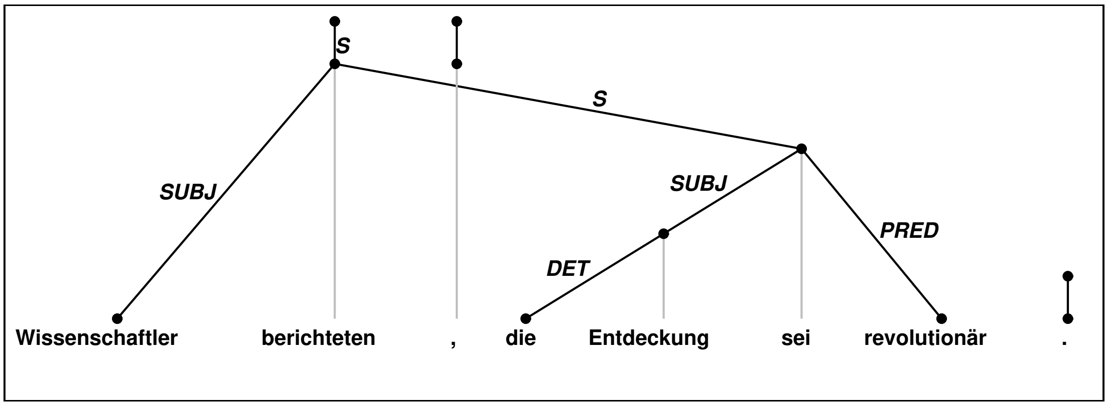




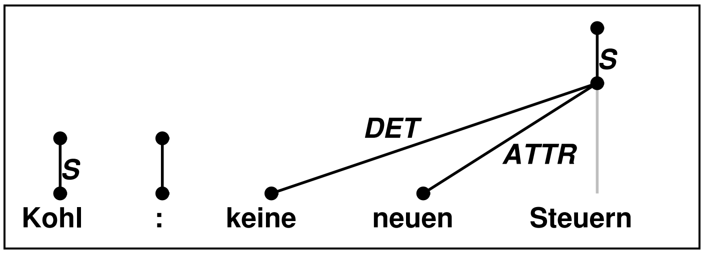
Ein Relativsatz ist immer ein Nebensatz, der mit einem Relativpronomen ("der") oder einem Fragewort ("womit") eingeleitet wird.

1.2.29 Das Label S

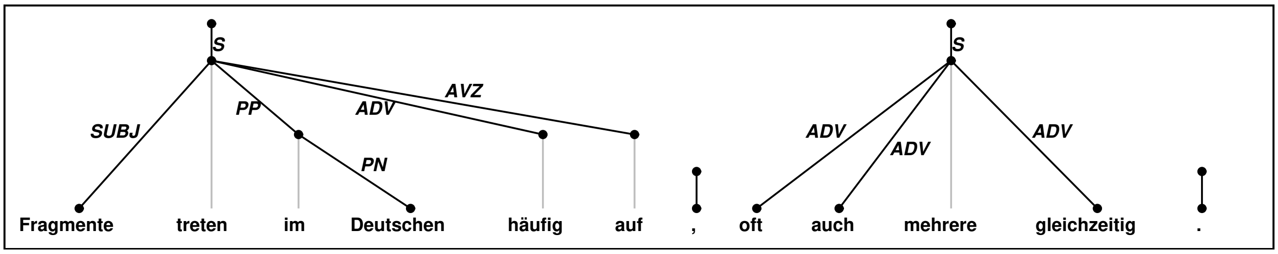
Das Label S bezeichnet das Wurzelwort eines vollständigen Satzes. Es macht dabei keinen Unterschied, ob es sich um einen Matrixsatz oder um einen untergeordneten Hauptsatz handelt.



S wird auch für alle anderen nicht untergeordneten Worte[1] verwendet, also für das dominierende Wort eines Fragmentes. Ein Fragment ist eine Konstruktion, die nicht in eine Satzstruktur eingeordnet werden kann, ohne harte Bedingungen zu verletzen. Typischerweise tauchen Fragmente in Überschriften auf:

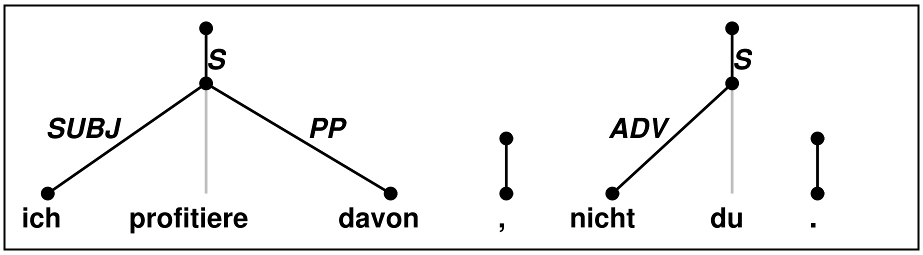


Oftmals werden Fragmente mit Komma abgetrennt und durch Adverben markiert:

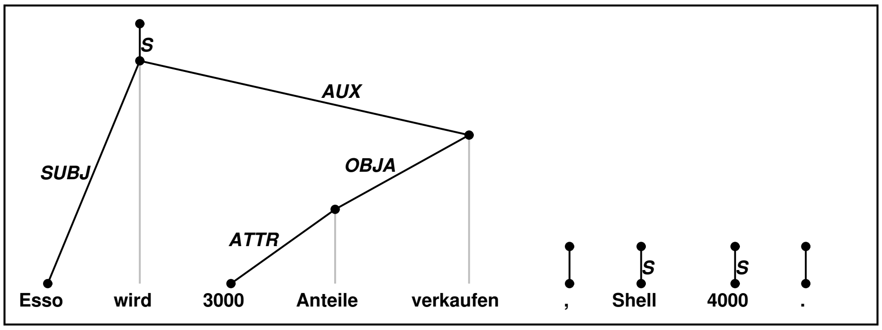


Eine weitere Quelle von Fragmenten sind elliptische Koordinationen. Oft wird etwa vor "nicht" das "und" fortgelassen:

[1] Satzzeichen gelten nicht als Worte; sie tragen daher auch kein Label. (Genaugenommen tragen sie den leeren String als Label.)



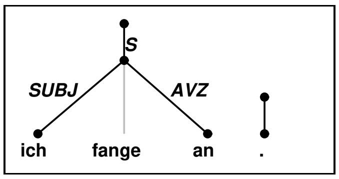
In anderen Fällen werden mehrere gleichlautende Worte aus einer Koordination weggelassen, weil sie sonst eine Wiederholung zur Folge hätten:

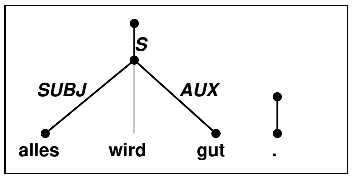


Solche Konstruktionen sind in einer rein deklarativen Dependenzgrammatik schlecht zu modellieren. Meistens müssen die verbleibenden Wörter als Fragmente angesehen werden.

1.2.30 Das Label SUBJ

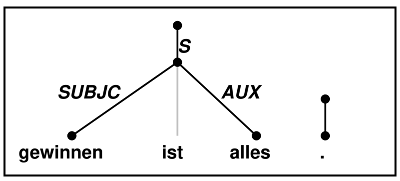
Als `SUBJ` wird das Subjekt eines finiten Verbs bezeichnet. Das ist immer irgendeine Art von NP:

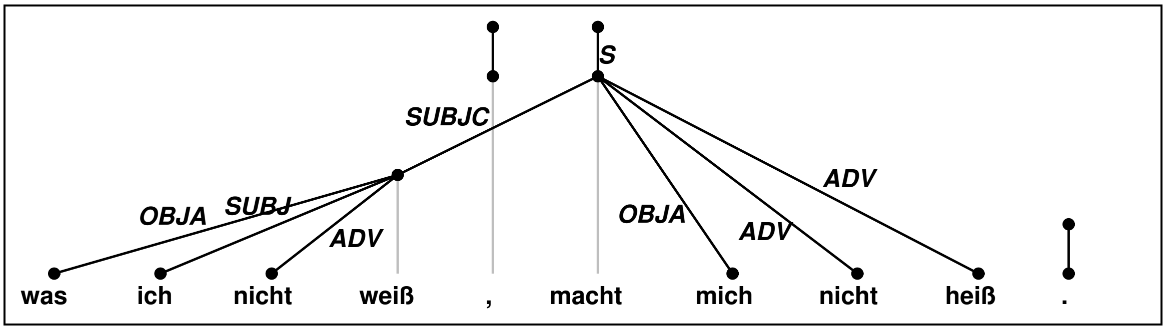


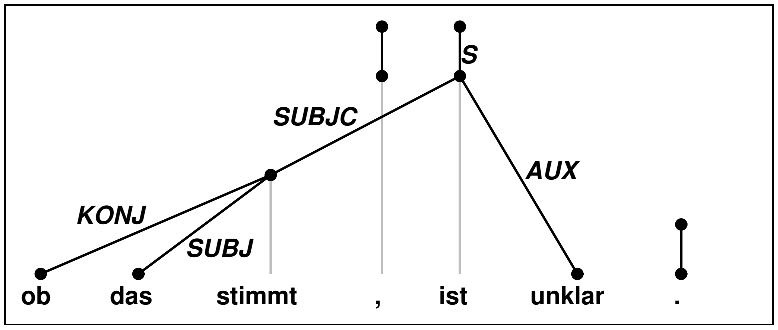


1.2.31 Das Label SUBJC

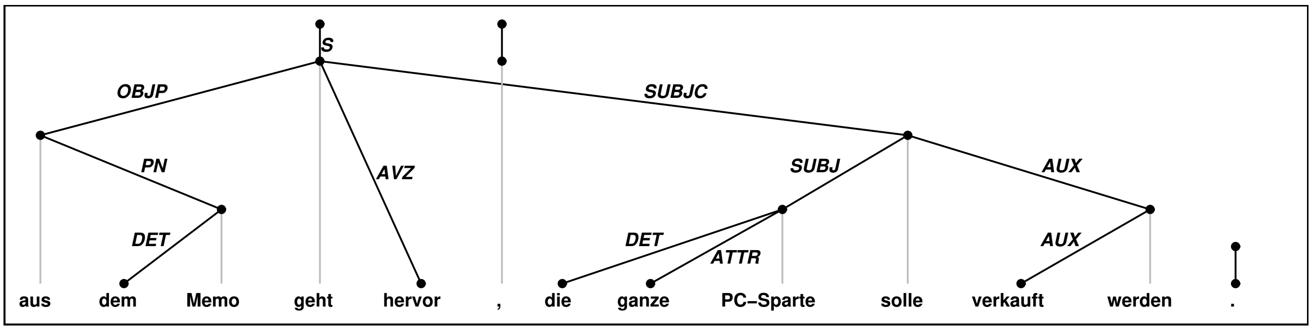
Es gibt jedoch auch ganze Subjektsätze, d.h. infinitive oder finite Verben, die die Subjektrolle ausfüllen. Diese werden als `SUBJC` bezeichnet.







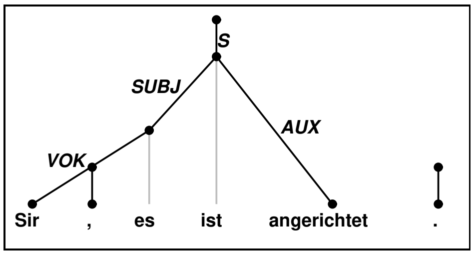
Ein finites Verb kann gewöhnlich nur dann die Subjektrolle ausfüllen, wenn es mit "dass", "ob" oder einem Fragewort verwendet wird. In seltenen Fällen kann ein ganz normaler (konjunktionsloser) Hauptsatz Subjektsatz sein. Das ist jedoch nur mit ganz bestimmten Verben möglich:

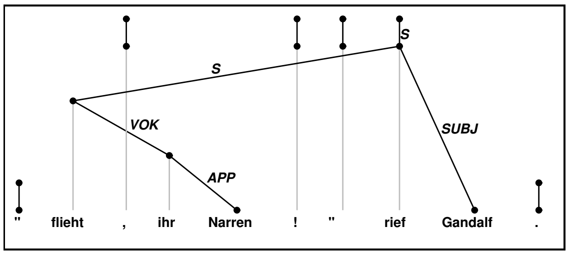


Alle diese Verben tragen das Feature `nimmt_Subjektsatz`.

1.2.32 Das Label VOK

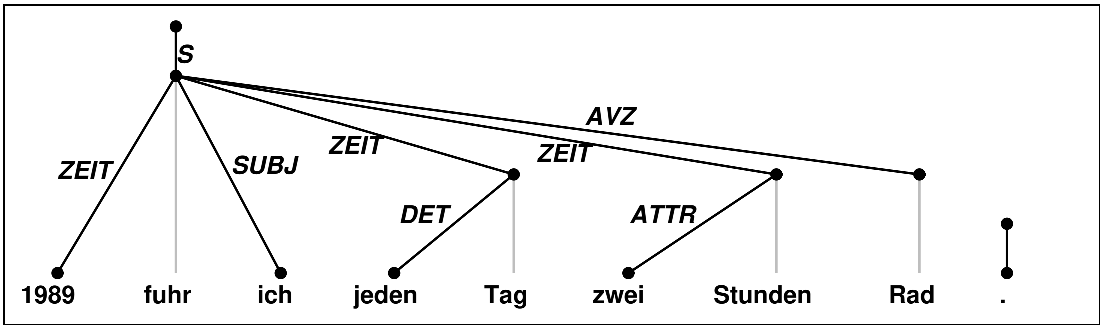
Das Label `VOK` verbindet Anreden mit dem Satz, zu dem sie gehören. Es wird jeweils am nächstliegenden Wort untergeordnet, egal welcher Kategorie dieses angehört.





1.2.33 Das Label ZEIT

Dieses Label bezeichnet konjunktionslose Zeitangaben. Solche Formulierungen stehen immer im Akkusativ und sind auf sehr wenige Nomen und Zahlen beschränkt.

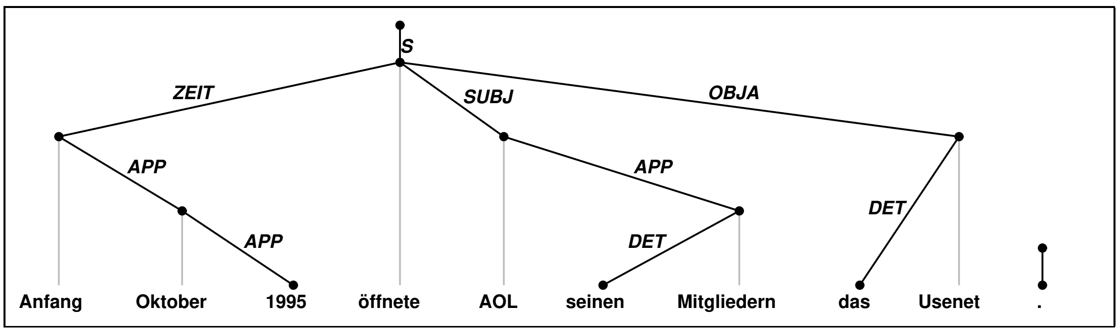


Nur folgende Worte können als Zeitangabe verwendet werden:

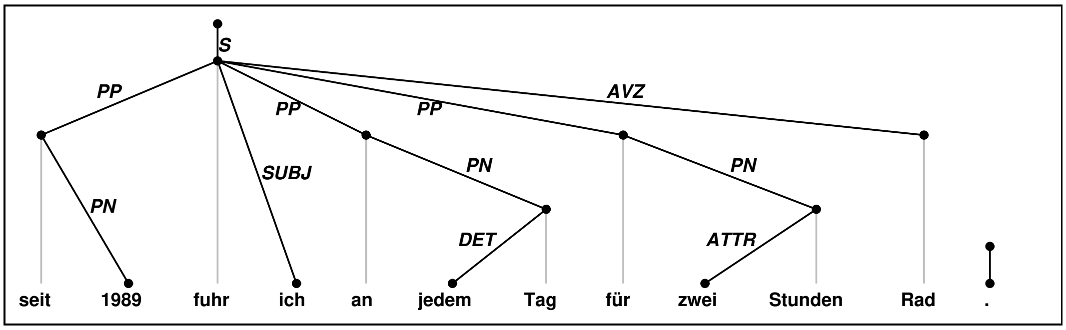
- Kardinalzahlen, wenn sie als Jahreszahlen erwendet werden
- "Anfang", "Mitte" und "Ende" in Verbindung mit anderen Zeitangaben
- Zeiteinheiten ("Tag", "Woche" etc.) und benannte Zeiten ("Montag", "Januar" etc.)

All diese Worte zeichnen sich dadurch aus, daß sie bestimmten semantischen Klassen angehören (feature `sort`).

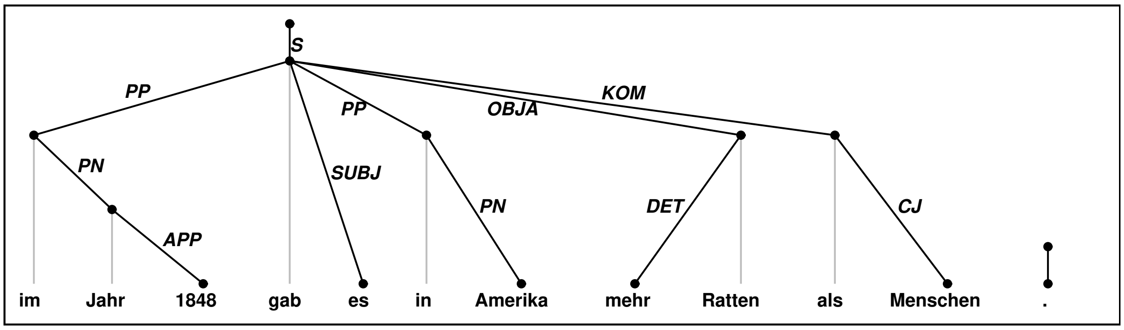
Wenn komplexe NP als Zeitangaben auftreten, ist nur die oberste Kante `ZEIT` und alle anderen normal `APP`.



Wenn eine solche Zeitangabe mit einer Präposition verwendet wird, ist sie nicht als `ZEIT` zu kennzeichnen, sondern als `PN`:



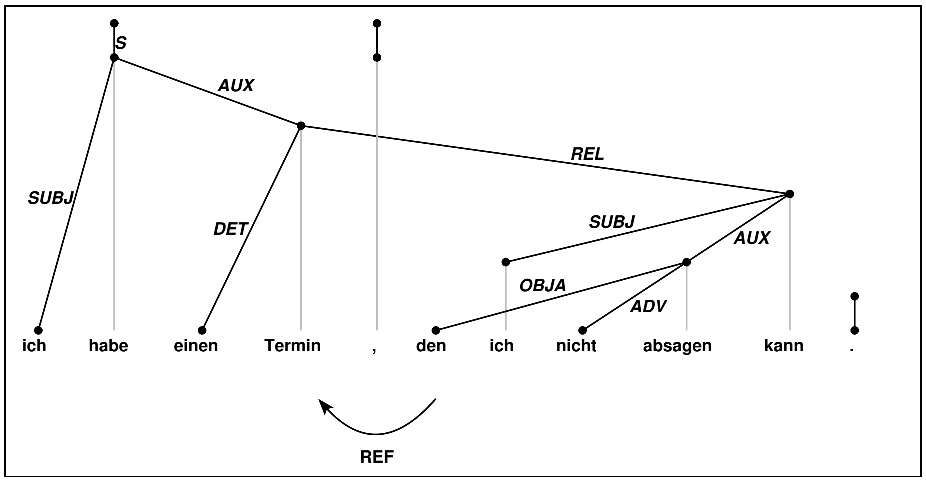
Jahreszahlen sind keine `ZEIT`-Modifikationen, wenn sie Hauptworte modifizieren, sondern normale Attribute:



1.3 Die Referenz-Ebene

Diese Analyseebene stellt Referenzbeziehungen zwischen Pronomen und ihren Antezedenten dar. Derzeit werden nur Relativpronomen berücksichtigt (obwohl prinzipiell auch Personalpronomen oder auch ganze Nominalphrasen auf andere Konstituenten referieren können).

Die einzige mögliche Kante auf dieser Ebene ist die vom Relativpronomen zu seinem Antezedent. Meistens ist dies das Nomen, das vom Relativsatz modifiziert wird:



1.4 Constraints

Obwohl CDG ein multiplikatives Bewertungsmaß verwendet, so daß mehrfache Fehler stärker ins Gewicht fallen können als ein einzelner starker Fehler, läßt sich doch auch das Verhalten von maximierenden Fehlermaßen annähern: wählt man scores in verschiedenen Größenordnungen, dann sind auch mehrere schwache Fehler immer noch besser als ein viel stärkerer. Diese Technik wird ausgiebig angewandt. Folgende Bereiche von scores treten auf:

1.0	nur debug-Constraints
0.9999 bis 0.99	Präferenz-Constraints
0.99 bis 0.5	heuristische Regeln, die öfter verletzt werden
0.5 bis 0.1	minder schwere Fehler
0.1 bis 0.001	schwere Fehler
0.0	Wohlgeformtheitsregeln der Grammatik selbst

Jedes Constraint ist in eine Gruppe verwandter Constraints eingeordnet, die ähnliche Phänomene modellieren. Folgende Gruppen gibt es:

- POS: Information des part-of-speech taggers
- agree: Kongruenz innerhalb von Phrasen
- category: Bedingungen an die Wortkategorie
- chunker: Information des chunk parsers
- debug: Constraints, die dem Autor zu Diagnosezwecken dienen.
- distance: Bevorzugung kurzer Anbindungen
- exist: Forderung nach der Existenz bestimmter Unterordnungen
- init: Strukturbedingungen, die niemals verletzt werden dürfen
- lastresort: willkürliche Disambiguierung in Ermangelung besserer Information
- lexical: besonderes Verhalten einzelner Worte

- order: Anordnung von verwandten Knoten
- proj: Projektivität von Syntaxstrukturen
- root: Wurzelanbindung
- sort: sortale Restriktionen
- vorfeld: Besetzung des Vorfelds

Viele Constraints fallen in mehrere dieser Klassen, aber CDG erlaubt keine Mehrfacheinordnung. Jedes Constraint ist deshalb der Klasse zugeteilt, der es am besten entspricht.

1.4.1 Die Constraintklasse `POS`

Diese Constraints nutzen Information des part-of-speech taggers, wenn vorhanden. Sie wird anhand der CDG-Funktion `pts()` gelesen.

Beispiel: der Score des Taggers wird direkt als Constraint-Score verwendet.

```
{X:SYN} : tagger : POS : [ pts(X@id) ] :
  pts(X@id) = 1.0
```

Die einzige sinnvolle Verwendung dieses Wertes ist die Forderung, daß er hoch sein soll. Statt den Wert selbst als score zu verwenden, kann man ihn auch in verschiedener Weise normalisieren.

1.4.2 Die Constraintklasse `agree`

Diese Constraints fordern die Übereinstimmung (Kongruenz) oder Existenz (Rektion) von Features innerhalb einer Phrase.

Beispiel: Determiner und Nomen kongruieren im Kasus.

```
{X!SYN} : DET_Kasus : agree : 0.1 :
  X.label = DET
  ->
  exists(X@case) & compatible(Kasus,X@case,X^case);
```

Das Subjekt steht im Nominativ.

```
{X!SYN} : Subjekt_Kasus : agree : 0.1 :
  X.label = SUBJ &
  exists(X@case)
  ->
  compatible(Kasus,X@case,nom);
```

Da nicht alle Lexikoneinträge voll spezifiziert sind, muß gewöhnlich `exists()` und `compatible()` verwendet werden.

Verletzungen der Kongruenz sind verhältnismäßig schwere Fehler.

1.4.3 Die Constraintklasse `category`

Diese Constraints fordern bestimmte Wortkategorien (feature `cat`). Beispiel: Das Subjekt ist eine NP (oder ein Subjektsatz).

```
{X!SYN} : Subjekt_Kategorie : category : 0.0 :
  X.label = SUBJ
  ->
  isa(X@,Nominal) | isa(X@,not_PP) | X@cat = ADJA;
```

Sie verwenden oft das Makro `isa()`.

1.4.4 Die Constraintklasse `chunker`

Die Constraints verwerten Information des chunk parsers. Sinnvollerweise verlangt man, daß Phrasenköpfe außerhalb und andere Worte innerhalb des Chunks angebunden werden:

```
{X!SYN} : chunk_2 : chunker : 0.0 :
 chunk_head(X@id)
 ->
 X^to > X@chunk_end |  X^from < X@chunk_start;
```

1.4.5 Die Constraintklasse `debug`

Diese Constraints haben immer den Wert 1.0, d.h. sie beeinflussen die Analyse nicht. Sie dienen lediglich dazu, anzuschlagen, wenn eine erlaubte, aber außergewöhnliche Situation eintritt, z.B. eine Ausnahmeregelung in einem anderen Constraint, die sonst unbemerkt bliebe. Beispielsweise wird jede Projektivitätsausnahme durch ein eigenes debug-Constraint scharf bewacht.

Die Aufgabe von Debug-Constraints läßt sich besser durch `search-annotations.pl` erfüllen.

1.4.6 Die Constraintklasse `dist`

Diese Constraints bestrafen lange Anbindungen leicht, so daß im Zweifelsfall die kürzere gewählt wird. Sie verwenden zumeist das Makro gradient().

Beispiel: von zwei PP-Anbindungen wähle die kürzere.

```
{X!SYN} : PP_Distanz : distance : gradient(100) :
  X.label = PN
  ->
  abs( distance( X@id, X^id ) ) <= 1;
```

1.4.7 Die Constraintklasse `exist`

Constraint, die die Existenz von Komplementen fordern. Gewöhnlich verwenden sie das Prädikat `has()`.

Beispiel: Wenn irgendein Wort ein drittes Komplement verlangt, bestrafe dessen Fehlen.

```
{X:SYN} : Existenz_3 : exist : [ X@args:3:obl ] :
  exists(X@args:3:obl)
  ->
  X@cat=VVPP |
  Partizipialadjektiv(X@) |
  X.label=REL |
  has(X@id, X@args:3:lab);
```

In einigen Fällen ist die Existenz eines normalerweise notwendigen Komplementes ausdrücklich verboten. Beispiel: ein Partizip, das von 'sein' oder 'werden' abhängt, gilt als Passivkonstruktion und darf keine Akkusativobjekt tragen.

```
{X!SYN\Y!SYN} : VVPP_obl2_exist_penalty : exist : 0.0 :
 X.label = OBJA &
 isa(Y@,Partizip) &
 (Y^infinitive = werden | Y^infinitive = sein)
 ->
 false;
```

1.4.8 Die Constraintklasse `help`

Diese Constraints sind Hilfsformeln, die von anderen Constraints über `has()` aufgerufen werden. Sie selbst sind gewöhnlich nicht aktiv, weil sie keine allgemeingültigen Regeln beschreiben.

```
{X:SYN} : find_daß        : help : 1.0 : exists(X@Objektsatz_Konjunktion);
```

1.4.9 Die Constraintklasse `init`

Diese Constraints stellen Strukturbedingungen, die nicht sinnvoll verletzt werden können. Sie haben stets die Bewertung 0.0.

Beispiel: Es kann es nicht zwei Subjekte zu einem Verb geben (höchstens eine Koordination aus zwei NP).

```
{X!SYN/\Y!SYN} : SUBJ_uniq : init : 0.0 :
 X.label = SUBJ
 ->
 Y.label != SUBJ;
```

1.4.10 Die Constraintklasse `lastresort`

Diese Constraints haben Bewertungen nahe 1. Sie dienen zur Disambiguierung, wenn keine andere Information vorliegt.

Beispiel: Wenn nichts anderes bekannt ist, nimm den Singular an.

```
{X:SYN} : pl : lastresort : 0.999 :
  exists(X@number) -> X@number!=pl;
```

1.4.11 Die Constraintklasse `lexical`

Diese Constraints beschreiben das Verhalten einzelner Lexeme genauer. Information dieser Art ist gewöhnlich lexikalisiert; z.B. gibt jedes Verb selbstbestimmt seinen Valenzrahmen vor, und die Grammatik regelt Verbkomplemente, indem sie auf diesen Rahmen zugreift statt auf die Wortform. In manchen Fällen ist das Verhalten aber so abweichend, daß eigene Constraints sinnvoller sind.

Beispiel: gewöhnlich sind alle Koordinationen rechtsverzweigend, daher muß eine `CJ`-Kante immer nach links angebunden werden. Ausnahmen hierfür gibt es für Trunkationen (erkennbar an der Kategorie `TRUNC`) und für einige bestimmte Konjunktionen, die oft weiter nach rechts verstellt werden, z.B. 'aber' (erkennbar am Feature `Verstellung`).

Die Konjunktion 'doch' hat ein noch komplizierteres Verhalten: gewöhnlich steht sie zwischen den Konjunkten, aber bei finiten Verben darf sie verstellt werden ('...ist er doch der beliebteste Alleinunterhalter Deutschlands'). Daher wird eigens für sie ein weiteres Constraint aufgestellt.

```
{X/SYN} : doch_C1 : lexical : 0.0 :
 X^cat = KON &
 X^word = doch &
 X.label = CJ
 ->
 isa(X@,finit);
```

1.4.12 Die Constraintklasse `order`

Diese Constraints schränken die Reihenfolge von verwandten Knoten ein.

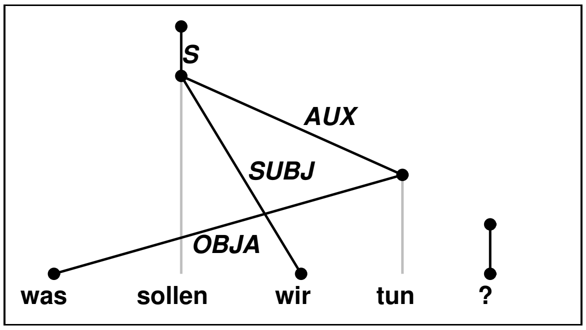
Beispiel: Der Determiner steht vor dem Bezugswort.

```
{X\SYN} : DET_order : order : 0.01 :
  X.label != DET;
```

1.4.13 Die Constraintklasse `proj`

Diese Constraints fordern die Projektivität von Syntaxstrukturen. Unglücklicherweise bestehen im Deutschen eine Reihe von Ausnahmen von diesem Prinzip.

Beispiel: Sinnvollerweise ordnet man das Subjekt dem finiten Verb unter, weil zwischen diesen beiden die Rektionsbeziehung besteht. Ebenso wird das Objekt dem Vollverb untergeordnet, weil dort die Valenzinformation verfügbar ist. Wenn aber das Objekt topikalisiert ist, kann es die Projektivität verletzen:



Es gibt eine Vielzahl solcher Situationen. Die meisten hängen entweder mit komplexen Verbgruppen oder mit bestimmten Worten zusammen, die die Satzstruktur beeinflussen ('wir ... alle'). Alle diese Möglichkeiten erfordern Ausnahmen in der sonst unverletzlichen Bedingung. Die typischen Bedingungen zur Herstellung von Projektivität (keine interne Anbindung, keine überlappende Anbindung) werden deshalb weiter je nach der Richtung der betroffenen Kanten unterschieden und als sechs verschiedene Constraints formuliert, von denen jedes mehrere Ausnahmen hat.

Die einzelnen Ausnahmen sind jeweils mit Beispielen versehen, vergl. die Constraintdefinitionen selbst.

Als Warnung vor übergenerierenden Ausnahmen gibt es außerden `debug`-Constraints, die bei jeder nichtprojektiven Struktur anschlagen.

1.4.14 Die Constraintklasse `punc`

Diese Constraints stellen Regeln auf, die syntaktische Beziehungen in Abhängigkeit von Satzzeichen regeln. Diese Regel meldet etwa ein fehlendes Komma für den Nebensatz:

```
{X!SYN} : 'Komma für Nebensatz fehlt' : punc : 0.1 :
  X.label = NEB
  ->
  between(X@id,X^id,"(),-");
```

Wenn Text ohne Satzzeichen geparst werden soll, muß diese Klasse deaktiviert werden.

Nur solche Constraints, die ohne Satzzeichen nicht richtig funktionieren, gehören dieser Klasse an. Regeln, die vom *Vorhandensein* eines Satzzeichens auf bestimmte Strukturen schließen, sind bei Input ohne Satzzeichen natürlich immer trivial erfüllt. Sie können daher auch in anderen Klassen stehen, etwa `order`.

1.4.15 Die Constraintklasse `root`

Diese Constraints regeln die Nicht-Anbindung von Worten. Beispiel: finite Verben sind legitimerweise Satzwurzeln, aber wenn es zwei finite Verben gibt, sollte das eine ein Nebensatz zum anderen sein.

```
{X|SYN, Y|SYN} : 'mehrere Hauptsätze' : root : 0.6 :
  isa(X@,finit) &
  isa(Y@,finit)
  ->
  false;
```

1.4.16 Die Constraintklasse `sort`

Diese Constraints stellen sortale Restriktionen auf.

Beispiel: ein Nebensatz mit Relativadverb kann nur durch ganz bestimmte Adverben ausgelöst werden ('da. . . wo').

```
{X:SUB\/Y!SYN} : v_mod_a : sort : 0.0 :
 Y.label = NEB &
 isa(Y@,finit) &
 isa(Y^,ADV)
 ->
 X! & (
    ( Y^word = so & (X^word = daß | X^word = dass) ) |
    ( Y^word = dann & X^word = wenn ) |
    ( Y^word = dort & X^word = wo ) |
    ( Y^word = da & X^word = wo )
 );
```

1.4.17 Die Constraintklasse `uniq`

Diese Constraints drücken Eindeutigkeit aus, zumeist unter den Labeln von nebengeordneten Worten.

```
{X!SYN/\Y!SYN} : 'doppeltes GRAD' : uniq : 0.0 :
  X.label = GRAD -> Y.label != GRAD;
```

1.4.18 Die Constraintklasse `zone`

Diese Constraints verbieten Anbindungen, die durch Kommas markierte Zonen in ungültiger Weise überschreiten. (Sie gehören nicht der Klasse `punc` an, weil sie auch für Input ohne Satzzeichen gelten.)

```
{X/SYN/\Y/SYN} : 'Objektsatz-Zone' : zone : 0.0 :
  X.label = OBJC &
  Y@from > X@from &
  between(X^id,X@id,",")
  ->
  between(X@id,Y@id,",:-");
```

1.4.19 Makros für Constraintformeln

In den Constraints häufig wiederkehrende Konstruktionen sind als Makros definiert. Folgende Definitionen gibt es derzeit:

edge(<Dependenzkante>,<String>): true, wenn das Label der Kante in der Hierarchie `Label` dem angegebenen Zweig angehört.

gradient(<Ganzzahl>): erzeugt einen Penalty-Term für ein Distanz-Constraint mit Gradient `N`.

isa(<Lexemknoten>,<Kategorie>): true, wenn der Lexemknoten der angegebenen Kategorie angehört.

left(<Lexemknoten>,<Lexemknoten>): true, wenn der erste Lexemknoten weiter links liegt als der zweite.

right(<Lexemknoten>,<Lexemknoten>): true, wenn der erste Lexemknoten weiter rechts liegt als der zweite.

contains(<Dependenzkante>,<Dependenzkante>): true, wenn die erste Kante die zweite vollständig enthält oder beide sich nicht berühren.

inside(<Lexemknoten>,<Dependenzkante>): true, wenn der Lexemknoten zwischen den Endpunkten der Dependenzkante liegt.

crosses(<Dependenzkante>,<Lexemknoten>): true, wenn der Lexemknoten zwischen den Endpunkten der Dependenzkante liegt.

enters(<Dependenzkante>,<Dependenzkante>): true, wenn der Modifikand der ersten Dependenzkante innerhalb der zweiten Kante liegt und der Modifikator nicht.

leaves(<Dependenzkante>,<Dependenzkante>): true, wenn der Modifikator der ersten Dependenzkante innerhalb der zweiten Kante liegt und der Modifikand nicht.

Deverbalkante(<Dependenzkante>): true, wenn die Kante ein Verb mit einem Nomen verbindet, das das Feature `nimmt_Objektsatz` besitzt.

Objektvertreter(<Lexemknoten>): true, wenn der Lexemknoten eines der Worte repräsentiert, die als Vergleichsworte Objekte ersetzen können, nämlich wie/PWAV oder als/KOUS.

Satztrenner(<Lexemknoten>): true, wenn der Lexemknoten eines der Satzzeichen repräsentiert, die Hauptsätze voneinander trennen (- : ;).

Partizipialadjektiv(<Lexemknoten>): true, wenn der Lexemknoten ein Adjektiv mit dem Feature partizipial1 oder partizipial2 repräsentiert.

1.5 Lexikon

Die Information über mögliche und notwendige Bindungen ist zum großen Teil lexikalisiert, d.h. in den einzelnen Lexikoneinträgen festgehalten. Hier werden die vorkommenden Features mit ihrer Bedeutung erläutert.

Die Werte von lexikalischen Features können im CDG-Formalismus die Datentypen String, Zahl oder Liste haben. Zusätzlich werden einige Features verwendet, deren Bedeutung nur darin liegt, ob daß sie überhaupt definiert sind, die also wie Boolean-Werte funktionieren.

Solche Features werden von der Grammatik nur durch `exists()` zugegriffen. Der Wert selbst, wenn vorhanden, lautet immer `yes`.

Folgende Features werden im Lexikon verwendet:

`Objektsatz_Konjunktion` (Boolean): Dieses Feature tragen nur die Konjunktionen 'daß' und 'ob'. Nur diese Konjunktionen können normalerweise einen Nebensatz einleiten, der als Objekt oder Subjekt verwendet wird (statt als normaler Nebensatz).

`Satzkonjunktion` (Boolean): Dieses Feature markiert Konjunktionen, die nur ganze Sätze reihen können, nicht einzelne Phrasen. So darf etwa in einer beigeordneten VP das Subjekt fehlen, wenn sie mit 'oder' gebildet ist, aber nicht, wenn sie mit 'denn' gebildet ist.

`Stoffnomen` (Boolean): Dies Feature markiert Nomen, die Stoffe o.ä. bezeichnen und daher im Singular ohne Artikel auftreten, z.B. 'Gold', 'Freiheit' oder 'Musik'. (Das schließt nicht aus, daß sie dennoch einen Plural bilden können: 'das Metall'/'viel Metall'/'drei Metalle').

`avz` (required|allowed|forbidden): Dieses Feature tragen finite Verben, um anzugeben, ob sie mit einem trennbaren Verbzusatz auftreten können. Beispielsweise hätte 'gibt' den Wert 'allowed', denn sowohl 'gibt' als auch 'gibt...auf' ist möglich. 'interessiert' hat den Wert 'forbidden', weil es (gemäß unserem Lexikon) keine Zusammensetzungen bildet. Den Wert 'required' hätte ein Verb, das nur als trennbares Verb auftritt.

`baseform` (String): Dies Feature gibt die Grundform eines nominalen Wortes an, also den Nominativ Singular.

`case` (nom|gen|dat|acc): Der syntaktisch ausgeprägte Kasus an einer nominalen Form. Kann unterspezifiziert sein.

`cat` (wie in Hierarchie 'Kategorien'): Die syntaktische Kategorie des Wortes. Dies ist immer eine der Kategorien des STTS.

`cat2` (wie `cat`) Einige Worte haben im Lauf der Zeit ihre syntaktische Kategorie geändert, verhalten sich aber immer noch wie Angehörige beider Klassen. In diesem Fall wird als `cat2` diejenige Kategorie angegeben, die das Wort ursprünglich hatte. Zum Beispiel ist 'anno' im Lateinischen Nomen, im Deutschen aber Adverb. Es trägt also die Attribute `cat=ADV` und `cat2=NN`.

`clitic` (Boolean): Nur gesetzt bei finitem Verb, das ein klitisches 'es' trägt und daher kein Subjekt braucht (etwa 'gibts'). Normalerweise sollten solche Formen als zwei Worte tokenisiert werden: 'gibt'+ ''s'.

`definite` (yes|no): Dieses Feature unerscheidet bestimmte von unbestimmten Artikeln, was Auswirkungen auf die erlaubten Adjektive hat. Auch einige Pronomen tragen es.

`degree` (positive|comparative|superlative): Der Steigerungsgrad von Adjektiven und einigen Pronomen.

`flexion` (weak|mixed|strong): Die Flektionsart von Adjektiven. Kann unterspezifiziert sein.

`gender` (fem|masc|neut): Das Genus von nominalen Formen. Kann unterspezifiziert sein.

`graded` (Boolean): Dieses Feature tragen Adverben, die jede Art von Komparativ modifizieren können ('erheblich', 'viel' etc.).

`infinitive` (String): Dies Feature bezeichnet den Infinitiv, von der das Verb abgeleitet ist. Dieses Feature unterscheidet z.B. die Formen von 'gehen' von den Formen von 'ausgehen'.

`subcat` (String): Dieses Feature unterscheidet verschiedene Typen innerhalb einer syntaktischen Kategorie. Beispielsweise sind alle Worte vom Typ `NE` genauer klassifiziert (Vorname, Nachname, Firmenname etc.), weil der Unterschied manchmal syntaktisch bedeutsam ist.

`leftpenalty` (Zahl): Die Strafe, die angewendet werden soll, wenn dieses Wort als Modifikator links steht.

`loctype` (lative|essive): Unterscheidet Richtungs- von Orts-Ausdrücken.

`linker_Teil` (Boolean): markiert die einleitenden Teile von mehrteiligen Konjunktionen ('weder', 'je').

`measurement` (Boolean): markiert Adjektive, die Gradangaben auch im Positiv zulassen (300 Mhz—GRAD—>schnell).

Ebenfalls markiert werden Adverben, die durch Zusammensetzung aus Nomen entstanden sind, die Gradangaben zulassen, z.b. '1000-mal', denn diese erlauben andere Modifikatoren als normale Adverben:

rund 1000-mal
*rund jeweils

`modifies` (Liste): Spezifiziert, welche Kategorien von Worten dieses Wort modifizieren kann, und mit welcher Bevorzugung.

`mood` (indicative|subjunctive1|subjunctive2): Der Modus eines finiten Verbs.

`nimmt_Objektsatz` (Boolean): Dies Flag zeigt an, daß das Wort irgendeine Art von Verb als Modifikator oder Komplement nehmen kann (was normalerweise nicht möglich ist). Nomen, Verben und Pronominaladverben ('dazu') können das tun. Das Wort muß in seinem Argumentrahmen genauer bestimmen, ob finite oder infinite Verben erlaubt sind.

`nimmt_Subjektsatz` (Boolean): Dies Flag zeigt an, daß ein Verb einen normalen Hauptsatz als Subjekt nehmen kann. (Hauptsätze mit 'daß' sind immer als Subjekt erlaubt; dieses Flag markiert nur solche Verben, bei denen auch Sätze *ohne* 'daß' erlaubt sind (verlauten, hervorgehen aus).)

`number` (sg|pl): Der Numerus von Verb- oder Nominalformen.

`obja` (String): Dies Feature gibt an, daß nur ein ganz bestimmtes Nomen als Akkusativobjekt in Frage kommt. Meistens betrifft dies die *figura etymologica*, wo Verb und Nomen denselben Stamm besitzen:

den Heldentod sterben
*die Schlacht sterben

`partizipial1` (Boolean): Dies Flag markiert Partizipien des Aktiv (die syntaktisch gesehen Adjektive sind).

`partizipial2` (Boolean): Dies Flag markiert deklinierte Partizipien des Passiv (die syntaktisch ebenfalls Adjektive sind).

`pattern` (String): Dies Feature zeigt an, aus welcher Regel für unbekannte Worte ein Lexikoneintrag hervorging. Es ist nicht nur informativ, sondern wird von der Grammatik benutzt, um z.B. verschiedene Arten von `TRUNC` zu unterscheiden.

`perfect` (sein|haben|bot): Dies Feature gibt das Hilfsverb an, mit dem ein `VVPP` verwendet werden muß. Einige Verben können sowohl mit 'sein' als auch mit 'haben' verwendet werden; diese tragen den unterspezifizierten Wert 'bot'.

`person` (first|second|third): Die syntaktische Person eines Nomens oder Verbs.

`phon` (String): Dies Feature gibt den abgetrennten Wortbestandteil eines TRUNC an (z.B. 'aus' für 'aus-').

`pos` (SOV|SVO): Dies Feature markiert den Satztyp (Nebensatzstellung oder Hauptsatzstellung), der bei einigen Verben direkt ablesbar ist.

`predicative` (Boolean): Dies Feature markiert einige Verben, die wie 'sein' verwendet werden, obwohl sie Vollverben sind (bleiben, scheinen). Solche Verben erlauben Pseudopassive (transitive Verben ohne Objekt): 'Das Spiel schien verloren.'

`prep` (vgl. Hierarchie 'Kasus'): Dies Feature nennt bei allen Adpositionen und Abarten davon (`APPRART`, `PROAV`) die zugrundeliegende Adpositionsform. Beispielsweise tritt des Verb 'abhängen' gern mit der Präposition 'von' zusammen; die tatsächliche Wortform kann aber ebenso gut 'vom' oder 'hiervon' lauten. Dies Feature kennzeichnet alle diese Varianten als zusammengehörig.

`rechter_Teil` (Boolean): markiert die rechten Teile von mehrteiligen Konjunktionen ('noch, 'desto').

`rightpenalty` (Zahl): Die Strafe, die angewendet werden soll, wenn dieses Wort als Modifikator rechts steht.

`sibilant` (Boolean): Dieses Feature markiert Eigennamen, die in Zischlauten enden. (Das ist deshalb von Bedeutung, weil diese Namen manchmal den Nominativ als Genitiv verwenden.)

`sort` (vgl. Hierarchie 'Sorten'): Rudimentäre Information über die Bedeutung eines Wortes.

`stress` (stressed|unstressed|su|us|none): Dies Feature unterscheidet Varianten von Verben, die sich nur in der Betonung unterscheiden. Beispielsweise ist das 'übersetzen' mit betontem Präfix ein Synonym für 'Gewässer überqueren', während das mit unbetontem Präfix 'dolmetschen' bedeutet.

`Substantivierung` (Boolean): Dies Feature markiert Nomen, die durch Substantivierung aus Adjektiven entstanden sind. Dies ist notwendig, weil diese sich in mancher Hinsicht noch wie Adjektive verhalten (z.B. kein Determiner notwendig).

`tense` (past|present): Das Tempus eines finiten Verbs.

`valence` (String): Dies Feature gibt den Valenzrahmen von Verben (und verwandter Worte wie etwa Verbaladjektive) an.

Die Valenz eines Verbs ist im wesentlichen ein Kürzel für das Label, mit dem sein Komplement bezeichnet wird. Folgende Kürzel werden verwendet:

Kürzel	Kantenlabel
A	`OBJA`
B	`OBJA2`
C	`OBJC`
D	`OBJD`
E	`PRED`
G	`OBJG`
I	`OBJI`
K	`KOM`
P	`OBJP`
Q	`PN`
X	`AUX`

Optionale Komplemente werden durch angefügtes '?' bezeichnet, alternative Komplemente (z.B. `OBJA` oder `OBJC`) werden durch Hintereinanderschreibung der betreffenden Kürzel ausgedrückt. Bei bitransitiven Verben werden zwei Kürzel durch '+' verbunden.

`value` (Zahl): Dies Feature gibt den numerischen Wert von ausgeschriebenen Zahlen an.

1.6 Hierarchien

Die meisten Verbände enthalten genau solche Knoten, die als Wert eines bestimmten Features auftreten können. Darunter fallen die Verbände 'Flexion', 'Gender', 'Kasus', 'Number', 'Person' und 'Präpositionen'.

Der Verband 'Kategorien' enthält alle möglichen Werte des Features `cat`. Dies sind gerade die im Stuttgart-Tübinger Tagset definierten Tags. Darüberhinaus enthält der Verband diverse Zwischenknoten, die verschiedene gröbere Klassifikationen bezeichnen. So ist z.B. die Kategorie `VVINF` sowohl 'Vollverb' als auch 'Infinitiv', und `VMFIN` sowohl 'Modalverb' als auch 'finit'.

Der Verband 'Nebenordnungen' gruppiert jeweils die Kategorien, die durch Koordinationen einander beigeordnet werden können. So ist zum Beispiel 'jetzt und in Zukunft' grammatisch, 'jetzt und die Zukunft' nicht.

Der Verband 'Sorten' enthält derzeit eine Minimalklassifikation von Worten nach sehr allgemeinen Bedeutungsklassen und muß vollständig abgeändert werden, bevor er nützlich sein kann.

Der Verband 'Label' zählt alle Label der Syntax-Ebene auf und unterscheidet sie nach Modifikator und Komplement (vgl. unter 'Analyseebenen').

2 Annotationsrichtlinien

Im Prinzip ist die gewünschte Struktur für jede Äußerung durch die Constraints der Grammatik definiert. Da die Constraints aber Fehler enthalten können und außerdem weiterentwickelt werden, ist hier ausdrücklich beschrieben, wie in Zweifelsfällen annotiert werden soll.

2.1 Einteilung in Worte

Normalerweise werden Worte durch Leerzeichen oder Satzzeichen voneinander abgegrenzt, wobei jedes Satzzeichen als ein Wort gilt.

- "Was wollt ihr?" schrie er. (neun Worte)

Es gibt jedoch einige Ausnahmen. Mehrteilige Namen gelten als mehrere Worte; wird aber ein Adjektiv aus einem solchen Namen gebildet, so gilt es als ein Wort.

- Bad Nauheim (zwei Worte)
- der Bad Nauheimer Kulturverein (drei Worte)

Apostrophe, die Auslassungen oder Genitiv anzeigen, gehören zum betreffenden Wort.

- Leibniz' Philosophie des Unabwendbaren (vier Worte)
- Das war's. (vier Worte)

Sind Anführungszeichen teilweise in Zusammensetzungen verschmolzen, so gehören beide Anführungszeichen zum Wort dazu.

- kurze "Jingle"-Einspielungen (zwei Worte)

Das betreffende gilt für modische Klammerungen.

- eine/n Schüler/in (zwei Worte)
- man(n) amüsiert sich (drei Worte)
- der mühsame 2:1(1:1)-Erfolg (drei Worte)

Satzzeichen, die zu Eigennamen gehören, gehören zum betreffenden Wort.

- Wham!, die Gruppe mit dem Ausrufezeichen (sieben Worte)

Komposita mit Bindestrichen gelten als ein Wort. Wenn ein solches Wort am Zeilenende getrennt wurde, ist es wieder zusammenzusetzen.

- Bayern, Baden-Württemberg und Mecklenburg-Vorpommern (vier Worte)

Arabische Zahlen gelten als ein Wort, auch wenn sie mit Kommas oder Punkten unterteilt sind.

- 1,000,000 Millionen Dollar (drei Worte)

2.2 Behandlung von fehlerhaftem Input

Grundsätzlich ist für jede Eingabe eine möglichst regeltreue Analyse zu suchen. Wenn ein Satz strukturell mehrdeutig ist, ist diejenige zu wählen, die deutlich weniger Regeln verletzt.

- Dreiviertel der Unternehmen in der Internet-Industrie werden/VAFIN aufgekauft/VVPP oder/KON scheitern/VVFIN. (nicht VVINF!)

Enthält der Eingabesatz unabweisbar Fehler, die offensichtlich auf Nachlässigkeit zurückzuführen sind, so wird diejenige Struktur annotiert, die dem wahrscheinlich intendierten Satz am ähnlichsten ist. Beispielsweise soll ein falsch geschriebenes Verb als ein unbekanntes Verb angesehen werden, ein falsch geschriebenes Nomen als unbekanntes Nomen etc.

- Der Versuch, 512 Linux-Systeme zu einem Cluster zusammenzufasssen/OBJI, ist in der Nacht zum Sonntag an der Uni Paderborn erfolgreich beendet worden.

Wenn ein Wort offensichtlich fehlt oder widerholt ist, so sind die nicht einzuordnenden Bestandteile als Fragment anzusehen.

- Zum ersten Mal ist den USA/S eine Gefängnisstrafe gegen einen Kinobesucher verhängt worden, der das Leinwandgeschehen mit seiner Videokamera aufgenommen hat.

Bei direkter Wiederholung wird dasjenige Wort in den Satz eingebaut, das seinem Regenten näher steht, das andere bleibt Fragment.

- Der Internet-by-Call-Tarif ohne Anmeldung und Grundgebühr soll von/S von/PP 6 auf 4,9 Pfennig gesenkt werden.

Bewirkt der Schreibfehler eine Kategorieänderung, durch die die ähnlichste Struktur eine harte Bedingung verletzt, so müssen eventuell mehrere Bestandteile als Fragment angesehen werden.

- Unabhängig vom Patch-Level steht also jedes System, dass/S mit Sasser infiziert ist/S, sperrangelweit offen.

2.3 Mehrdeutigkeiten zwischen syntaktischen Kategorien

Die Einteilung von Worten in syntaktische Kategorien folgt den Richtlinien des STTS (vgl. 'Guidelines für das Tagging deutsche Textcorpora mit STTS'). Wo dieses Dokument unvollständig oder widersprüchlich ist, gelten folgende Richtlinien. Niemals sollte eine Kategorie nur deshalb gewählt werden, weil der Tagger sie vorhersagt!

2.3.1 ADJA oder NN?

Praktisch alle Adjektive können substantiviert werden. Gemäß STTS werden sie als ADJA eingeordnet, wenn sie klein geschrieben sind, sonst als NN.

- Die kleinen/ADJA hängt man und die großen/ADJA läßt man laufen.
- Die Roten/NN haben nichts Großes/NN bewirkt.

Attributiv gebrauchte Adjektive sind immer ADJA, ob sie groß oder klein geschrieben sind.

- die hohe/ADJA Steuer
- der Hohe/ADJA Rat

2.3.2 ADJD oder VVPP?

Partizipien werden laut STTS als ADJD eingeordnet, wenn sie adverbial gebraucht sind, sonst als VVPP.

- Der Feind wurde gestellt/VVPP.
- Der Feind ist gestellt/VVPP.
- Die Szene wirkt gestellt/ADJD.
- Die Szene ist gestellt/ADJD.

Wann liegt nun ein adverbialer Gebrauch vor? Wenn die Bedeutung offensichtlich passivisch ist, ist auch bei ‘sein’ ein Partizip anzunehmen:

- Der Tisch ist verrückt/VVPP. (= Man hat den Tisch verrückt)
- Die Tante ist verrückt/ADJD. ($\neq$ Man hat die Tante verrückt)

In vielen Fällen ist die Unterscheidung nicht klar. Das STTS gibt eine lange Liste mit Partizipien an, die fallweise ADJD sein können; diese ist weder vollständig noch ausschließlich. Keine Angabe macht es zu folgender Konstruktion:

- Das waren kleine Stifte, gedrechselt/VVPP aus Buchenholz.

In solchen Sätzen, die aus Relativsätzen verkürzt wurden, wählen wir stets VVPP.

2.3.3 ADV oder APPR?

Einige quantifizierende Adverben haben Homonyme unter den Präpositionen. Das sind die Worte ‘bis’, ‘gegen’, ‘unter’, ‘über’, und ‘zwischen’. Die Adverb-Lesart tritt nur auf, wenn das Wort direkt vor einer Zahl steht. Auch dann kann aber immer noch die Präpositions-Lesart eintreten. Daher muß die Satzaussage geprüft werden: kann der Satz paraphrasiert werden durch ‘nicht mehr als’, ‘ungefähr’, ’mehr als’, ‘weniger als’ oder ‘von...bis’, so liegt die Adverb-Lesart vor. Also:

- Hamburg besitzt über/ADV tausend Brücken
- Hamburg verfügt über/APPR tausend Brücken.

2.3.4 ADV oder KON?

Gemäß STTS sind die Konjunktionen ‘aber’, ‘doch’, ‘denn’ und ‘jedoch’ als KON zu bezeichnen, wenn sie zwischen zwei Konstituenten stehen, jedoch als ADV, wenn sie in die zweite eingeschoben sind. Diese Unterscheidung wird befolgt, obwohl sich durch die Umstellung keinerlei Sinnänderung ergibt.

- Wir haben den Krieg gewonnen, aber/KON den Frieden verloren.
- Wir haben den Krieg gewonnen, den Frieden aber/ADV verloren.

Die zweite Konstituente selbst wird aber in beiden Fällen als ‘KON’ untergeordnet.

2.3.5 ADV oder PTKVZ?

Die Unterscheidung zwischen freien Adverben und abgetrennten Verbzusätzen ist schwer zu treffen, weil diese aus jenen entstanden sind. Grundsätzlich ist zu prüfen, ob das Verb in der Nebensatzstellung mit der Partikel zusammen oder getrennt geschrieben würde:

- Wir werden auch kommen ⇒ Wir kommen auch/ADV
- Wir werden mitkommen ⇒ Wir kommen mit/PTKVZ

Da die Getrennt- und Zusammenschreibung aber ihrerseits sehr uneinheitlich geregelt ist, liefert dieser Test meistens keine eindeutige Antwort. Es ist dann zu prüfen, ob die Kombination von Verb+Partikel zu einer anderen als der kompositionalen Bedeutung oder Verwendung führt. Beispiel:

- Wir bauen den Schrank zusammen/ADV (= gemeinsam)
- Wir bauen den Schrank zusammen/PTKVZ (= aus Einzelteilen)

Die Veränderung kann auch darin bestehen, daß sich der Valenzrahmen des Verbs ändert:

- Ich werde dich zur Königin machen.
- *Ich werde dich zur Königin weitermachen.

Also:

- Ich mache weiter/PTKVZ.

Insbesondere auf diesem Gebiet enthält das Lexikon viele Lücken. Wenn also ein zusammengesetztes Verb mit deutlicher Sinnänderung fehlt, ist es in `Verben.txt` einzutragen.

2.3.6 APZR oder PTKVZ?

Verschiedene Partikel können sowohl zur Präposition als auch zum Verb treten:

- Sie stürzte mit ausgebreiteten Armen auf ihn zu/APZR.
- Die Tür fiel zu/PTKVZ.
- Sie irrten vierzig Jahre lang herum/PTKVZ.
- Er segelte um den Kontinent herum/APZR.

Wenn eine Präposition auftritt, zu der die Zirkumposition gehört, dann liegt APZR vor. Wenn keine Präposition auftritt, liegt entweder PTKVZ oder ADV vor; vgl. den Abschnitt ‘ADV oder PTKVZ?’.

2.3.7 PRELS oder PWS?

Das Wort ‘was’ kann sowohl Frage- als auch Relativpronomen sein. In der direkten Frage ist es Fragepronomen und im Relativsatz mit Satzreferenz Relativpronomen:

- Was/PWS soll das bedeuten? Es taget ja schon!
- Es tagt schon, was/PRELS ich sehr merkwürdig finde.

Bei Relativsätzen, die sich auf NP beziehen, ist zu prüfen, ob sie aus der betreffenden Gesamtfrage hervorgegangen sind oder ob das ‘was’ semantisch leer ist:

- Ich kann das, was/PRELS geschehen ist, nicht ungeschehen machen.
  (impliziert *nicht* die Frage ‘Was ist geschehen?’)
- Sie fragte ihn, was/PWS geschehen sei.
  (= Sie fragte: “Was ist geschehen?”)

2.3.8 VAFIN oder VVFIN?

Die Verben ‘sein’, ‘haben’ und ’werden’ werden immer als Auxiliarverben eingeordnet, egal ob sie mit oder ohne anderes Verb stehen.

- Ich bin/VAFIN, der ich bin/VAFIN.
- Wir sind/VAFIN lange umhergezogen.
- Wir sind/VAFIN müde.

2.3.9 VMINF oder VMPP?

Die Formen ‘wollen’, ‘können’ etc. werden auch dann als Infinitiv eingeordnet, wenn sie syntaktisch die Rolle eines Partizips einnehmen.

- Er hat nicht kommen können/VMINF.
  (= Er hat nicht zu kommen vermocht/VVPP.)

2.4 Mehrdeutigkeiten zwischen Labeln

2.4.1 ADV oder AVZ?

AVZ wird für alle Worte der Kategorie PTKVZ verwendet und ADV für Worte der Kategorie ADV; diese Entscheidung ist also zurückzuführen auf die Frage ‘ADV oder PTKVZ?’ (siehe oben).

2.4.2 ADV oder PP?

Präpositionaladverben wie 'dazu', 'hiermit' etc. sind Verschmelzungen von Präposition und Adverb. Logischerweise könnten sie also sowohl ADV als auch PP sein. Der Einheitlichkeit halber wird immer PP verwendet.

2.4.3 ADV oder PRED?

Als Prädikativum (PRED) wird nur ein Komplement bezeichnet, das in etwa eine Gleichsetzung ausdrückt:

- Wasser ist naß/PRED.
- Ich bin hier/PRED.
- Das ist die Wahrheit/PRED.

Es gibt einige Tests, die das Prädikativum von Objekten, adverbialen Bestimmungen etc. unterscheiden. Zum Beispiel muß es immer im Nominativ stehen:

- Ich bin der Schnitter/PRED Tod.
- Ich bin des Todes/OBJG.

Es alterniert mit den anderen Komplementen von 'sein':

- Ich bin gut/PRED.
- Ich bin gut/ADV im/PRED Rennen. (= liege in Führung)
- Ich bin gut/PRED im/PP Rennen. (= kann gut rennen)

Dennoch ist die Grenze fließend:

- Ich bin naß/PRED.
- Ich werde naß/PRED.
- Ich mache den Wal naß/PRED
- Ich halte den Wal naß/PRED.
- Ich halte den Wal für/OBJP naß/PN.

Wir erlauben das Label PRED an verschiedenen anderen Verben außer 'sein'. Richtlinie ist dabei, daß das Verb die elementare Aussage 'A ist B' ausdrückt oder sie zumindest impliziert. Daher gilt 'den Wal naß machen' als Prädikativum, denn es impliziert 'der Wal ist naß', während 'schön singen' nichts dergleichen impliziert (der talentierte Sänger kann häßlich sein wie die Nacht).

2.4.4 ADV oder S?

Wirkliche Interjektionen (PTKANT, ITJ) sind stets S. Sie sind immer Satzwurzeln, außer es gibt ein übergeordnetes Aussageverb. Nur wenn die Wortstellung deutlich macht, daß es sich um ein adverbial gebrauchtes Wort der Kategorie ADV handelt, ist ADV zu verwenden.

- Ja/S, das Wandern ist des Müllers Lust.
- Oh/S, was haben wir denn da?
- Ich bin ja/ADV nicht blöd.
- Er sagte ←ja/S.

2.4.5 APP oder DET?

Wenn Zahlausdrücke aus mehreren Worten bestehen, werden alle adjektivischen Bestandteile als ATTR oder DET dem Nomen untergeordnet.

- Die/DET ganzen/ATTR sieben/ATTR Weltwunder sind verschwunden.

Das gilt auch dann, wenn inhaltlich das eine Zahlwort das andere modifizieren könnte:

- Mehrere/DET hunderttausend/ATTR Menschen/SUBJ waren gekommen.

Wenn Zahlworte aber syntaktisch Nomen (wie 'Millionen' etc.), sind sie selbst DET zum Hauptnomen. Vorangehende Betandteile werden dem Zahlnomen untergeordnet.

- Die/DET 41/ATTR Millionen/DET Dollar waren nur eine Anzahlung.

Dagegen gelten Kombinationen von Maßnomen+Stoffnomen als Apposition:

- Trinken Sie eine/DET Tasse/OBJA Kaffee/APP.

2.4.6 APP oder KON?

In komplexen Nominalphrasen kommt gewöhnlich APP zum Einsatz, egal wieviele Bestandteile sie haben.

- der Inprise Application/APP Server/APP Plus/APP

KON wird nur verwendet, wenn eindeutig verschiedene Referenten bezeichnet werden:

- Es kamen der Bundeskanzler, Gerhard/APP Schröder/APP, der Bundespräsident/KON, Roman/APP Herzog/APP, und andere.

2.4.7 ATTR oder DET?

DET wird für Artikel und attributive Pronomen gebraucht. Alle anderen Bestandteile von NP, die nicht Nomen im weiteren Sinne sind (NN, NE, FM, TRUNC) sind als ATTR zu bezeichnen.

- die/DET fünf/ATTR besten/ATTR deutschen/ATTR Teilnehmer
- mein/DET lieber/ATTR Schwan
- jeder/DET gute/ATTR Künstler

Wenn sowohl Artikel als auch attributives Demonstrativpronomen stehen, ist der Artikel DET und das Pronomen ATTR.

- ein/DET jeder/ATTR Bürger
- manch/ATTR ein/DET Unternehmen

Treten zwei attributive Pronomen auf, so sind sie als Beiordnung anzusehen:

- dieses/DET unser/KON Land

2.4.8 AUX oder OBJI?

Infinitive, ob mit oder ohne Erweiterung, sind stets AUX, wenn sie am Hilfsverb stehen.

- Ich habe zwei Pferde in diesem Rennen laufen/AUX.
- Ich habe deshalb viel zu tun/AUX.
- Das betroffene Bundesland ist anzuhören/AUX.
- Gestern war ich eislaufen/AUX.

Wenn sie am Vollverb stehen, sind sie stets OBJI.

- Hör auf, mich anzulügen/OBJI.
- Laß uns nicht streiten/OBJI.

Das gilt auch für die Vollverben, die den Hilfsverben in ihrer Funktion sehr ähnlich sind ('lassen', 'scheinen', 'brauchen').

2.4.9 DET oder SUBJ?

Zahlworte, deren Kategorie NN ist, werden als Determiner angesehen. Haben sie selbst einen Artikel, so entstehen zwei DET-Kanten untereinander.

- zwei/DET Millionen/DET Dollar/SUBJ wurden eingenommen.
- Zehntausende/DET Siedler/SUBJ ließen sich dort nieder.

Wenn Zahlwort und Nomen durch Präposition getrennt sind, ist jedoch das Nomen PN.

- Millionen/SUBJ von/PP Mark/PN wurden verschwendet.

Ist das Zahlwort nachgestellt, so ist es stattdessen APP.

- Die Orks strömten aus dem schwarzen Tor, zehntausend/APP mindestens.

2.4.10 EXPL oder OBJA?

Steht das 'es' vor einem Objektsatz oder Subjektsatz, der zum vorigen Verb gehört, so ist es EXPL.

- Ich liebe es/OBJA.
- Ich liebe es/EXPL, daß ich als König geboren wurde/OBJC.
- Es/SUBJ ist mir gleichgültig.
- Es/EXPL ist mir gleichgültig, daß das Volk mich haßt/SUBJC.

Wenn kein kein Objektsatz, sondern ein normaler Nebensatz folgt, so ist es SUBJ (oder OBJA, wenn das Verb transitiv ist).

- Es/SUBJ ist mir gleichgültig, weil ich die Macht im Land habe/NEB.

Auch wenn der Nebensatz inhaltlich die Funktion des Objektsatzes übernimmt, ist dennoch OBJA zu wählen.

- Ich liebe es/OBJA, wenn ein Plan funktioniert/NEB.
- Ich liebe es/EXPL, daß der Plan funktioniert hat/OBJC.

2.4.11 EXPL oder SUBJ?

'es' bei unpersönlichen Verben ist stets SUBJ, nicht EXPL.

- Es/SUBJ regnet.

2.4.12 GMOD oder SUBJ?

Nominalphrasen mit einem Eigennamen im Genitiv werden manchmal auf diesen Namen verkürzt. In diesem Fall übernimmt der Eigenname die Funktion des fortgefallenen Wortes.

- Peters Noten waren schlecht, und Marias/GMOD Noten/SUBJ waren nicht viel besser.
- Peters Noten waren schlecht, und Marias/SUBJ waren nicht viel besser.

Das kann nicht nur die Funktion SUBJ sein, sondern z.B. auch OBJA oder PN.

- Meine Handschrift ist nicht die beste,
  aber Gunnars/OBJA kann ich überhaupt nicht lesen.
- Hast du dein Auto wieder oder kommst du mit Erikas/PN?

2.4.13 KOM oder KONJ?

Das Wort 'als' kann Worte vieler Kategorien beiordnen, fast wie eine Konjunktion. Wenn es einen ganzen Nebensatz einleitet, kann es sowohl über- als auch untergeordnet werden. Ist mit 'als' ein Vergleich ausgedrückt, so wird es als KOM dem Hauptsatz untergeordnet. Ist Gleichzeitigkeit ausgedrückt, so wird es als KONJ dem Nebensatzverb untergeordnet.

- Ich schrak auf, als/KONJ sie hereinkam/NEB.
- Ich weiß mehr, als/KOM du glaubst/CJ.

Die Vergleichsbedeutung tritt nur mit Komparativen auf; umgekehrt kann aber ein Hauptsatz mit Komparativ auch einen als-Nebensatz tragen.

- Ich wurde wütender, als/KOM ich je gewesen war.
- Ich wurde ruhiger, als/KONJ wir die Grenze hinter uns hatten/NEB.

2.4.14 NEB oder OBJC?

Das Label OBJC wird nur vergeben, wenn der Regent diese Funktion verlangt. Das sind nur einige Verben, deverbale Nomen und Adjektive.

- Ich verlange, daß man die Schlüssel der Stadt aushändigt/OBJC!
- Ich bin sicher, daß man sich mir fügen wird/OBJC.
- Die Frage, ob wir Spenden angenommen hätten/OBJC, ist irrelevant.

Wenn der untergeordnete Satz als Paraphrase von 'so daß' oder 'egal ob' anzusehen ist, ist er stattdessen NEB:

- Wir werden siegen, ob die Aristokratie uns unterstützt/NEB oder nicht.
- Wir tanzen, daß die Wände wackeln/NEB.

Deutliches Indiz für NEB ist, daß zu dem Satz noch das Objekt treten kann:

- Wir tanzen Samba, daß die Wände wackeln/NEB.
- *Wir verlangen eine Abrüstung/OBJA, daß abgerüstet wird/OBJC.

2.4.15 NEB oder REL?

Ein Nebensatz ist REL (oder OBJC), wenn er ein Relativpronomen oder Fragepronomen in seinem Skopus hat, sonst NEB.

- Wenn alle Stricke reißen/NEB, haben wir immer noch den Bunker.
- Wird/NEB dieses Recht nicht wahrgenommen, so erlischt es.
- Eine Sparte, die hohe Umsätze erbringt/REL.
- Der Ort, wo wir uns trafen/REL.
- Ein Bursche, wie du einer bist/REL.

Ist ein Relativpronomen in mehrere Sätze verschachtelt, so erzwingt es das Label REL nur für den untersten, weil es nicht im Skopus der anderen Sätze steht.

- Ich weiß/S nicht, ob das klug ist/OBJC, was du vorhast/REL.

2.4.16 OBJA oder OBJA2?

Wenn ein Verb zwei Akkusative nimmt, so ist derjenige das OBJA2, der im normalen Hauptsatz weiter rechts steht.

- Das kostet dich/OBJA den Führerschein/OBJA2!
- “Penner”/OBJA2 nennst du mich/OBJA?

2.4.17 OBJC oder REL?

Untergeordnete Sätze mit Frageworten (PWAV, PWAT) sind Objektsätze, wenn der Regent es verlangt, sonst Relativsätze.

- Ich will wissen, was passiert ist/OBJC!
- Es ist eine Brücke zusammengebrochen, was niemand überrascht/REL.
- Ich fragte, wie es ihr gehe/OBJC.
- Es ging ihr gut, wie ich hörte/REL.

Hierbei gilt derselbe Test wie wie bei der Frage ‘PRELS oder PWS?’.

2.4.18 OBJC oder S?

Untergeordnete Sätze sind immer dann S, wenn sie Hauptsätze sind, also Verbzweitstellung vorliegt. Alle anderen untergeordneten Sätze sind OBJC, NEB oder REL.

- Er sagte, das sei/S in Kambodscha nicht ungewöhnlich.
- Er sagte, daß das nicht Ungewöhnliches sei/OBJC.

Nur wenn ein Nebensatz gar nicht (oder auf einer anderen Sprachebene) untergeordnet ist, wird er als S bezeichnet.

- Was aber auch niemand überzeugte/S.
- Er sagte: "Weil ich dich liebe/S."

2.4.19 OBJC oder SUBJC?

Im Passiv wird das Objekt des Hauptverbs syntaktisch zum Subjekt. Ebenso wird der Objektsatz des Hauptverbs im Passiv zum Subjektsatz, ist also immer SUBJC und nicht OBJC.

- Die Regierung behauptet, die Verhältnisse hätten/OBJC sich gebessert.
- Oft wurde behauptet, die Verhältnisse hätten/SUBJC sich gebessert.

2.4.20 PP oder OBJP?

Präpositionalobjekte sind solche PP, die nicht frei zu einem Verb treten, sondern mit ihm zusammen eine nicht kompositionale Bedeutung erzeugen. (TIGER-Annotationsrichtlinien: "Präpositionalobjekte sind Präpositionalphrasen, die infolge eines Abstraktionsprozesses an das Verb gebunden sind.")

Im allgemeinen ist die Unterscheidung zwischen OBJP und PP schwierig, weil sie von semantischen Kriterien abhängt. Ein Test zur Unterscheidung von PP und OBJP ist beispielsweise, daß in einem Präpositionalobjekt die ursprüngliche Bedeutung der Präposition bis zur Bedeutungslosigkeit verblaßt (z.B. 'an' in 'glauben an' hat keine Ortsbedeutung). Dieser Test ist aber nur einseitig anwendbar, denn auch in vielen freien PP ist die Bedeutung verblaßt; so hat z.B. das 'im' in 'im Grunde' auch keine Ortsbedeutung mehr.

Ein anderes Kriterium ist, daß ein Präpositionalobjekt bei einem eigentlich transitiven Verb das Akkusativobjekt vertreten kann:

- Ich spiele den Mitspieler/OBJA an.
- Ich spiele auf/OBJP einen wunden Punkt an.
- Ich kaufe das Auto/OBJA.
- *Ich kaufe auf/OBJP der Straße.

Dieser Test funktioniert nur in einer Richtung, denn viele Präpositionalobjekte sind nicht transitiv (z.B. 'zurückgreifen'+'auf').

Wenn die Ersetzung der PP durch ein Pronominaladverb zu einer Sinnänderung führt, handelt es sich mit Sicherheit um eine normale PP:

- Ich arbeite an/OBJP meinem Bericht.
- Ich arbeite an/PP meinem Schreibtisch.
- Ich arbeite daran/OBJP. (= an dem Bericht, nicht dem Schreibtisch)

Wenn ein Verb ohne seine Präposition ungrammatisch wird, handelt es sich sicher um ein Präpositionalobjekt:

- Sie fielen über/OBJP ihn her.
- *Sie fielen her.

Wenn ein Verb eindeutig Richtungs- oder Ortsbedeutung hat, muß ein Präpositionalobjekt derselben Lokationsklasse angehören, während freie PP dieser Beschränkung nicht unterliegen:

- Er spielt auf/OBJP den Versailler Vertrag an.
- *Er spielt auf/OBJP dem Versailler Vertrag an.
- Er spielt auf/PP dem Kongress auf/OBJP den Versailler Vertrag an.
- Er saugt an/OBJP dem Strohhalm.
- *Er saugt an/OBJP den Strohhalm.
- Er saugt am/PP Anfang an/OBJP dem Strohhalm.

Präpositionalobjekte müssen eindeutig sein, während freie PP mehrfach auftreten können.

- Wir entscheiden den Fall mit/PP Sicherheit am/PP Montag.
- *Wir entscheiden über/OBJP den Fall über/OBJP andere Fragen.

2.4.21 S oder KON?

Stehen zwei ganze Sätze ohne Konjunktion nebeneinander, so wird der zweite dem ersten durch KON untergeordnet.

- Sagen/S Sie nichts, ich weiß/KON die Antwort schon.

Wenn die Sätze durch : oder ; getrennt sind, werden stattdessen zwei Satzwurzeln annotiert.

- Der Staat kann/S Feiertage festlegen; das Nähere regelt/S ein Bundesgesetz.

2.4.22 S oder PAR?

Das Label PAR wir nur verwendet, wenn die normale Unterordnung durch S einen nichtprojektiven Baum erzeugen würde. PAR wird also verwendet, wenn der Matrixsatz zwischen zwei Teilen des untergeordneten Satzes steht.

- "Das", sagte/PAR er, "ist/S mein Neffe."
- "Das ist/S mein Neffe", sagte/S er.

PAR wird nicht für andere Einschübe verwendet. Überzählige oder durch Komma markierte Konstituenten erfüllen entweder dieselbe Funktion wie die primäre Konstituente, oder sie sind dieser als APP oder KON beigeordnet.

- es werden auch penible/ATTR (eingehende/ATTR) Untersuchungen durchgeführt
- besonders bedeutsame Investitionen der Länder und Gemeinden (Gemeindeverbände/APP)

NP-Einschübe, die eine VP erläutern, können überhaupt nicht repräsentiert werden und sind S.

- Sie zerreißt ihre Maske, einer/S der wenigen bewegenden Momente des Stücks.

2.4.23 SUBJ oder PRED?

In kurzen Sätzen können Subjekt und Prädikat miteinander verwechselt werden. Das Subjekt ist jeweils das Wort, das die Aussage des Kontextes weiterführt, während das Prädikat neue Information liefert.

- Ich/SUBJ bin ein guter Hirt/PRED.
- Angeber/PRED sind sie/SUBJ alle.

Läßt sich eine von zwei NP zu einem Adjektiv umformulieren, so ist diese das Prädikat.

- Angeberisch/PRED sind sie alle.

2.4.24 SUBJ oder SUBJC?

Nominalphrasen jeder Art sind SUBJ, wenn sie Subjektbedeutung haben. Das gilt auch für substantivierte Infinitive, die der Kategorie NN angehören.

- Spaß/SUBJ ist wichtig.
- Das/SUBJ ist wichtig.
- Meins/SUBJ ist aber besser als deins.
- Das Surfen/SUBJ macht viel Spaß.

SUBJC wird nur für echte Verben in Subjektposition verwendet.

- Mir macht surfen/SUBJC mehr Spaß.
- Einen thermonuklearen Sprengkörper zu konstruieren/SUBJC ist kein Kinderspiel.

2.4.25 SUBJC oder REL?

Relativsätze können das Subjekt vertreten, sowohl solche mit ‘wer’ etc. als auch solche mit ‘der’ etc. Diese Anbindung ist immer zu wählen, wenn der Hauptsatz kein eigenes Subjekt hat und eine Paraphrase durch ‘die, welche oder ‘derjenige, der’ möglich ist.

- Wer wagt/SUBJC, gewinnt. (= derjenige, der wagt)
- Die Gott lieben/SUBJC, werden sein wie die Sonne. (= die, welche Gott lieben)

Anderenfalls ist REL oder OBJC zu wählen.

- Wer wagt/REL, der gewinnt.
- Wer wagt/OBJC, stellt die Börsenaufsicht durch Stichproben fest.

2.4.26 ZEIT oder OBJA?

Die Funktion ZEIT ist nur für Zeitausdrücke erlaubt. Entscheidend ist, ob der Satz durch adverbiale Zeitbestimmung paraphrasiert werden kann oder nicht.

- Ich laufe jeden Tag/ZEIT. (= an jedem Tag, also temporal)
- Ich laufe den Boston-Marathon/OBJA (= beim Boston-Marathon, also lokal)

Wenn beides möglich ist, ist dennoch OBJA zu wählen, wenn das Objekt sonst fehlen würde:

- Ich verbringe den Winter/OBJA im Süden.

2.5 Mehrdeutigkeiten der Struktur

2.5.1 Anbindung von ADV

- An Vollverb oder Hilfsverb?

  Tritt ein Adverb zwischen den Bestandteilen der Verbklammer auf, also im Mittelfeld, soll es stets an das Vollverb angebunden werden.

  - Ich habe gestern→ ferngesehen.

  Das gilt selbst dann, wenn es dabei das Subjekt überkreuzt:

  - Damals hat häufig→ (der Chef) die Sitzung geleitet.

  In allen anderen Fällen wird die kürzeste mögliche Anbindung gewählt. Steht das Adverb im Vorfeld, so modifiziert es das finite Verb:

  - Damals→ hat häufig der Chef die Sitzung geleitet.

Stehen aber sowohl Adverb als auch Vollverb im Vorfeld, so modifiziert das Adverb stattdessen das Vollverb:

- Selbst→ besucht hat er ihn zwar nie.

Steht das Adverb im Nachfeld, so modifiziert es das Vollverb:

- Ich habe ein neues Auto gekauft, ←endlich.

Liegt eine Inversion in der Verbgruppe vor, so wird das finite Verb modifiziert:

- Ich wurde gefragt, wie→ (ich das) habe verantworten können.

Verbgruppen, die nicht mit Hilfsverben, sondern durch Vollverben gebildet werden, erlauben die Unterordnung sowohl links als auch rechts, je nach der Satzbedeutung.

- Ich plane ←heute, zu verreisen. (Die Reise findet morgen statt.)
- Ich plane, heute→ zu verreisen. (Die Reise findet heute statt.)

Richtlinie: wenn ein Komma den Nebensatz markiert, muß das Adverb diesseits des Kommas angebunden sein.

- An ADV oder VVFIN?

  Stehen mehrere Adverben nebeneinander, modifizieren gewöhnlich beide dasselbe Verb oder Adjektiv. Nur wenn ein Adverb sich deutlich auf den Bedeutungsbeitrag des anderen bezieht, modifizieren sie einander.

  Hierbei gelten folgende Regeln:

  Wenn ADV zwei Adverben oder verwandte Wörter verbindet, ist meistens entweder das zweite Wort ein Adjektiv

  - Schnelles Handeln ist ←hierbei sehr→ wichtig/ADJD.

  oder das erste Wort ein Fokusadverb ('auch', 'schon', 'nur', 'erst'):

  - Nur→ hier finden wir die richtige Umgebung.

  Echte Adverben modifizieren einander nur, wenn beide Zeitbedeutung oder beide Ortsbedeutung haben.

  - Wie ist die Lust da→ oben?
  - Wir müssen noch→ heute aufbrechen.

  Ein Beweis dafür, daß ein Adverb das andere modifiziert, ist, daß die Adverbgruppe unverändert ins Vorfeld gestellt werden kann:

  - Auch heute ist die Krankheit noch nicht besiegt.
  - *Nicht genug schützen sich die meisten Einwohner.

  Adverben modifizieren einander immer von links nach rechts, außer bei einigen Worten, die explizit postmodifizierend sind ('bereits', 'etwa', 'genug').

  - Ich habe erst→ spät davon erfahren.
  - Ich wurde nicht früh ←genug unterrichtet.

2.5.2 Anbindung von ATTR

Stehen mehr als zwei Adjektive am selben Nomen, so sind sie alle nebenzuordnen. Wenn eine Konjunktion auftritt, verbindet sie nur die beiden benachbarten Adjektive, die anderen modifizieren direkt das Nomen.

- allgemeine, freie, gleiche ←und ←geheime Wahlen

2.5.3 Anbindung von APP

Wenn mehrere Worte eine NP bilden, modifiziert jedes Hauptwort (NN, NE, FM) das vorige Hauptwort. Das gilt auch, wenn es sich um fremdsprachliche Worte handelt, die erkennbar eine andere Struktur haben.

- Die Zeitung '← Het/APP ←financieele/APP ←dagblat/APP'

2.5.4 Anbindung von AUX

Die Reihenfolge der Unterordnung in der Auxiliargruppe richtet sich danach, welche Funktion die einzelnen Worte haben:

1. Tempus (VAFIN, VMFIN)
2. Modalität (VMINF)
3. Passiv (VAINF, VAPP von 'sein', 'werden')
4. Inhalt (VVINF, VVPP)

Die normale Reihenfolge ist also

- Als er geschlagen→ worden→ war...

Diese Struktur wird auch dann beibehalten, wenn die Verbgruppe umgestellt wird, selbst dann wenn der Syntaxbaum dadurch nichtprojektiv wird.

- Er kann nicht geschlagen worden sein.
- Das ist ein Fehler, den er nicht hätte übersehen dürfen.

2.5.5 Anbindung von GMOD

Wenn ein mehrteiliger Eigenname als Genitivmodifikator verwendet wird, ist dennoch der linke Bestandteil GMOD und die anderen APP. In diesem Fall kann also auch ein Wort GMOD sein, das selbst eindeutig im Nominativ steht.

- Onkel/GMOD ←Toms/APP Hütte

2.5.6 Anbindung von KOM

Vergleichsaussagen werden häufig durch einen Positiv (mit 'wie') oder Komparativ (mit 'als') ausgelöst, der tief in den Hauptsatz verschachtelt ist. Für diese Erscheinung wird keine Ausnahme von der Projektivität gemacht. Die KOM-Modifikation muß also so weit wie nötig nach oben verschoben werden.

- Ich bin müde wie ein Hund. (tief)
- Wir haben größere Probleme als das Fernsehprogramm (einmal angehoben).
- Es wurde berichtet, daß die Raketen mehr eigene Flugzeuge abgeschossen hätten als feindliche. (dreimal angehoben)

2.5.7 Anbindung von KON

In Nebenordnungen ist immer das erste Wort dem zweiten übergeordnet, egal ob eine Konjunktion auftritt oder nicht.

- Birnen, ←Bohnen ←und Speck
- Theseus ←und Ariadne

Es gibt nur zwei Situationen, in denen die Beziehung andersherum verläuft. Zum einen sind das die linken Teile von zusammengesetzten Konjunktion ('weder', 'entweder'):

- Weder→ Tod ←noch Teufel

Zum anderen werden verkürzte Verben dem vollständigen Verb untergeordnet, obwohl sie links stehen:

- Wir müssen das System aus- und→ umbauen.

Dies gilt nur für Verben. Andere verkürzte Wortteile werden normal behandelt:

- Der Aus- ←und Umbau ist anstrengend.
- Sein Gesicht wurde weiß- ←und rotglühend.
- beamten- ←und ←angestelltenrechtliche Fragen

2.5.8 Anbindung von NEB

Nebensätze, die von komplexen Verbgruppen abhängen, werden hoch angebunden, also an das finite Verb. Es wird nur dann das Vollverb modifiziert, wen das finite Verb nicht vorhanden oder z.B. hinter einer Koordination unzugänglich ist.

Sind mehrere finite Verben erreichbar, so entscheidet die Satzbedeutung, ob die kurze oder lange Anbindung zu wählen ist.

- Der Vorstand versprach, die Firma zu erhalten, wenn sich ein Käufer findet.
  (Der Erhalt hängt vom Käufer ab, das Versprechen nicht, also kurze Anbindung.)
- Wir zögern damit, eine Entscheidung zu treffen, weil wichtige Daten fehlen.
  (Das Fehlen begründet das Zögern, nicht die Entscheidung, also lange Anbindung.)

2.5.9 Anbindung von OBJA

Das Akkusativobjekt modifiziert stets das Vollverb, zu dem es gehört. Bei Konstruktionen mit 'lassen' modifiziert das Subjekt des untergeordneten Satzes 'lassen', das Objekt des untergeordneten Satzes modifiziert das andere Vollverb.

- Ich lasse (den) ←Kunden warten.
- Ich lasse das Auto→ warten.

2.5.10 Anbindung von OBJD

Wenn das Dativobjekt eine Valenz des Verbs ist, modifiziert es stets das Vollverb. Ist es allerdings ein ethischer (spontaner) Dativ, so gelten dieselben Regeln wie für adverbiale Bestimmungen, d.h. es wird stets die kurze Anbindung gewählt.

2.5.11 Anbindung von PP

- an komplexe NP

  Modifiziert eine Präposition eine komplexe Nominalphrase, die aus mehreren Nomen mit APP zusammengesetzt ist, so wird sie kurz angebunden, also an das letzte Nomen. Nur wenn der Sinn diese Möglichkeit eindeutig ausschließt, wird lang angebunden.

  - Der Kanzler, Herr Gerhard Schröder von der SPD (kurz)
  - Er war ein Mann, keine erbärmliche Maus, mit stolzem Sinn und starkem Geist. (lang)

- an Adjektive

  Präpositionen modifizieren nur dann Adjektive, wenn sie offensichtlich inhaltlich zu diesen gehören.

  - Wir betrachten ←(mit Genugtuung) die (von Sklaven)→ aufgebaute Pyramide.

Wenn die Präposition einem prädikativ gebrauchten Adjektiv folgt, ist sie meistens als dazugehörig anzusehen.

- Das ist zu hoch ←für mich.
- Mir war leichter ←ums Herz.

- innerhalb von Verbgruppen

  Hier gelten dieselben Regeln wie für Adverben (vgl. 'Anbindung von ADV').

- an Nomen oder Verb

  Kann eine Präposition sowohl an ein Nomen als auch an ein Verb modifizieren, so ist die Unterordnung nach der Satzbedeutung auszuwählen.

  - Man werde die Anteile←an France Telecom verkaufen.
    (= Man hält Anteile an France Telecom und will sie verkaufen.)
  - Man werde die Anteile an→ France Telecom verkaufen.
    (= Man hält Anteile, und France Telecom will sie kaufen.)

  Präpositionen modifizieren keine weiter rechts stehenden Nomen, außer wenn diese fragmentarisch oder beigeordnet sind.

- an Präpositionen

  Präpositionen modifizieren nur dann andere Präpositionen, wenn sie verblaßte Adjektive tragen, die den Charakter von Adverben haben. Anderenfalls sind sie nebenzuordnen, auch wenn dadurch z.B. das Vorfeld überfüllt wird.

  - Vor allem im Winter sind die Lebensbedingungen hart.
  - *Im Winter in Sibirien sind die Lebensbedingungen hart.

  Präpositionen dürfen außerdem dann andere Präpositionen modifizieren, wenn diese fragmentarisch oder beigeordnet sind.

  - Im Zweifel für den Angeklagten!
  - Er ist voller Ehrgeiz und mit Sicherheit auf unserer Seite.

2.5.12 Anbindung von REF

Kanten auf der Ebene REF ordnen nur Relativpronomen unter. Sie modifizieren jeweils dasjenige Wort, das den betreffenden Relativsatz trägt. Wenn mehrere Relativsätze beigeordnet sind, modifizieren beide Relativpronomen dasselbe Bezugswort.

Andere Pronomen werden nicht durch REF untergeordnet, auch wenn das Bezugswort im selben Satz zu finden wäre.

2.5.13 Anbindung von REL

Relativsätze modifizieren gewöhnlich das Nomen, das der Referent ist. In mehrteiligen NP wird das nächste Nomen modifiziert, also gewöhnlich das letzte. Relativsätze, die ganze Sätze erläutern, modifizieren das finite Verb.

Ist unklar, ob NP oder VP der Referent ist, so modifizieren die Relativpronomen der 'das'-Reihe eher Nomen und die der 'was'-Reihe eher Verben.

- Er hat eine Entscheidung getroffen, die nicht einfach war.
  ('Entscheidung' als Referent)
- Er hat eine Entscheidung getroffen, wofür wir dankbar sind.
  ('hat' als Referent)

2.5.14 Anbindung von SUBJ

Das Subjekt steht stets an demjenigen Verb, zu dem es inhaltlich gehört.

- Daß kein Integrationstest→ erfolgte→, war nur eines der Versäumnisse.

Sind mehrere Sätze koordiniert, so modifiziert jedes Subjekt das Verb seines Teilsatzes.

- Jesus→ schläft, was soll ←ich wachen?

Sind zwei Verben eng koordiniert, d.h. ohne zweites Subjekt oder Objekt, dann modifiziert das Subjekt das übergeordnete Verb.

- Das war (und ist unsere vordringlichste) ←Aufgabe.

2.5.15 Anbindung von VOK

Die Anrede modifiziert das nächste Wort (nicht Satzzeichen) links von ihr. Steht sie ganz vorn, so modifiziert sie stattdessen das nächste Wort rechts.

- "Flieht, ←ihr Narren!" rief Gandalf.
- Sir→, es ist angerichtet.

Wenn sie als erstes in einer direkten Rede steht, so modifiziert sie das nächste Wort, auch wenn der Matrixsatz vorausgeht.

- Er sagte, "Junge→, komm her."

2.5.16 Anbindung von ZEIT

Wenn eine absolute Zeitangabe zusammen mit einer Präposition auftritt, modifiziert die Zeitangabe das Verb und die Präposition die Zeitangabe.

- zwei Tage ←nach dem Unfall wurde die Anklage erhoben.

2.5.17 Projektivitätsregeln

Grundsätzlich sind nur projektive Syntaxstrukturen erlaubt. Ausnahmen werden nur für eine bestimme Gruppe von Phänomenen gemacht, die alle mit der Verbletztstellung der Klammerstellung zusammenhängen.

1. Komplexe Verbphrasen
   (a) mit Topikalisierung:
      - Was sollen wir tun?

      Hier kollidieren die Regeln für Verbgruppen (VVINF unter VMFIN) mit der Regel für Vollverben (OBJA nur an Vollverb). Statt eine von beiden aufzugeben, erlauben wir die nichtprojektive Anbindung des Objektes. Das gilt auch für OBJD, OBJG, etc. Insbesondere gilt es für OBJP, aber **nicht** für PP oder KOM:
      - Hierzu hat das Verfassungsgericht ein Urteil erlassen.

      Hier muß die Präposition das finite Verb modifizieren, obwohl inhaltlich die Alternative etwas passender wäre.

   (b) mit Scrambling:
      - Er betrachtet es als eine Genugtuung, daß er sich doch noch hat habilitieren können.
2. Koordinationen und Appositionen

   Sowohl markierte als auch unmarkierte Koordinationen können nach rechts extraponiert werden, auch wenn dadurch die Projektivität verletzt wird:
   - Ich habe eine triumphale Rückkehr im Sinn und keinen Bittgang.
3. Relativsätze und Objektsätze

   Diese Art von untergeordneten Sätzen ist fast immer extraponiert. Nichtprojektivität kommt zustande, wenn das Bezugswort in eine komplexe Verbphrase verschachtelt ist.
   - Man muß dem Angeklagten das Recht einräumen, sich zu äußern.
   - Ich habe ein Angebot erhalten , das ich nicht ablehnen konnte.

2.5.18 Ellipsen

Elliptische Sätze sind das größte Repräsentationsproblem für die Grammatik. In einigen Fällen dürfen die Kinder von fortgefallenen Worten ersatzweise deren Regent modifizieren, in anderen Fällen müssen sie zum Fragment erklärt werden.

- von Nomen

  Fällt ein Nomen weg, so kann ein attributives Adjektiv dessen Rolle einnehmen. Es kann SUBJ, OBJA etc. sein und durch Artikel als DET modifiziert werden. Ist kein Adjektiv vorhanden, so sind evtl. verbleibende Vorkommen von 'der' etc. immer Demonstrativpronomen, nicht Artikel.

  Ein Eigenname als Genitivattribut kann ebenfalls die Funktion des entfallenen Nomen einnehmen (vgl. 'GMOD oder SUBJ?').

  Fällt ein Subjekt-Demonstrativpronomen weg, das einen Relativsatz trägt, so kann dieser ersatzweise als SUBJC untergeordnet werden.

- von Verben

  Fällt ein finites Verb weg, so werden Subjekt und Objekte zu Fragmenten. Es gibt **keine** Unterordnung zwischen ihnen. Dagegen können Präposition und Adverben ersatzweise die nächste NP modifizieren (wie im Fall von Koordinationen).

  Fällt das Hilfverb eines Relativsatzes weg, so kann ein verbleibendes Partizip ersatzweise den Referenten als ADV modifizieren.

  Fällt ein infinites Vollverb weg, so modifizieren alle adverbialen Bestimmungen (ADV, PP, KOM) stattdessen das Hilfsverb. Ihre Label bleiben unverändert, d.h. die fehlende AUX-Kante wird nicht ersetzt.

  In Fällen wie 'falls nötig' können Konjunktionen ersatzweise das Adjektiv als KONJ modifizieren. Das Adjektiv darf den Hauptsatz als NEB modifizieren.

  Wird ein Relativsatz auf das Fragewort verkürzt, so darf dieses als S dem Hauptsatz untergeordnet werden:

  – Ich weiß auch nicht ←warum/S.

2.6 Mehrdeutigkeiten der Morphologie

2.6.1 nom oder acc?

Wenn die syntaktische Funktion von Fragmenten eindeutig zu erkennen ist, ist die Variante zu wählen, die im ganzen Satz gestanden hätte.

- Hände/acc hoch! (= Nehmt die Hände/OBJA hoch!)
- Mann/nom über Bord! (= Ein Mann ist über Bord gefallen!)

Wenn die ursprüngliche Funktion nicht zu rekonstruieren und die Form mehrdeutig ist, ist vorzugsweise der Nominativ oder der Singular zu wählen.

- Piraten/nom!
- Streber/sg!

2.7 Einzelne Konstruktionen

2.7.1 'als ob'

Die Fügung 'als ob' gilt als eine Konjunktion. 'als' modifiziert 'ob' als KONJ und braucht kein Argument. Dasselbe gilt für Wendungen wie 'ohne daß'.

2.7.2 'je... desto'

Ein Nebensatz kann mit 'je' eingeleitet werden, wenn der Hauptsatz 'desto' verwendet. Beide Worte modifizieren als ADV den Komparativ, neben dem sie stehen. Das Verb des Nebensatzes modifiziert wie üblich das Verb des Hauptsatzes.

- Je/ADV näher der Berg kam, desto/ADV schwerer wurde der Ring.

2.7.3 'mehr als'

Die Ausdrücke 'mehr als' und 'weniger als' sind normale Paare von Pronomen und Vergleichswort, wenn sie adverbial vorkommen.

- Ich liebe dich mehr/ADV als/KOKOM mein Leben.

'mehr' kann auch attributiv vorkommen; auch dann ist ein späteres 'als' möglich.

- Wir haben mehr/PIAT Autos verkauft als/KOKOM die Konkurrenz.

Wenn 'mehr als' aber innerhalb der Adjektivphrase auftritt, wird es behandelt wie ein mehrteiliges Adverb, d.h. 'mehr' modifiziert als Adverb das Zahlwort, und 'als' verliert sein Komplement. Dadurch ist gewährleistet, daß das Nomen der Kopf der gesamten Phrase bleibt und z.B. zur Kongruenzüberprüfung verfügbar ist.

- Wir haben mehr/ADV als/KOKOM ein Mittel/OBJA, dich zum Reden zu bringen.
- Die Belange mehr/ADV als/KOKOM eines Landes/GMOD sind betroffen.

Diese Konstruktion kann auch an Verben auftreten, die Zahlsinn besitzen.

- Damit kann die Bank die Zahl ihrer Privatkunden mehr/ADV als/KOKOM verdoppeln.

Exakt dasselbe gilt für die Fügung 'bis zu': 'bis' modifiziert das Zahlwort, und 'zu' verliert sein Argument.

- Es sind bis/PP zu/PN 3000/ATTR Klagen/SUBJ zu befürchten.

Wenn aber die NP keine andere Rolle ausfüllt, ist sie normal als PN zu bezeichnen:

- Es droht eine Gefängnisstrafe bis/PP zu/PN sechs/ATTR Monaten/PN.

2.7.4 'so... auch'

Wenn ein Nebensatz mit 'so' eingeleitet wird, modifiziert 'so' als ADV das Adjektiv, vor dem es steht. Das Nebensatzverb modifiziert wie üblich das Hauptsatzverb.

- So/ADV eifrig/ADV wir auch suchten, wir fanden nichts.

2.7.5 'wir alle'

Die quantifizierenden Pronomen 'alle', 'beide' etc. können ein vorausgehendes Pronomen modifizieren, wenn sie direkt folgen. Stehen sie weiter entfernt, modifizieren sie stattdessen das finite Verb als ADV.

- Wir ←alle sind Sünder.
- Höhepunkte gab (es) ←viele.

2.7.6 Idiomatische Ausdrücke

Datumsangaben gelten als Kombination von Attribut und Apposition.

- der 21.→ Dezember ←1995

Idiomatische Ausdrücke und Konstruktionen werden soweit möglich behandelt, als wären sie freie Kombinationen.

- Sie schicken Tonband/OBJA nach/PP Tonband/PN an die Verlage.
- Gott/OBJD sei/S Dank/SUBJ, daß das vorbei ist.

Wenn keine reguläre Unterordnung möglich ist, müssen die betreffenden Worte als Fragmente angesehen werden.

- Die Hersteller müssen auf/PP Teufel/VOK komm/S raus/AVZ verkaufen.

3 Ungelöste Probleme

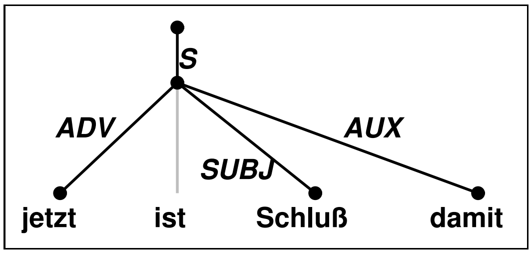
Fast alle regelmäßig vorkommenden Phänomene des Deutschen können angemessen modelliert werden. Es verbleiben jedoch Ausnahmen; einige davon sind grundsätzlich nicht im Dependenzformat zu lösen, andere wurden bislang wegen extremer Seltenheit nicht eingearbeitet.

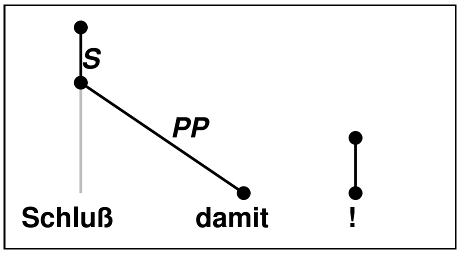
Wenn ein einigermaßen regelhaft in akzeptablem Deutsch auftretendes Phänomen nicht richtig modelliert werden kann oder von der Grammatik hart bestraft wird, so ist das auf jeden Fall ein Fehler in der Grammatik. Alle diese Fehler sollten hier vermerkt werden, nicht im `TODO` oder sonstwo.

Bekannter Fehler 1: Unzulängliche Behandlung von Ellipsen

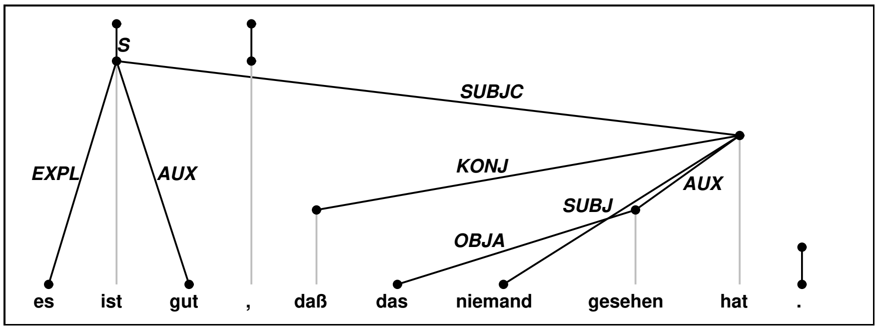
Das schwerste und häufigste Problem stellen Ellipsen dar. Fast jede Konstituente eines regelhaften Satzes kann fortgelassen werden. Insbesondere kann der Kopf einer Phrase fehlen, obwohl andere Teile der Phrase verbleiben. Will man dem Parser nicht erlauben, spontan zusätzliche virtuelle Anbindungen zu erlauben, so müssen diese Worte als Fragment behandelt werden.

Beispielsweise werden Hilfsverben oft fortgelassen. Hierbei fallen zwei Fragmente an. In manchen Fällen kann hilfsweise das eine unter das andere untergeordnet werden (der Parser wählt diese Möglichkeit deshalb, weil ein Fragment billiger ist als zwei). Die Grammatik erlaubt beispielsweise die Unterordnung von Adverben und Präpositionen unter Nomen:

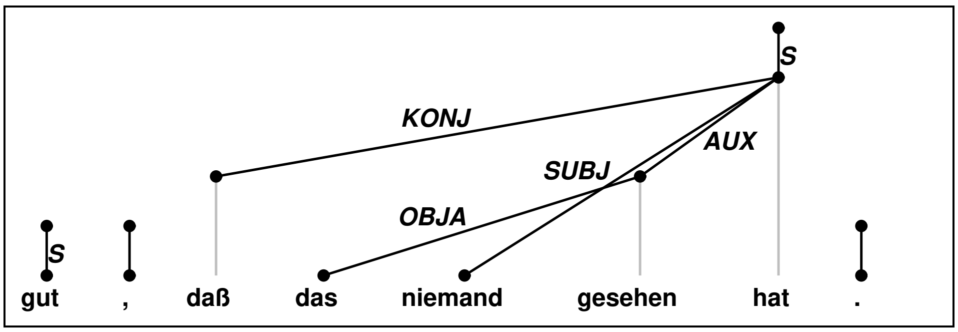




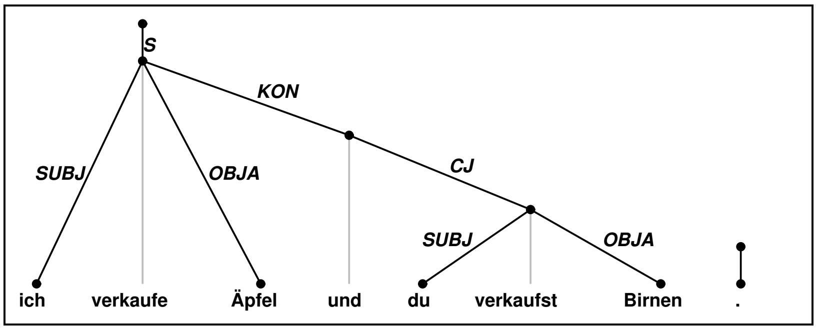
Oft fällt das Auxiliarverb weg, wenn es einen Subjektsatz mit einem Prädikat verbindet:



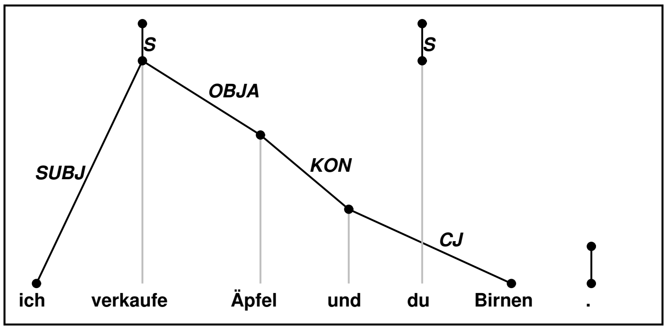
In diesem Fall kann die Grammatik nur zwei Fragmente melden.



Sehr oft fällt die wörtliche Wiederholung eines Wortes fort. Das ist oft der Fall, wenn mehrere ähnliche Sätze beigeordnet werden.

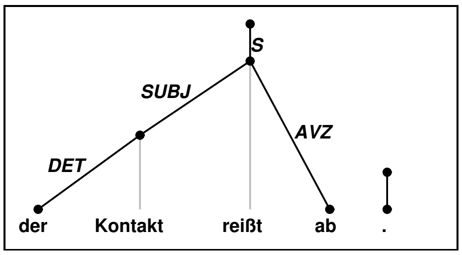


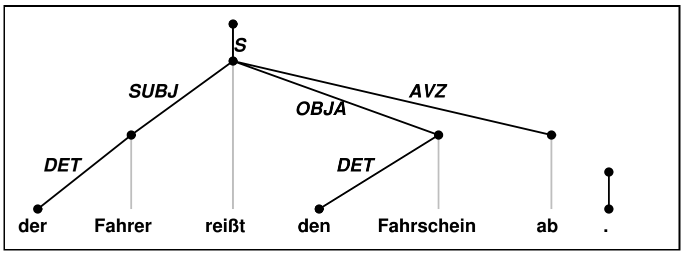
Dadurch wird eines der beigeordneten Worte heimatlos und kann nur als Fragment analysiert werden. Meistens ist das das Subjekt des zweiten Satzes:



Vgl. eventuell das Label `NP2`.

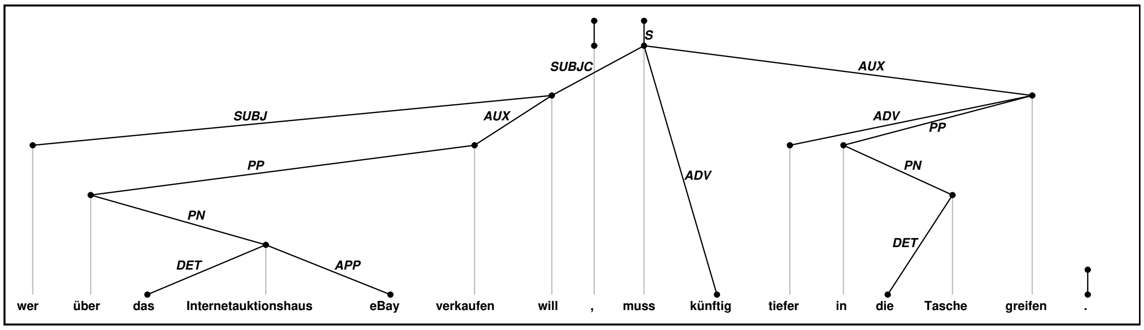
Oft fällt das Objekt eines Verbs fort, wobei die Grenze zwischen falschem und richtigem Gebrauch sehr schwer zu bestimmen ist. Manche Verben werden gleichermaßen transitiv und intransitiv gebraucht:





In diesem Fall muß es entweder zwei verschiedene Lexikoneinträge geben, oder die Strafe für das Fehlen muß einen sehr hohen Wert erhalten. (Derzeit wird immer die zweite Variante gewählt, damit das Lexikon möglichst klein bleibt.)

Nun gibt es aber auch Worte, die zwar transitiv sind, aber dennoch manchmal intransitiv gebraucht werden:



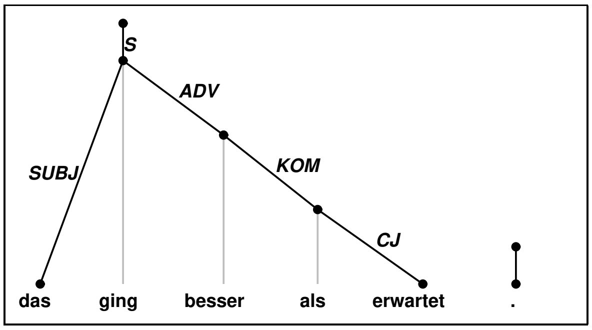
Wenn man die überwiegende Verwendung weiterhin als bevorzugt modellieren will, werden solche Sätze schlecht bewertet.

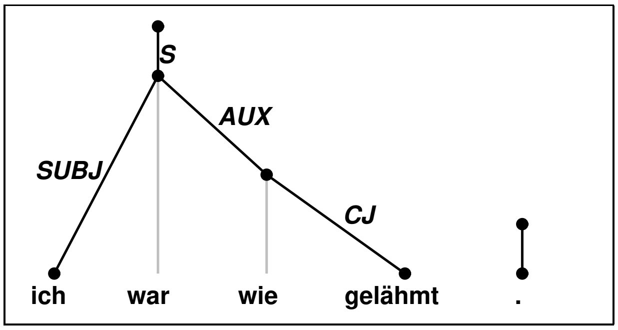
Oft wird eine NP alleinstehend verwendet, wenn sie dazu dient, die vorige Aussage zusammenzufassen:

‘Ich traf Neil O’Flannagan, [und das war] eine Ehre, die zuvor nur dreizehn Menschen hatten.’

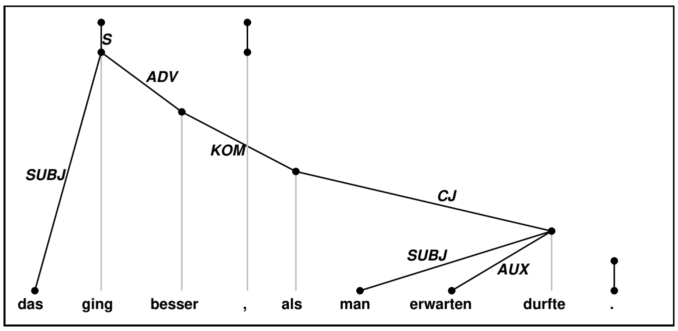
Bekannter Fehler 2: Pseudopassiv durch `KOKOM` wird nicht erkannt

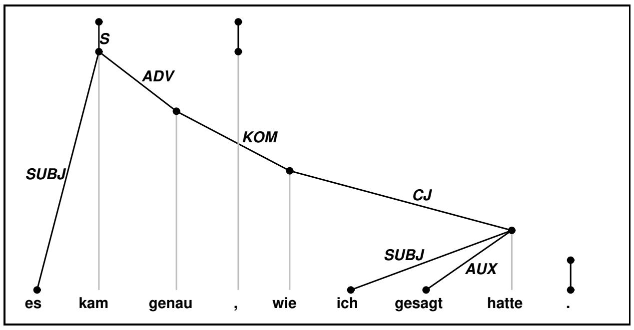
Objekte können auch dann fehlen, wenn eine Konstruktion mit ‘wie’ oder ‘als’ gebildet wird:





Hierfür wird eine Ausnahme von der Existenzbedingung gemacht, wenn ein Partizip direkt von ‘wie’ oder ‘als’ dominiert wird. Das Vergleichswort kann aber auch weiter entfernt stehen:



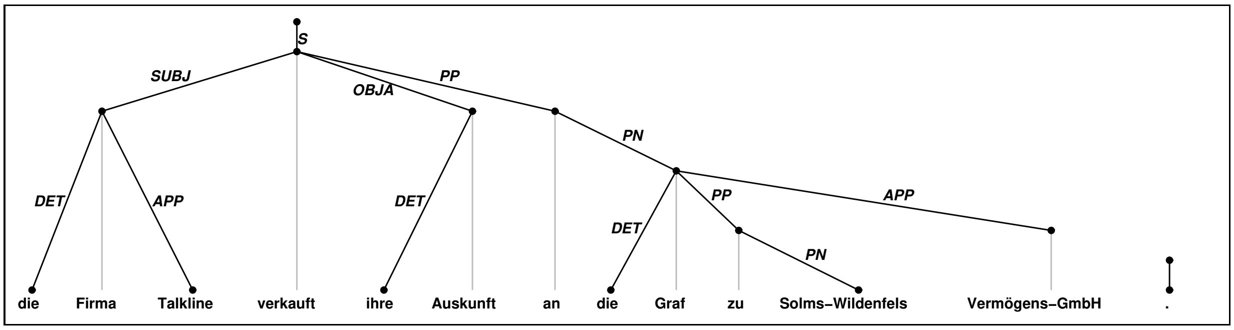


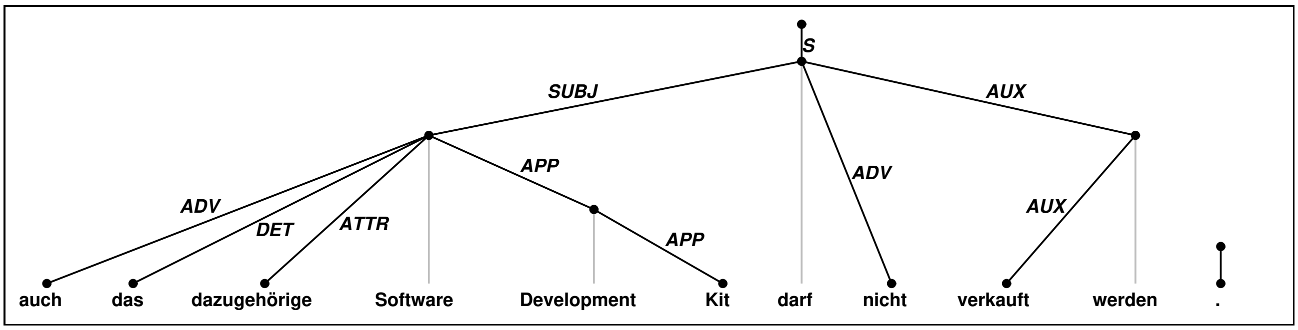
In diesen Fällen kann die Ausnahmesituation nicht erkannt werden, und die Grammatik meldet einen Fehler.

3.1 Probleme bei der Tokenisierung

Einige Probleme ich nicht als Fehler in der Grammatik an, sondern sie müßten im Tokenizer behoben werden.

Wenn Eigennamen aus normalen deutschen Worten zusammengesetzt sind, ergeben sich oft Kongruenzfehler, wenn sie Wort für Wort analysiert werden ('die Graf', 'das Software'):





Solche Namen sollten als ein einziges Lexem aufgefaßt werden, was aber eine Komponente für *named entity recognition* voraussetzt.

Alternativ könnte man den Namen anders analysieren, so daß die sinnvollere Unterordnung gewählt werden kann ('die GmbH', 'das Kit'). Dazu müßte aber die strenge Rechtsverzweigung aufgegeben werden, was die Mehrdeutigkeit sehr erhöhen würde.

Allgemein ist leider die Tendenz zu beobachten, zusammengesetzte Hauptworte (wie im Englischen) durch Leerzeichen zu bilden statt durch Bindestriche:

“Ich wollte die Charaktere *der Manson Familie* als Repräsentanten derselben Archetypen zeigen, die man in der griechischen Tragödie findet.”

Derartige Komposita können nicht erkannt werden und führen dann zumindest zu Flektionsfehlern.

Wo der Bindestrich doch auftritt, wird er bisweilen durch falsche Tokenisierung wieder vernichtet:

“In 13 der 16 Stadtteilparlmente sitzen jeweils 19 Beiräte , nur in dreien (Kalbach , Harheim und Nieder-Erlenbach) kommen jeweils neun *“ Vor-Ort “ -Politiker* zu den monatlichen Sitzungen .”

Hier sollte die Zusammensetzung als ein Nomen behandelt werden.

3.2 Verschiedenes

Diverse seltene Konstruktionen sind nicht richtig modelliert. Alle folgenden Probleme sind ungelöst.

Bekannter Fehler 3: Metagrammatische Benutzung kann nicht richtig verarbeitet werden

Beispiel: ‘Der Online-Versand verkauft Adolf Hitlers *Mein Kampf.*’

Hier besteht Kasus-Inkongruenz, weil das Akkusativobjekt ein Buchtitel ist, der in sich natürlich Nominativ ist. Ähnlich wie bei mehrteiligen Namen müßte der Titel als ein einziges zwar intern strukturiertes, aber nicht nach außen kongruierendes Lexem angesehen werden. Hier macht sich bemerkbar, daß die Dependenzgrammatik keine Zwischenknoten kennt.

Bekannter Fehler 4: Akkusativus absolutus ist nicht vorgesehen

‘Sie stehen sich *Auge in Auge* gegenüber.’

‘Er kam kurz darauf zurück, *einen bärtigen Mann im Schlepptau.*’

Ganz selten existiert auch im Deutschen ein Akkusativus absolutus. Wie soll man diese Konstruktion erlauben, ohne beliebige Beziehungen der Art $\mathrm{N} \bullet\!\!\longrightarrow \mathrm{V}$ zuzulassen?

Bekannter Fehler 5: Genitivus absolutus ist nicht vorgesehen

‘Solange das Denken in einer Epoche nicht auf sein eigenes Niveau gelangt, kann man *strengen Sinnes* nicht vom Fortschritt in der Philosophie reden.’

Ein Genitivus absolutus ist ebenfalls möglich, aber nur mit ganz bestimmten Konstruktionen, die man wohl alle eigens aufzählen müßte, um sie zu erlauben.

Bekannter Fehler 6: Idiomatische Ausdrücke verletzen Syntaxregeln

Idiomatische Ausdrücken können nur dann behandelt werden, wenn sie zwar semantisch anormal sind, aber den normalen Syntaxregeln gehorchen, also Wendungen wie ‘ins Kraut schießen’.

Wenn ein Idiom aber auch syntaktisch anormal ist, kann es gewöhnlich nicht vollständig analysiert werden:

'Die Chipschmieden sind gezwungen, *auf Teufel komm raus* zu verkaufen.'

Bekannter Fehler 7: Im Funktionsverbgefüge können sich Valenzrahmen ändern

Die möglichen Argumente von Verben sind jeweils im Verb selbst als Valenzrahmen kodiert. Das stellt sicher, daß etwa ein `OBJA` nur mit einem Verb auftreten darf, daß als transitiv markiert wird. Dasselbe gilt für alle anderen Nominalobjekte.[2] Dabei ist dem Phänomen Rechnung getragen, daß zusammengesetzte Verben sich anders verhalten als ihre Grundformen, selbst dann, wenn sie im Satz getrennt geschrieben werden.

Nun gibt es aber Verben, deren Valenzrahmen sich auch dann ändert, wenn sie in bestimmten festen Fügungen verwendet werden. Beispiel:

- Ich mache Spaß(OBJA).
- Ich mache dich(OBJA) glücklich(PRED).
- Ich mache dich(OBJA) zum König(PRED).
- *Ich mache zu gehen(OBJC).

Aber:

- Ich *mache Anstalten*, zu gehen(OBJC).

In diesem Fall könnte man das OBJA als Komplement zu 'Anstalten' ansehen. Aber auch andere Verben zeigen dies Verhalten:

'Ich *habe es schwer*, mich durchzusetzen.'

'Ich *bin froh*, daß das vorbei ist'.

Diese Art der Valenzänderung ist derzeit nicht repräsentierbar.

Bekannter Fehler 8: Der Objektsatz mit 'was' ersetzt das Objekt

Jedes Akkusativobjekt kann durch einen Objektsatz mit 'was' ersetzt werden:

'Was da auf dem Boden herumlag, hatten Angestellte des Hessen-Centers selber heruntergeklopft, um eben einem derartigen Steinschlag vorzubeugen.'

Die Grammatik erkennt diese Konstruktion nicht als Objekt; stattdessen ist der daß-Satz ein normaler Relativsatz, und die Valenz bleibt unerfüllt.

Bekannter Fehler 9: Dialekt kann nicht erfaßt werden

Dialekte des Deutschen enthalten viele abweichende Formen, oft aber auch normale. Daher können sie nicht pauschal als `FM` abgeschrieben werden:

" wolle mern reilasse ? "

[2] Ein Dativobjekt darf allerdings frei auftreten (ethischer Dativ).