

Lösungsvorschlag für das Übungsblatt 8.

Aufgabe 1.

a)

a)

Zeile	Erzeugungsschritt	Regel
1	S	
2	NP VP	R1
3	Det A N VP	R6
4	Any A N VP	L3
5	Any intelligent N VP	L4
6	Any intelligent cat VP	L2

Nach den Regeln kann der Satz weiter nicht erzeugt werden, deswegen ist eine neue Regel hinzuzufügen: (R9) VP -> V AdvAdv SAdv

7	Any intelligent cat V AdvAdv SAdv	R9
8	Any intelligent cat dies AdvAdv SAdv	L1
9	Any intelligent cat dies really SAdv	L6
10	Any intelligent cat dies really well	L5

b)

Zeile	Erzeugungsschritt	Regel
1	S	
2	NP VP	R1
3	Det A N VP	R6
4	A A N VP	L3
5	A awful N VP	L4
6	A awful linguist VP	L2
7	A awful linguist V NP	R2
8	A awful linguist seeks NP	L1
9	A awful linguist seeks Det A N	R6
10	A awful linguist seeks a A N	L3
11	A awful linguist seeks a nice N	L4
12	A awful linguist seeks a nice idea	L2

c)

Zeile	Erzeugungsschritt	Regel
1	S	
2	NP VP	R1
3	Det A N VP	R6
4	A A N VP	L3
5	A nice N VP	L4
6	A nice unicorn VP	L2
7	A nice unicorn V NP AdvAdv SAdv	R4
8	A nice unicorn finds NP AdvAdv SAdv	L1
9	A nice unicorn finds Det A N AdvAdv SAdv	R6
10	A nice unicorn finds any A N AdvAdv SAdv	L3
11	A nice unicorn finds any good N AdvAdv SAdv	L4
12	A nice unicorn finds any good idea AdvAdv SAdv	L2
13	A nice unicorn finds any good idea really SAdv	L6
14	A nice unicorn finds any good idea really quickly	L5

b)
Any linguist sleeps every unicorn well.

Zeile	Erzeugungsschritt	Regel
1	S	
2	NP VP	R1
3	Det N VP	R5
4	Det N V NP SAdv	R3
5	Det N V Det N SAdv	R5
6	Any linguist sleeps every unicorn well	L3,L2,L1,L3,L2,L5

Nach der PSG sind alle Verben sozusagen transitiv, da einem Verb eine Nominalphrase folgt.

Lösung – Einführung von transitiven und intransitiven Verben.

Aufgabe 2.

a)

PSG'	PSG
Transitive und intransitive Verben: VP-> V2 NP VP->V1	Alles zusammen
Nominalphrasen: Unterscheidung von Eigennamen (kein Determinator vorne), Nomen (vorne ist ein Determinator und mind. ein Adjektiv)	Alle Nomen folgen einem Determinator
Adjektivphrasen können rekursiv angewendet werden (mehrere Adjektive nacheinander)	Keine AP, können nur einzeln angewendet werden (R6,R7)
Adverbialphrasen können rekursiv angewendet werden	Keine AdvP, können nur einzeln angewendet werden (R3,R4)

b)

Zeile	Erzeugungsschritt	Regel
1	S	
2	NP VP	R1
3	Det N1 VP	R7
4	Det AP N0 VP	R8
5	Det A AP N0 VP	R13
6	Det A A N0 VP	R12
7	Det A A N0 VP AdvP	R4
8	Det A A N0 V AdvP	R2
9	Det A A N0 V AdvAdv AdvP	R11
10	Det A A N0 V AdvAdv SAdv	R10
11	Every nice intelligent unicorn sleeps really well	Lexikon

b)

Zeile	Erzeugungsschritt	Regel
1	S	
2	NP VP	R1
3	N2 VP	R5
4	N2 VP AdvP	R4
5	N2 V2 NP AdvP	R3
6	N2 V2 N2 AdvP	R5
7	N2 V2 N2 AdvAdv AdvP	R11
8	N2 V2 N2 AdvAdv SAdv	R10
9	Jack greets Mary really quickly	Lexikon

c)

Zeile	Erzeugungsschritt	Regel
1	S	
2	NP VP	R1
3	N2 VP	R5
4	N2 V2 NP	R3
5	N2 V2 NP PP	R9
6	N2 V2 Det N1 PP	R7
7	N2 V2 Det AP N0 PP	R8
8	N2 V2 Det A N0 PP	R12
9	N2 V2 Det A N0 P NP	R14
10	N2 V2 Det A N0 P Det N1	R7
11	N2 V2 Det A N0 P Det AP N0	R8
12	N2 V2 Det A N0 P Det A AP N0	R13
13	N2 V2 Det A N0 P Det A A N0	R12
14	Bill finds any awful cat with a nice intelligent unicorn	Lexikon

d)

Zeile	Erzeugungsschritt	Regel
1	S	
2	NP VP	R1
3	Det N1 VP	R7
4	Det AP N0 VP	R8
5	Det A AP N0 VP	R13
6	Det A A AP N0 VP	R13
7	Det A A A N0 VP	R12

8	Det A A A N0 VP AdvP	R4
9	Det A A A N0 V1 AdvP	R2
10	Det A A A N0 V1 AdvAdv AdvP	R11
11	Det A A A N0 V1 AdvAdv AdvAdv AdvP	R11
12	Det A A A N0 V1 AdvAdv AdvAdv SAdv	R10
13	A good intelligent nice linguist sleeps really really quickly	Lexikon

c)

Zusätzliche Regel:

(R15) S -> S C S

(R16) VP -> VAdv VP

Zusätzlicher Eintrag im Lexikon:

(L10) C -> {and} (C – für Konjunktion)

(L5) V2 -> {greets, beats, seeks, finds, loves, kisses} – "loves" und "kisses" sind hinzugefügt

(L2) N2 -> {Jack, Mary, Bill, Rose} – "Rose" ist hinzugefügt

(L11) VAdv -> {never} (VAdv – für Adverbien, die vor dem Verb geschrieben werden)

a)

Zeile	Erzeugungsschritt	Regel
1	S	
2	S C S	R15
3	NP VP C NP VP	R1,R1
4	N2 VP C N2 VP	R5,R5
5	N2 V2 NP C N2 V2 NP	R3,R3
6	N2 V2 N2 C N2 V2 N2	R5,R5
7	Jack loves Mary and Bill loves Rose	L2,L5,L2,L10,L2,L5,L2

b)

Zeile	Erzeugungsschritt	Regel
1	S	
2	NP VP	R1
3	N2 VP	R5
4	N2 VAdv VP	R16
5	N2 VAdv V2 NP	R3
6	N2 VAdv V2 N2	R5
7	Jack never kisses Mary	L2,L11,L5,L2

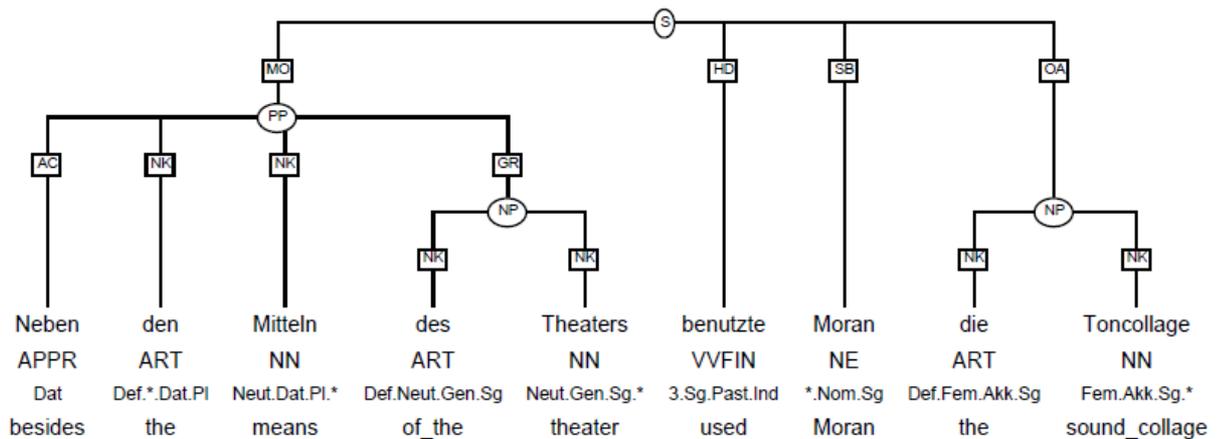
Aufgabe 2.

a)

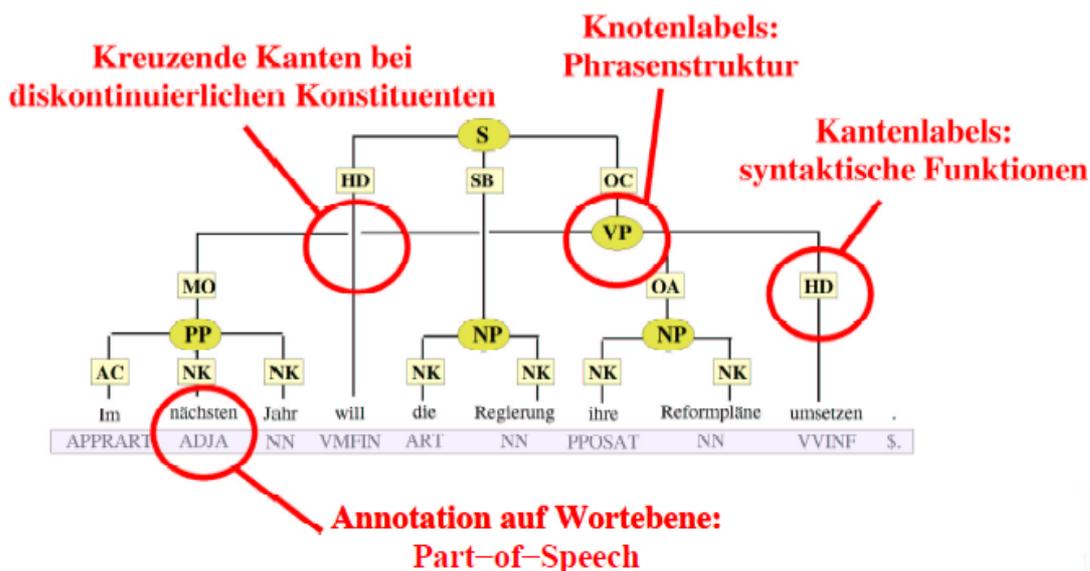
TIGER: Linguistic Interpretation of a GERman Corpus

Ebenen der Annotation

- PoS-Tagging: Wortklassen (Artikel, Nomen, attributives Adjektiv etc.)
- " Phrasenstruktur: Nominalphrasen, Verbalphrasen, etc.
- " Grammatiche Funktionen: Subjekt, Akkusativobjekt, etc.
- " *secondary edges*: grammatische Funktionen
- " Morphologische Information (Kasus, Numerus, etc.)



Ebenen der Annotation: Beispiel



aus Ihren Beiträgen:

Ein Überblick über verfügbare Ressourcen und Annotationsformate bei der TIGERTreebank zeigt, dass als Korpus für diese Datenbank vollständige Texte aus der Frankfurter Allgemeinen Zeitung genommen werden. Die annotierten Sätze werden weiterhin in einer MySQL Datenbank abgespeichert und Informationen in bestimmten Tabellen gesichert. Das Format der Ausgabe ist für den Menschen gut lesbar und die Maschine ebenso gut ausführbar in einer zeilen-orientierten und ASCII-basierten Format erstellt. Als weiteres Format wird TIGER XML genutzt und typischerweise in header und body unterteilt, wobei der header die Meta-Informationen über den Korpus beinhaltet, so dass eine Erklärung von tags für die Morphology, den POS und nicht-terminalistische Knoten und Kanten ausgegeben werden kann, und der body einen azyklischen Graphen, der als Datenmodell zum kodieren von linguistischen Annotationen dient. Als Hilfsmittel für eine graphische Oberfläche bei der Erstellung von Annotationen wird das tool Annotate genutzt. Ein weiteres Hilfsmittel zur Analyse von

Lemmata und Morphologischen Tags stellt das tool TIGERMorph als weitere graphische Benutzerschnittstelle zur Verfügung.

Für die syntaktische Annotation besitzt der TIGER Korpus zwei verschiedenen Methoden, und zwar die interaktive Annotation und das LFG Parsing Übungsblatt 8 Simone Winter 12.Dezember 2011 7

Im Vergleich der Syntaxform der TIGER Treebank mit der PennTreebank für die englische Sprache aus dem Originalpaper lässt sich feststellen, dass bei der PennTreebank die syntaktischen Kategorien nicht einheitlich erscheinen. Es fehlt bei der Form an Untergliederung bei den Tagsets und somit lässt sich die Annotation nicht so schön in verschiedenen Levels finden wie bei der TIGERTreebank. Weiterhin ist die Struktur recht unübersichtlich und sehr verschachtelt.

b)

Allgemein hin kann man zur Problematik syntaktischer Ressourcen und Annotationen sagen, dass es noch teilweise an einer befriedigenden Repräsentation für die Satzzergliederung fehlt. Schwierig stellt sich das Fehlen bestimmter Regeln gerade in der Nutzung von Adverbien heraus. Es gibt keine Regeln für Adverbien-Nutzung am Satzanfang oder in Kombination mit Verbalphrasen fehlen Regeln, die den Einsatz von Adverbien bestimmen. Schwierig ist auch immer noch der fehlende Gebrauch von Vergleichskorpora aus anderen Ressourcen zur Analyse, welche auch teilweise noch immer nur einen bestimmten Teil genau taggt und richtig analysiert.