

TIGER Morphologie-Annotationsschema

Berthold Crysmann	crysmann@dfki.de
Silvia Hansen-Schirra	hansen@coli.uni-sb.de
George Smith	smithg@rz.uni-potsdam.de
Dorothea Ziegler-Eisele	ziegler@coli.uni-sb.de

Projekt-Info: <http://www.ims.uni-stuttgart.de/projekte/TIGER/>

5.April 2005

DRAFT – Bitte nicht kopieren!

Projekt TIGER

Aufbau eines linguistisch interpretierten Korpus des Deutschen

Universität des Saarlandes

FR 8.7 Computerlinguistik

Universität Stuttgart

Institut für Maschinelle Sprachverarbeitung

Universität Potsdam

Institut für Germanistik



Inhaltsverzeichnis

1	Vorwort	4
2	Flexionskategorisierungen und ihre Werte	4
3	Nomina	5
4	Adjektive	10
5	Zahlen	11
6	Verben	12
7	Artikel	13
8	Pronomina	13
9	Adverbien	22
10	Konjunktionen	22
11	Adpositionen	22
12	Partikel	23
13	Sonstige	24
14	Kurzfassung zum Nachschlagen	26
A	Literatur	27

Inhaltsverzeichnis

1 Vorwort

Das vorliegende Schema dient der Morphologie-Annotation im Treebank-Projekt TIGER. Es ist als Zusatz zu den Richtlinien für das Part-of-Speech-Tagging und die Syntax-Annotation zu sehen, da es auf diesen Regeln aufbaut und sie als Basisinformation voraussetzt. Angestrebt wird eine deskriptive, theorie-neutrale Annotation.

Das Schema baut auf dem Stuttgart-Tübingen-Tagset (Thielen et al., 1999) auf. Es mussten allerdings auch einige Änderungen eingeführt werden, die sich durch Schwierigkeiten beim Annotieren ergaben. Hierbei wurde versucht, die Regeln den konkret auftretenden Problemen anzupassen, gleichzeitig aber die Konsistenz und Systematik mit den anderen Regeln zu wahren. Oft konnte keine Lösung gefunden werden, die allen Fällen gerecht wird. Das vorliegende Schema stellt daher einen Kompromiss dar. Einige strittige Punkte, zu denen (bis jetzt) keine einheitliche Regelung gefunden wurde, sind im Schema mit Anmerkungen, Fragen u.ä. festgehalten. Aus diesem Grund ist dieses Morphologie-Annotationsschema als vorläufige Fassung zu sehen, die keinen Anspruch auf Vollständigkeit erhebt.

Der Aufbau des vorliegenden Annotationsschemas richtet sich nach der Struktur des STTS.

2 Flexionskategorisierungen und ihre Werte

Folgende Tabelle zeigt die Attribute, die bei der morphologischen Annotation mit TigerMorph vergeben werden, die möglichen Werte, die sie annehmen können, und bei welchen PoS-Tags die Attribute annotiert werden. Grundlage dafür ist das STTS, doch sind einige Abweichungen erkennbar.

Attribut	mögliche Werte	verwendet bei
Genus	Masc, Fem, Neut	ADJA, ART, APPRART, NE, NN, PDS, PDAT, PIAT, PIS (teilweise), PPER, PPOSAT, PPOSS, PRELS, PRELAT, PWAT, PWS
Kasus	Nom, Gen, Dat, Acc	ADJA, ART, APPRART, NE, NN, PDAT, PDS, PIAT, PIS, PPER, PPOSAT, PPOSS, PRELS, PRELAT, PRF, PWAT, PWS
Numerus	Sg, Pl	ADJA, ART, APPRART, NE, NN, PDAT, PDS, PIAT, PIS, PPER, PPOSAT, PPOSS, PRELS, PRELAT, PRF, PWAT, PWS, V.FIN, V.IMP
Grad	Pos, Comp, Sup	ADJA, ADJD
Person	1, 2, 3	VVFIN, VAFIN, VMFIN, PPER, PRF
Tempus	Pres, Past	VVFIN, VAFIN, VMFIN
Modus	Ind, Subj	VVFIN, VAFIN, VMFIN
Nichtfinitheit	Inf, Psp, Imp, Infzu	VVINF, VAINF, VMINF, VVPP, VAPP, VMPP, VVIMP, VAIMP, VVIZU

Nicht annotiert werden mit TigerMorph die Attribute Flexion bei NN und ADJA und Definitheit bei ART. Priorität hat die Annotation aus dem morphosyntaktischen Kontext, d.h. wir versuchen immer, Ambiguitäten aus dem Kontext aufzulösen. Nur wenn es nicht gelingt, im gegebenen Kontext dem Attribut einen eindeutigen Wert zuzuweisen, soll der Wert * zugewiesen werden. Als Kontext wird für die Annotation nur der jeweilige Satz herangezogen.

3 Nomina

NN (Appellativa)

Morphologische Merkmale: Kasus, Numerus, Genus

Kasus, Numerus und Genus müssen immer angegeben werden. Pluraliatantum (Nomen, die keine Singularformen haben), z.B. *Kosten*, erhalten für das Merkmal Genus den Wert *, da das Genus hier nicht definiert ist. Ebenfalls den Wert * erhalten nominalisierte Adjektive und Partizipien im Plural, bei denen das Genus nicht entscheidbar ist wie z.B. *die Grünen, die Betrogenen, die Alten*. Einige NPs unterliegen aus den im nachfolgenden erklärten Gründen einer Sonderbehandlung.

1. Maßangaben

Maßeinheiten in Maßangaben wie z.B. *Liter, Meter, Kilogramm, Mark, Dollar, Blatt, Mann* werden hinsichtlich Kasus und Numerus wie folgt annotiert:

Kasus und Numerus sollen, wenn nicht eindeutig markiert, aus dem syntaktischen Kontext erschlossen werden. Wenn eine aus dem Kontext erwartbare Markierung fehlt, soll für Kasus und Numerus der Wert * vergeben werden. Steht nach der Maßeinheit eine Artangabe (meist ein Stoffsubstantiv wie *Kaffee, Stahl, Benzin*), wird diese als appositionelles Attribut zur Maßeinheit behandelt und erhält für Kasus den Wert Nom.

Beispiele:

ein Liter_{Nom.Sg.Masc} Benzin_{Nom.Sg.Neut}
 fünf Liter_{Nom.Pl.Masc} Bier_{Nom.Sg.Neut}
 Karl trinkt täglich 2 Liter_{Acc.Pl.Masc}
 die Höhe von 600 Kilometern_{Dat.Pl.Masc}
 die Höhe von 600 Kilometer_{*.*.Masc}
 Karl kauft 500 Blätter_{Acc.Pl.Neut} Schreibmaschinenpapier_{Nom.Sg.Neut}
 Karl kauft 500 Blatt_{*.*.Neut} Schreibmaschinenpapier_{Nom.Sg.Neut}
 11 Männer_{Nom.Pl.Masc} gingen leer aus
 11 Mann_{*.*.Masc} kamen durch
 Wer den Dollar_{Acc.Sg.Masc} nicht ehrt
 Die Dollars_{Nom.Pl.Masc} sitzen locker
 Die Kosten belaufen sich auf 50000 Dollar_{*.*.Masc}
 Gib mir mal meine 20 Mark_{*.*.Fem} Taschengeld_{Nom.Sg.Neut}

2. NPs mit engen Appositionen

Die Kasuszuweisung bei Nominalphrasen mit engen Appositionen gestaltet sich differenzierter, denn es tauchen hier innerhalb der Nominalphrasen unflektierte Formen auf, die auch als unveränderlicher Nominativ annotiert werden sollen.

(a) NP mit Artikel

den Bundeskanzler_{Acc. Sg. Masc} Helmut_{Nom. Sg. Masc} Kohl_{Nom. Sg. Masc}
des Bundeskanzlers_{Gen. Sg. Masc} Helmut_{Nom. Sg. Masc} Kohl_{Nom. Sg. Masc}
Helmut Kohl unveränderlich Nominativ Artikel und *Bundeskanzler* flektieren

(b) NP ohne Artikel

Kollegen erlebten Bundeskanzler_{Nom. Sg. Masc} Helmut_{Nom. Sg. Masc} Kohl_{Acc. Sg. Masc}
Bundeskanzler_{Nom. Sg. Masc} Helmut_{Nom. Sg. Masc} Kohls_{Gen. Sg. Masc} Abtritt
Helmut_{Nom. Sg. Masc} Kohls_{Gen. Sg. Masc} Auftritt
Bundeskanzler bzw. *Helmut* unveränderlich Nominativ *Kohl* flektiert

(c) NP aus reduziertem Kopulasatz

Wer kennt schon Helmut_{Acc. Sg. Masc} den_{Acc. Sg. Masc} Bundeskanzler_{Acc. Sg. Masc}
der Auftritt von Helmut_{Dat. Sg. Masc} dem_{Dat. Sg. Masc} Bundeskanzler_{Dat. Sg. Masc}
der Auftritt Helmut_{Gen. Sg. Masc} des_{Gen. Sg. Masc} Bundeskanzlers_{Gen. Sg. Masc}
Hier herrscht zwischen dem Kern der NP und dem appositivem Attribut Kasusidentität,
d.h. sowohl *Helmut* als auch *der Bundeskanzler* kongruieren.

(d) Maßangaben, die aus dem obsoleten Genitivus partitivus (*ein Glas Weines*) hervorgehen:
sechs Monate Gefängnis, zu *sechs Monaten Gefängnis verurteilt*, mit *Hilfe eines Glases Wein*.
Die Artangabe (*Gefängnis*, *Wein*) steht unveränderlich im Nominativ, die Maßangabe
(*sechs Monate* bzw. *ein Glas*) flektiert.

zu sechs Monaten_{Dat. Pl. Masc} Gefängnis_{Nom. Sg. Neut} verurteilt
mit Hilfe eines Glases_{Gen. Sg. Neut} Wein_{Nom. Sg. Masc}

3. Akkusative der Zeit

Als Akkusative der Zeit betrachten wir *das erste Mal*, *das erste Jahr*, *diesen Montag*, *einen Tag* (vor Neujahr) und analoge Fälle. Hier wird als Wert für Kasus ACC vergeben. Dagegen ist bei *Anfang/Mitte/Ende Januar* unklar, welchen Kasus die Einheiten tragen. Wir vergeben deshalb für *Anfang*, *Mitte*, *Ende* und auch für *Januar* den Kasus-Wert *.

Lemmatisierung: Als Grundform wird das Nomen im Nominativ Singular eingetragen (starke Form), außer bei Pluraliatantum (Nomen, die nur im Plural vorkommen). Hier steht als Lemma der Nominativ Plural.

Bei der Lemmatisierung von NNs erfassen wir drei Typen der Konversion, d.h. Substantive, die aus Infinitiven abgeleitet sind, solche aus Adjektiven und drittens Substantive, die aus Partizipien gebildet wurden. Bei den Substantiven aus Partizipien wollen wir noch die Unterscheidung nach Partizip 1 und Partizip 2 treffen. Hier wollen wir als Lemma eine verbale bzw. adjektivische Basis annotieren. Wir versuchen eine Lösung zu finden, die einerseits bei der Verwertung des Korpus den Zugriff auf die Nähe einer nominalen Form zum adjektivischen oder verbalen Paradigma ermöglicht, und andererseits wenig Abgrenzungsprobleme verursacht. Das Kriterium ist nicht semantische Identität, sondern Formgleichheit.

Für die Lemmatisierung der abgeleiteten NNs ergibt sich somit folgende Unterscheidung:

1. Substantive, die aus Adjektiven abgeleitet werden, bekommen die schwache Form des Adjektivs im Nominativ Singular Maskulin als Lemma. Das entscheidende Kriterium ist, dass es im gegenwärtigen Deutsch ein Adjektiv gleicher Form gibt.

Beispiele:

ein Alter Lemma: alte
ein Grüner Lemma: grüne
der Ältere Lemma: alte

2. Substantivierte Infinitive bekommen den verbalen Infinitiv als Lemma.

Beispiele:

das Lesen Lemma: lesen
das Fahren Lemma: fahren

3. Substantive, die aus einem Partizip 1 abgeleitet werden, erhalten die unflektierte Form des Partizips 1 als Lemma.

Beispiele:

der Lesende Lemma: lesend
der Fahrende Lemma: fahrend
der Vorsitzende Lemma: vorsitzend

4. Substantive, die aus einem Partizip 2 abgeleitet werden, erhalten als Lemma die starke Form des Partizips 2 (im Maskulin Singular).

Beispiele:

ein Betroffener Lemma: betroffener
der Angeklagte Lemma: angeklagter

Beispiele:

riefen die Demonstranten _{NN:Nom.Pl.Masc}	Lemma: Demonstrant
Stadtverordneter der Grünen _{NN:Gen.Pl.*}	Lemma: grüne
Premierminister wagt Wirtschaftsreform _{NN:Acc.Sg.Fem}	Lemma: Wirtschaftsreform
Hoffnungen der Betrogenen _{NN:Gen.Pl.*}	Lemma: betrogener
Hunderte Beamte _{NN:Nom.Pl.Masc}	Lemma: Beamter
stecken in den Staus _{NN:Dat.Pl.Masc}	Lemma: Stau
ein Blick _{NN:Nom.Sg.Masc} zeigt	Lemma: Blick
ein Darlehen in Höhe von 245 Mio. Dollar _{NN:*.*.Masc}	Lemma: Dollar
US-Präsident _{NN:Nom.Sg.Masc} George _{NE:Nom.Sg.Masc} Bush _{NE:Nom.Sg.Masc}	
des Finanzministers _{NN:Gen.Sg.Masc} Theo _{NE:Nom.Sg.Masc} Waigel _{NE:Nom.Sg.Masc}	
um die Hauptstadt _{NN:Acc.Sg.Fem} Lagos _{NE:Nom.Sg.Neut}	

NE (Eigennamen)

Morphologische Merkmale: Kasus, Numerus, Genus

Wir wollen so weit wie möglich die morphologischen Merkmale Kasus, Numerus und Genus aus dem syntaktischen Kontext erschließen. Falls ein Merkmal nicht bestimmbar ist, wird der Wert * vergeben.

Kasus: Die Bestimmung des Kasus wird häufig möglich sein, da NEs oder deren übergeordnete PNs meist eine Funktion im Satz bzw. in der Nominalphrase oder Präpositionalphrase haben, die mit einem bestimmten Kasus einhergeht.

Numerus: Für Numerus wird in der Regel Sg als Wert vergeben (Default-Regel). Wir gehen davon aus, dass Eigennamen singuläre Entitäten bezeichnen. Ausnahmen sind am syntaktischen Kontext erkennbar (Verbkongruenz, Pluralartikel).

Genus: Das Genus ist bei manchen Typen von Eigennamen eindeutig bestimmbar. Vornamen sind z.B. in der Regel eindeutig Masc oder Fem. Städte- und Ländernamen sind in der Regel eindeutig Neut (Ausnahmen wie *die Schweiz*, *die Türkei* sind an ihrem bestimmten Artikel zu erkennen). Dagegen kann alleinstehenden Familiennamen (z.B. *Kohl*), Firmenbezeichnungen ohne bestimmten Artikel (z.B. *Siemens*) sowie Eigennamen, die nur im Plural vorkommen (z.B. *die Niederlande*), kein eindeutiges Genus zugewiesen werden, sie erhalten somit für Genus den Wert *.

Komplexe Eigennamen:

Problematisch ist die morphologische Annotation bei komplexen, d.h. mehrgliedrigen, Eigennamen, die in der Regel Firmenbezeichnungen oder auch geographische Eigennamen sind. Hauptproblem ist der unklare Status fremdsprachlicher Einheiten mit Pos-Tag NE.

- Mehrgliedrige geographische Eigennamen

Bsp.: *New York*, *Tel Aviv*, *Bad Homburg*, *Santiago de Chile*

Kasus und Numerus sollen nur dort annotiert werden, wo auch eine Markierung zu sehen ist (Test: den Eigennamen als postnominalen Genitiv verwenden, z.B. *die Größe New Yorks*). In der Regel ist dies das letzte Glied (Kopf / Kern des Eigennamens), während die voranstehenden Glieder keine Flexion zeigen und dementsprechend den Wert * für Kasus und Numerus erhalten. Es ergibt sich also:

Beispiele:

*New*_{NE:*. . .*} *York*_{NE:Nom.Sg.Neut}
in *Tel*_{NE:*. . .*} *Aviv*_{NE:Dat.Sg.Neut}
die Einwohner *Bad*_{NE:*. . .Neut} *Homburgs*_{NE:Gen.Sg.Neut}

Die unterschiedliche Annotation von *New* in *New York* und *Bad* in *Bad Homburg* liegt darin begründet, dass *Bad* ein Substantiv des Deutschen ist, *New* aber nicht. Dieser Unterschied soll dadurch ausgedrückt werden, dass bei Erstem der jeweilige Genuswert annotiert wird, bei Zweitem wird dagegen auch für Genus der Wert * vergeben. Auch Eigennamen wie *Santiago*, *Beaujolais* usw. können Substantive des Deutschen sein und damit einen bestimmten Wert für Genus bekommen, auch wenn sie nicht als Kopf der Phrase fungieren. Maßgeblich für die Entscheidung, ob es sich um ein Substantiv des Deutschen handelt, sollte der Eintrag in ein repräsentatives Lexikon (z.B. Duden Deutsches Universalwörterbuch) sein.

Der Fall *Santiago de Chile* ist komplizierter, weil hier der erste Bestandteil den Kern des Eigennamens bildet (was der jeweilige Schreiber aber nicht mitvollziehen muss!) - man könnte also entweder *die Einwohner Santiago de Chiles* oder *die Einwohner Santiagos de Chile* finden bzw. bilden. Wir annotieren natürlich je nachdem, was wir vorfinden:

Beispiele:

die Einwohner *Santiago*_{NE:*. . .Neut} *de*_{NE:*. . .*} *Chiles*_{NE:Gen.Sg.Neut}
die Einwohner *Santiagos*_{NE:Gen.Sg.Neut} *de*_{NE:*. . .*} *Chile*_{NE:*. . .Neut}

Wenn nicht zu sehen ist, welchen Bestandteil der Schreiber als Kern/Kopf des Eigennamens angesehen hat, gilt als Default-Regel, dass der letzte Bestandteil als Kern/Kopf des Eigennamens anzusehen ist:

Beispiele:

in Santiago_{NE:*. *. *. Neut} de_{NE:*. *. *. *} Chile_{Dat. Sg. Neut}

- Mehrgliedrige Eigennamen von Firmen u.ä.

Bsp.: *Universal Pictures, Viking Penguin, Minnesota Mining und Manufacturing Company, Warner Bros, Harper's Magazine, Beaujolais nouveau*

Auch hier gilt, dass Kasus und Numerus nur dort annotiert werden, wo auch eine Markierung zu sehen ist. Dazu ist es wiederum nötig, Kern/Kopf des Eigennamens zu bestimmen. Einen Hinweis könnte der Genitiv-Test (siehe oben) geben, eventuell weist auch der Schreiber durch Gebrauch eines bestimmten Artikels darauf hin. Wiederum gilt: Wenn keine Hinweise gegeben sind, gilt als Default-Regel, dass der letzte Bestandteil als Kern/Kopf des Eigennamens anzusehen ist:

Beispiele:

das Hollywood-Studio Universal_{NE:*. *. *. *} Pictures_{NE:Nom. Sg. *}
mit Viking_{NE:*. *. *. *} Penguin_{NE:Dat. Sg. *}, einem US-Verlag
mit der amerikanischen Minnesota_{NE:*. *. *. Neut} Mining_{NE:*. *. *. *} und_{NE:*. *. *. *}
Manufacturing_{NE:*. *. *. *} Company_{Dat. Sg. Fem}
mit Warner_{NE:*. *. *. *} Bros_{NE:Dat. Sg. *} handelseinig
in Harper's_{NE:*. *. *. *} Magazine_{NE:Dat. Sg. *}

Am Beispiel *Beaujolais nouveau* wird deutlich: Insbesondere bei fremdsprachlichen Eigennamen aus vertrauten oder eng verwandten Sprachen ist manchmal offensichtlich, dass nicht der letzte Bestandteil Kern/Kopf des Eigennamens ist. In solchem Fall sollte der Bestandteil, der eindeutig Kern/Kopf des Eigennamens ist, auch als solcher annotiert werden:

Beispiele:

Beaujolais_{NE:Nom. Sg. Masc} nouveau_{NE:*. *. *. *} des Beaujolais'_{NE:Gen. Sg. Masc} nouveau_{NE:*. *. *. *}

Das Beispiel *Beaujolais nouveau*, bei dem der erste Bestandteil eindeutig Kern/Kopf des Eigennamens ist, leuchtet ein, weil *Beaujolais* schon Eingang in das deutsche Lexikon gefunden hat. Andere Beispiele mögen weniger eindeutig sein, was die Frage nach Kern/Kopf betrifft:

Beispiele:

das Kölner Unternehmen Gastro_{NE:Nom. Sg. *} Unlimited_{NE:*. *. *. *} oder
das Kölner Unternehmen Gastro_{NE:*. *. *. *} Unlimited_{NE:Nom. Sg. *}

Für die erste Annotation spräche, dass *Unlimited* als nachgestelltes Attribut erkannt wird. Andererseits ist aber der Status von *Gastro* zweifelhaft. Für diesen und andere Zweifelsfälle gilt, dass im Zweifel die Default-Regel in Kraft tritt, d.h. dass der letzte Bestandteil eines mehrgliedrigen Eigennamens als dessen Kern/Kopf angesehen wird. Im Beispiel *Gastro Unlimited* wäre die zweite Annotation also richtig.

Lemmatisierung:

Als Lemma wird die Form des Nominativ Singular eingetragen. Da die meisten Eigennamen höchstens im Genitiv eine Flexionsendung haben, ist dies meist auch die Oberflächenform. Bei Eigennamen, die keine Singularform haben, wird die Form des Nominativ Plural eingetragen.

Beispiele:

Konzernchefs in den USA _{NE:Dat.Pl.*}	Lemma: USA
Maiers _{NE:Nom.Pl.*} kommen	Lemma: Maier
Sabines _{NE:Gen.Sg.Fem} Chef	Lemma: Sabine
dem Unternehmen in Ohio _{NE:Dat.Sg.Neut}	Lemma: Ohio
nach Angaben der Menschenrechtskomitees CDF _{NE:Nom.*.*}	Lemma: CDF
auf den Philippinen _{NE:Dat.Pl.*}	Lemma: Philippinen
Der Kohl _{NE:Nom.Sg.Masc} liebt Saumagen	Lemma: Kohl
Die Kohl _{NE:Nom.Sg.Fem} liebt Sekt	Lemma: Kohl
Kohl _{NE:Nom.Sg.Masc} hat versagt, was er aber nicht zugeben will	Lemma: Kohl
Kohl _{NE:Nom.Sg.*} reiste ab	Lemma: Kohl

FM (fremdsprachliches Material)

Hier handelt es sich um eine nicht sonderlich gut abgegrenzte Klasse. Durch das Tag FM ist schon deutlich: dieses Vorkommnis ist keine Form des Deutschen. Weiter als das analysieren wir nicht. FM erhält folgerichtig keine morphologische Annotation und kein Lemma.

Beispiele:

wünschten Ross for President_{FM:--} Lemma: --

4 Adjektive

ADJD (adverbiales oder prädikatives Adjektiv)

Morphologische Merkmale: Grad

Hier wird nur für das Merkmal Grad ein Wert eingetragen, alle anderen Merkmale werden nicht annotiert. Das Merkmal Grad kann immer eindeutig bestimmt werden, der Wert * kommt hier nicht vor.

Lemmatisierung: als Lemma wird der Positiv eingetragen. Bei Partizipien aus Verben, die adjektivisch gebraucht werden, steht als Lemma die Kurzform des Partizips.

Beispiele:

ein Unternehmen erfolgreich _{ADJD:Pos} leiten	Lemma: erfolgreich
Probleme sind doch weit größer _{ADJD:Comp}	Lemma: groß
Probleme sind gravierender _{ADJD:Comp} als gemeinhin angenommen	Lemma: gravierend

ADJA (attributives Adjektiv)

Morphologische Merkmale: Grad, Kasus, Numerus, Genus

Das Flexionssystem des attributiv verwendeten Adjektivs ist komplex, und umfasst neben der Deklination nach Kasus, Numerus und Genus auch die Komparation. Wir folgen hierbei dem STTS und behandeln die Komparation als Flexion nach Steigerungsgrad. Ein Wert für das Merkmal Grad muss immer angegeben werden, der Wert * kommt nicht vor. Die Bestimmung richtet sich nach der Erscheinungsform, nicht nach der Semantik des Adjektivs.

Adjektive deklinieren nach verschiedenen Flexionsmustern in verschiedenen Kontexten. Deswegen und auch wegen zahlreicher Synkretismen sind wir in besonderer Weise auf den syntaktischen Kontext angewiesen, um Werte für Kasus, Numerus und Genus zu bestimmen. Dies ist angemessen, denn in der Deklination des Adjektivs werden Kongruenzbeziehungen ausgedrückt. Wir streben an, bei deklinierenden Wörtern Werte für die Merkmale Kasus, Numerus und Genus zu annotieren, sofern dies im gegebenen syntaktischen Kontext möglich ist. Wenn für einzelne Merkmale kein Wert bestimmt werden kann, wird der Wert * annotiert.

Wir weichen bewusst vom STTS ab, indem wir keine Werte für das Merkmal Flexion vergeben. Wir sehen das System der Deklination des Adjektivs im syntaktischen Kontext als Gegenstand empirischer Untersuchung und wollen nicht von vornherein bestimmen, welches Flexionsparadigma in welchem Kontext vorkommt.

Für Adjektive, die nicht deklinieren können oder müssen, wie z.B. *rosa*, *lila*, *sexy* usw., wird als Wert für Kasus, Numerus und Genus * annotiert, sie erhalten lediglich einen Wert für den Steigerungsgrad (POS). Stadtadjektive wie *Berliner* und verwandte Herkunftsbezeichnungen wie *Schweizer* sowie *siebziger* (Jahre) usw. deklinieren ebenfalls nicht und werden auf gleiche Weise behandelt. An dieser Stelle wird darauf hingewiesen, dass die Endung *-er* weder als Deklinationsendung noch als Komparationsendung zu deuten ist.

Lemmatisierung: Als Lemma wird der Positiv bei prädikativem Gebrauch (unflektierte Form) des Adjektivs eingetragen. Hat ein Adjektiv diese Kurzform nicht (z.B. *der zweite Wagen*), setzen wir die starke Form des Nominativ Singular Maskulinum als Lemma ein. Bei Herkunftsbezeichnungen wie *Berliner Sonntagszeitung* steht als Lemma *Berliner*. In Fällen, wo wir nur schwer entscheiden können, ob ein Lexem wie z.B. *früher* oder *weiter* kompariert wird oder nicht, setzen wir stets *früh* und *weit* als Lemma ein und annotieren das Adjektiv als kompariert.

Für ein als ADJA verwendetes Partizip II wird als Lemma die Kurzform des Partizips eingetragen.

Beispiele:

prächtiger _{ADJA:POS.Nom.Sg.Masc}	Sonnenschein	Lemma:prächtigt
texanische _{ADJA:POS.Nom.Sg.Masc}	Milliardär	Lemma: texanisch
um die größte _{ADJA:Sup.Acc.Sg.Fem}	Volkswirtschaft	Lemma: groß
ein optimales _{ADJA:POS.Nom.Sg.Neut}	Konzept	Lemma: optimal
der zweite _{POS.Nom.Sg.Masc}	Punkt	Lemma: zweiter
im lila _{ADJA:POS.*.*.*}	Kleid	Lemma: lila
mit überzeugten _{ADJA:POS.Dat.Pl.Masc}	Neonazis	Lemma: überzeugt
für die Dritte _{ADJA:POS.Acc.Sg.Fem}	Welt	Lemma: dritter
ein weiteres _{ADJA:Comp.Nom.Sg.Neut}	Anzeichen sei	Lemma: weit
während früherer _{ADJA:Comp.Gen.Pl.Masc}	Streiks	Lemma: früh
das Berliner _{ADJA:POS.*.*.*}	Haushaltsloch	Lemma: Berliner
in den siebziger _{ADJA:POS.*.*.*}	Jahren	Lemma: siebziger

5 Zahlen

CARD Bei Cardinalzahlen erfolgt keine morphologische Annotation.

Ordinalzahlen erhalten das POS-tag ADJA und werden auch als solche behandelt. Ebenfalls unter

die (Zahl-)Adjektive fallen Ausdrücke wie *auf halbem Weg, mit Ausnahme zweier Berliner*. In Ausdrücken wie *mit einem Kollegen* erhält das PoS-Tag ART und wird folglich morphologisch auch wie ein Artikel behandelt.

Lemmatisierung: Als Lemma wird die Nennform eingetragen. In den meisten Fällen ist dies auch die Oberflächenform. Ausnahmen sind z.B. *zweier, dreier* usw.

Beispiele:

500 _{Card:--} Großunternehmen	Lemma: 500
1962 _{Card:--}	Lemma: 1962
zwei _{Card:--} Jahren	Lemma: zwei

6 Verben

VVFIN, VAFIN, VMFIN (finites Verb voll, aux oder modal)

Morphologische Merkmale: Person, Numerus, Tempus, Modus

Alle Attribute müssen angegeben werden. Synkrete Formen werden soweit wie möglich anhand des syntaktischen Kontexts disambiguiert. Wenn Oberflächenform und syntaktischer Kontext nicht ausreichen, um zwischen Indikativ und Konjunktiv zu unterscheiden, wird Indikativ als Default annotiert.

Lemmatisierung: als Lemma wird der Infinitiv eingetragen

Beispiele:

Ob diese Fähigkeiten ausreichen _{VVFIN:3.Pl.Pres.Ind}	Lemma: ausreichen
In Indien kann _{VMFIN:3.Sg.Pres.Ind} der Mittelstand	Lemma: können

VVIMP, VAIMP (Imperativ voll oder aux)

Morphologische Merkmale: Person, Numerus, Modus

Das Attribut Person erhält immer den Wert 2 und das Attribut Modus immer den Wert Imp. Für das Attribut Numerus werden die Werte Sg oder Pl vergeben. Pragmatische Imperative (Höflichkeitsformen) sehen wir nicht als Imperative an.

Lemmatisierung: als Lemma wird der Infinitiv eingetragen

Beispiele:

geh _{VVIMP:2.Sg.Imp} doch nach draußen	Lemma: gehen
Laßt _{VVIMP:2.Pl.Imp} Honecker in Frieden	Lemma: lassen

Aber:

schauen _{VVIMP:3.Pl.Pres.Ind} Sie	Lemma: schauen
--	----------------

VVINF, VVIZU, VAINF, VMINF (Infinitiv voll, mit *zu*, aux oder modal)

Hier wird nur der Wert Inf bzw. Inf zu bei Verben mit eingegliedertem *zu* eingetragen.

Lemmatisierung: Als Lemma wird der Infinitiv (ohne *zu*) eingetragen.

Beispiele:

es gilt Computer abzusetzen_{VVIZU: Inf zu} Lemma: absetzen
die Wahl zu gewinnen_{VVINFIN: Inf} Lemma: gewinnen

VVPP, VMPP, VAPP (Partizip Perfekt voll, aux oder modal)

Hier wird nur der Wert *PSP* für *past participle* eingetragen.

Lemmatisierung: Als Lemma steht der Infinitiv.

Beispiele:

davon sind selbst seine Kritiker überzeugt_{VVPP: PSP} Lemma: überzeugen
seien gestrichen worden_{VAPP: PSP} Lemma: werden
nicht geäußert _{VVPP: PSP} hat Lemma: äußern

Anmerkung: Das Partizip Präsens annotieren wir je nach morphosyntaktischer Funktion. In den meisten Fällen bedeutet das eine Annotation wie ein Adjektiv.

Beispiele:

etwas stärker konjunkturstimulierend_{ADJD: POS} wirken können Lemma: konjunkturstimulierend
mit dem amtierenden_{ADJA: POS. Dat. Sg. Masc} Präsidenten Lemma: amtierend

7 Artikel

ART (unbestimmter und bestimmter Artikel)

Morphologische Merkmale Kasus, Numerus, Genus

Kasus, Numerus und Genus werden stets angegeben, auch wenn sie teilweise nur aus dem Kontext erschlossen werden können. Nicht annotiert wird im Unterschied zu STTS das Attribut Definitheit.

Lemmatisierung: als Lemma wird die Zitierform eingetragen, also der Nominativ Maskulin Singular.

Beispiele:

er wäre ein_{ART: Nom. Sg. Masc} prächtiger Diktator Lemma: ein
die_{ART: Nom. Pl. Masc} Konzernchefs halten nicht viel von Lemma: der
sie lehnen den_{ART: Acc. Sg. Masc} Milliardär ab Lemma: der

8 Pronomina

PPER (irreflexibles Personalpronomen)

Morphologische Merkmale: Person, Kasus, Numerus, Genus

Person, Kasus, Numerus, Genus werden immer zugewiesen. Das Genus muss bis auf die Formen der 3.P.Sg. aus dem Satzkontext erschlossen werden. Ist das nicht möglich, erhält das Attribut Genus den Wert *.

Unpersönliches *es* in Wendungen wie:

es gilt...

es handelt sich...

es ist schwer zu sagen...

wird aus dem Kontext nach seinen Kongruenzeigenschaften disambiguiert. Das Attribut Kasus erhält dabei seinen Wert aus dem Kontext, die Werte für die Attribute Person, Numerus und Genus bleiben stets konstant. So erhält das Attribut Person stets den Wert 3, Numerus stets den Wert Sg und Genus immer den Wert Neut.

Lemmatisierung: Als Lemma wird die Oberflächenform eingetragen

Beispiele:

ich _{PPER:1.Nom.Sg.*} glaube kaum	Lemma: ich
machen ihnen _{PPER:3.Dat.Pl.*} besonders zu schaffen	Lemma: ihnen
es _{PPER:3.Nom.Sg.Neut} ist wirklich schwer zu sagen	Lemma: es

PRF (reflexives Personalpronomen)

Morphologische Merkmale: Person, Kasus, Numerus

Person und Numerus müssen immer angegeben werden. Beim Attribut Kasus gibt es nur die Werte Gen, Acc und Dat.

Bei unpersönlichen reflexiven Verben bekommt *sich* beim Attribut Person den Wert 3, bei Kasus den Wert * und bei Numerus den Wert Sg zugewiesen.

Lemmatisierung: Als Lemma wird für alle (*mir, dir, sich, einander, uns, euch, meiner, deiner...*) *sich* eingetragen.

Beispiele:

Zwar können sich _{PRF:3.Dat.Pl} die Topmanager einen Plan vorstellen	Lemma: sich
während 78 Prozent sich _{PRF:3.Acc.Pl} für Bush aussprachen	Lemma: sich
handelt es sich _{PRF:3.Acc.Sg}	Lemma: sich
wenn man sich _{PRF:3.Acc.Sg} mit der Vergangenheit befasst	Lemma: sich

PPOSAT, PPOSS (attribuierendes und substituierendes Possessivpronomen)

Morphologische Merkmale: Kasus, Numerus, Genus

Kasus und Numerus müssen immer angegeben werden. Sie richten sich nicht nach dem Besitzer, sondern nach dem nachfolgenden Nomen. Das Merkmal Genus erhält ebenfalls den Wert des nachfolgenden Nomens, außer in Fällen, in denen das Genus nicht definiert ist. Dann erhält das Pronomen den Wert * für das Attribut Genus. Das ist z.B. der Fall bei manchen Pluralformen.

Lemmatisierung: Als Lemma steht der Nominativ Singular Maskulin.

Beispiele:

mit deinen _{PPOSAT:Dat.Pl.Fem} Sachen	Lemma: dein
von seinen _{PPOSAT:Dat.Pl.*} Beschäftigten verlange er	Lemma: sein
gibt Auskunft über seine _{PPOSAT:Acc.Sg.Fem} Wirtschaftspolitik	Lemma: sein
das sei nicht seiner _{PPOSS:Nom.Sg.Masc}	Lemma: seiner
meinem _{PPOSS:Dat.Sg.*} hat es auch nicht geschadet	Lemma: meiner

PDAT, PDS (attribuierendes und substituierendes Demonstrativpronomen)

Morphologische Merkmale: Kasus, Numerus, Genus

Kasus und Numerus müssen immer angegeben werden. Bei PDAT und noch seltener bei PDS kann der Wert für das Attribut Genus der Pluralformen dem syntaktischen Kontext entnommen werden. Falls nicht, annotieren wir den Wert *. Achtung: Die Formen *deren* und *dessen* sollen immer als PDS und syntaktisch als Genitivattribut (AG) annotiert werden (es sei denn, sie fungieren als Relativpronomen, siehe dazu Abschnitt PRELAT). Dementsprechend erhalten sie für Kasus immer den Wert Gen. Die Form *dessen* erhält für Numerus immer den Wert Sg. Numerus und Genus von *deren* bzw. Genus von *dessen* müssen aus dem Satzkontext erschlossen werden oder erhalten dafür den Wert *.

Lemmatisierung: Als Lemma steht der Nominativ Singular Maskulin.

Beispiele:

diese _{PDAT:Nom.Pl.Fem} Fähigkeiten verhelfen ihnen	Lemma: dieser
dessen _{PDS:Gen.Sg.*} Fähigkeiten verhelfen ihnen	Lemma der
aber das _{PDS:Nom.Sg.Neut} ist nicht unser System	Lemma: der
sachlich begründete Fragen - die _{PDS:Acc.Pl.Fem} gibt es durchaus -	Lemma: der
auf deren _{PDS:Gen.*.*} Fragen gibt es keine Antwort	Lemma: der
die Frau ist verrückt, auf deren _{PDS:Gen.Sg.Fem} Tour falle ich nicht mehr rein	Lemma: der

PIAT, PIS (attribuierendes und substituierendes Indefinitpronomen)

Morphologische Merkmale: Kasus, Numerus, Genus

Die Klasse der Indefinitpronomina zeigt ein sehr uneinheitliches Flexionsverhalten.

Eine Teilklasse von ihnen ist nichtflektierbar, d.h. sie besitzen nur eine Form in ihrem Wortparadigma (wie z.B. die auf *-lei* endenden *mancherlei*, *keinerlei* usw. oder *mehr*, *nichts*). Diese nichtflektierbaren Einheiten erhalten für Kasus und Numerus den Wert *, das Genus ist in einigen Fällen feststellbar (z.B. *nichts* - Neut).

Die nichtflektierbare Form *man* wird immer mit Nom.Sg.* annotiert.

Bei den Indefinitpronomina *viel* und *wenig* sind zwei homonyme Formen zu unterscheiden, eine nichtflektierbare und eine flektierbare: *vieles* vs. *viel Papier wurde verschwendet*. Die nichtflektierbare Form erhält deshalb das Morph-Tag *.*.* und das Lemma *viel*, die flektierbare Form für Kasus, Numerus und Genus die jeweilig entsprechenden Werte und wie alle flektierbaren Pronomina als Lemma die (starke) Form des Nom.Sg.Masc *vieler*.

Die Indefinitpronomina *solcher*, *mancher*, *aller* besitzen die Kurzformen *solch*, *manch*, *all*. Diese werden wie die nichtflektierbaren Indefinitpronomina behandelt.

Lemmatisierung: Als Lemma steht der Nominativ Singular Maskulin beim artikellosen Gebrauch. Existiert keine Singularform, wird der Nominativ Plural Maskulin als Lemma eingetragen, z.B. *beide, mehrere, viele*.

In den folgenden Tabellen ist die Annotation für jedes Indefinitpronomen, getrennt nach seinem jeweiligen Vorkommen als PIAT oder PIS, eingetragen. Unter jedem Lemma steht dabei jeweils ein morphologisch annotiertes Beispiel.

Liste zur morphologischen Annotation der Indefinitpronomina

Flektierbar

PIAT

aller

alles_{PIAT:Nom.Sg.Neut} Gute

beide

beiden_{PIAT:Dat.Pl.Neut} Kindern

einiger

einiges_{PIAT:Nom.Sg.Neut} Bier

etlicher

etlicher_{PIAT:Nom.Sg.Masc} Unsinn

irgendein

irgendein_{PIAT:Nom.Sg.Masc} Hund

irgendwelcher

irgendwelcher_{PIAT:Nom.Sg.Masc} Mist

jeder

jedem_{PIAT:Dat.Sg.Neut} Kind

jedweder

jedwedem_{PIAT:Gen.Sg.Masc} Erfolgs

jeglicher

jegliches_{PIAT:Nom.Sg.Neut} Geschäft

kein

kein_{PIAT:Nom.Sg.Neut} Brot

PIS

alle

alles_{PIS:Nom.Sg.Neut} ist hin

beide

Beides_{PIS:Nom.Sg.Neut} gefällt uns

einer

Der einen_{PIS:Dat.Sg.Fem} traue ich

einige

Es gefiel so einigen_{PIS:Dat.Pl.*}

etlicher

Etlicher_{PIS:Nom.Sg.Masc} wird so alt

irgendeiner

Irgendeiner_{PIS:Nom.Sg.Masc} kam rein

irgendwelcher

Nimm irgendwelchen_{PIS:Acc.Sg.Masc}

irgendwer

Irgendwem_{PIS:Dat.Sg.Masc} kann man immer helfen.

jeder

Jede_{PIS:Nom.Sg.Fem} ist willkommen

jedweder

Jedweder_{PIS:Nom.Sg.Masc} kann bleiben

jeglicher

Jegliches_{PIS:Nom.Sg.Neut} braucht seine Zeit

jemand

Jemand_{PIS:Nom.Sg.Masc} sollte es tun

keiner

Glaube einfach keinem_{PIS:Dat.Sg.*}

mancher
manchem_{PIAT:Dat.Sg.Masc} Mann
mehrere
mehreren_{PIAT:Dat.Pl.Neut} Kindern
meister
der meiste_{PIAT:Nom.Sg.Masc} Müll

sämtlicher
sämtliches_{PIAT:Nom.Sg.Neut} Gerümpel

solcher
solcher_{PIAT:Nom.Sg.Masc} Schmerz

ebensolcher
ebensolcher_{PIAT:Nom.Sg.Masc} Schmerz

vieler
vieler_{PIAT:Nom.Sg.Masc} Kaffee

ebensovieler
ebensovieler_{PIAT:Nom.Sg.Masc} Kaffee

zuvieler
zuvieler_{PIAT:Nom.Sg.Masc} Unsinn

weniger
nur weniger_{PIAT:Nom.Sg.Masc} Wein

wenigster
wenigster_{PIAT:Nom.Sg.Masc} Wein

allerwenigster
allerwenigster_{PIAT:Nom.Sg.Masc} Müll

Nicht-Flektierbar

PIAT

all
all_{PIAT:*. *.* *} sein Glück

allerhand
allerhand_{PIAT:*. *.* *} Kram

mancher
Er hat manche_{PIS:Acc. *.* *} gesehen

mehrere
Mehrere_{PIS:Nom.Pl. *} fehlten

meister
Die meisten_{PIS:Nom.Pl. *} sind gut
Sie hat am meisten_{PIS:*. *.* *}

niemand
Ich bin niemandem_{PIS:Dat.Sg.Masc} böse

solcher
Solche_{PIS:Acc.Pl. *} brauchen wir

ebensolcher
Ebensolchen_{PIS:Acc.Sg.Masc} will ich

viele
Erna kauft viele_{PIS:Acc.Pl. *}

ebensoviele
Hilde klaut ebensoviele_{PIS:Acc.Pl. *}

zuvieler
Karin kriegt zuviele_{PIS:Acc.Pl. *}

weniger
Nur weniger_{PIS:Nom.Sg.Masc} ist noch da

wenigster
Nimm die wenigsten_{PIS:Acc.Pl. *}

allerwenigster
Gib das allerwenigste_{PIS:Acc.Sg.Neut}

PIS

allerhand
Das ist ja allerhand_{PIS:*. *.* *}

allerlei
allerlei_{PIAT:***} Krimskrams

beiderlei
beiderlei_{PIAT:***} Geschlechts

(ein) bißchen
das bißchen_{PIAT:***} Haushalt

derlei
derlei_{PIAT:***} Schabernack

etwas
etwas_{PIAT:***} Zeit

was
was_{PIAT:***.Neut} Nettos

genug
genug_{PIAT:***} Geld

genügend
genügend_{PIAT:***} Beweise

irgendetwas
irgendetwas_{PIAT:***.Neut} Saures

irgendwas
irgendwas_{PIAT:***.Neut} Süßes

lauter
lauter_{PIAT:***} ehrenwerte Leute

manch
manch_{PIAT:***} einer

mancherlei
mancherlei_{PIAT:***} Arten

mehr
mehr_{PIAT:***} Geld

nichts
von nichts_{PIAT:***.Neut} anderem

allerlei
Dort gibt es allerlei_{PIS:***} zu kaufen

beiderlei
Wir brauchen beiderlei_{PIS:***}

(ein) bißchen
Wartet noch ein_{Art:***} bißchen_{PIS:***}

derlei
Derlei_{PIS:***} ist nicht zu gebrauchen

etwas
Er hat etwas_{PIS:***.Neut} gehört

was
Wir müssen was_{PIS:***.Neut} tun

genug
Es gibt genug_{PIS:***} zu tun

ihresgleichen
Die Show sucht ihresgleichen_{PIS:***}

irgendetwas
Sie suchen nicht irgendetwas_{PIS:***.Neut}

irgendwas
Geben Sie mir irgendwas_{PIS:***.Neut}

jedermann
Jedermann_{PIS:***} ist aufgefordert

man
Man_{PIS:Nom.Sg.*} sieht sich.

mancherlei
Es wird so mancherlei_{PIS:***} vermutet

mehr
Es wird immer mehr_{PIS:***} verlangt

nichts
Sie wollte nichts_{PIS:***.Neut}

nix
 von nix_{PIAT:*. *.Neut} anderem
 (ein) paar
 die paar_{PIAT:*. *. *.} Probleme
 solch
 solch_{PIAT:*. *. *.} ein Ärger
 solcherlei
 solcherlei_{PIAT:*. *. *.} Methoden
 soetwas
 soetwas_{PIAT:*. *.Neut} Gemeines
 sowas
 sowas_{PIAT:*. *.Neut} Komisches
 viel
 viel_{PIAT:*. *. *.} Knete
 ebensoviel
 ebensoviel_{PIAT:*. *. *.} Knete
 zuviel
 zuviel_{PIAT:*. *. *.} Knete
 vielerlei
 vielerlei_{PIAT:*. *. *.} Spielarten
 wenig
 wenig_{PIAT:*. *. *.} Aussagekraft
 ein wenig
 ein_{Art:*. *. *.} wenig_{PIAT:*. *. *.} Brot
 weniger
 weniger_{PIAT:*. *. *.} Zeit als gestern
 zweierlei
 zweierlei_{PIAT:*. *. *.} Maß

nix
 Er hat nix_{PIS:*. *.Neut}
 (ein) paar
 Davon gibt es schon ein_{Art:*. *. *.} paar_{PIS:*. *. *.}
 solcherlei
 Solcherlei_{PIS:*. *. *.} wurde oft berichtet
 soetwas
 Soetwas_{PIS:*. *.Neut} sagt man nicht
 sowas
 Sowas_{PIS:*. *.Neut} gehört sich nicht
 viel
 Das ist ganz schön viel_{PIS:*. *. *.}
 ebensoviel
 Er wollte ebensoviel_{PIS:*. *. *.}
 zuviel
 Sie wollte zuviel_{PIS:*. *. *.} auf einmal
 vielerlei
 Vielerlei_{PIS:*. *. *.} gab es zu sehen
 wenig
 Sie gab immer wenig_{PIS:*. *. *.}
 ein wenig
 Er schläft jetzt ein_{Art:*. *. *.} wenig_{PIS:*. *. *.}
 weniger
 Heute gibt es noch weniger_{PIS:*. *. *.}
 zweierlei
 Zweierlei_{PIS:*. *. *.} ist zu beachten

Die Indefinitpronomina *anderer*, *reichlicher*, *übriger*, *einzelner*, *einziger*, *erster*, *ersterer*, *letzter*, *letzterer*, *weiterer* sollen wie folgt annotiert werden:

1. Die Formen von *übrig* sollen immer als Adjektive angesehen werden.
2. *reichlich* soll nur dann als PIAT annotiert werden, wenn es unflektiert direkt vor einem Substantiv steht.

Beispiele:

reichlich _{ADJD:POS}	vorhanden	Lemma: reichlich
ein reichliches _{ADJA:POS.Nom.Sg.Neut}	Mahl	Lemma: reichlich
reichlich _{PIAT:*. *.*}	Fernwärme	Lemma: reichlich

3. Die Indefinitpronomina *anderer, einzelner, einziger, erster, ersterer, letzter, letzterer, weiterer* sollen einheitlich wie folgt behandelt werden:

In attributiver Funktion vor einem Nomen sollen sie als ADJA getaggt werden, in substituierender Funktion (ohne Bezugsnomen) dagegen als PIS.

Beispiele:

ein anderer _{ADJA:POS.Nom.Sg.Masc}	Manager	Lemma: anderer
unter anderem _{PIS:Dat.Sg.Neut}		Lemma: anderer
gegen Erich Honecker und andere _{PIS:Acc.Pl.*}		Lemma: anderer
Die anderen _{PIS:Nom.Pl.*}	sind schuld	Lemma: anderer
in einzelnen _{ADJA:POS.Dat.Pl.Masc}	Bereichen	Lemma: einzelner
im einzelnen _{PIS:Dat.Sg.Neut}		Lemma: einzelner
sein einziger _{ADJA:POS.Nom.Sg.Masc}	Sohn	Lemma: einziger
der einzige _{PIS:Nom.Sg.Masc}	der	Lemma: einziger
deren ersten _{ADJA:POS.Acc.Sg.Masc}	Entwurf	Lemma: erster
nur wer als erster _{PIS:Nom.Sg.Masc}		Lemma: erster
im ersteren _{ADJA:Comp.Dat.Sg.Masc}	Fall	Lemma: erster
was erstere _{PIS:Acc.Pl.*}	angeht	Lemma: ersterer
in letzter _{ADJA:POS.Dat.Sg.Fem}	Minute	Lemma: letzter
das ist das letzte _{PIS:Nom.Sg.Neut}		Lemma: letzter
das letztere _{ADJA:Comp.Nom.Sg.Neut}	Thema	Lemma: letzter
letzteren _{PIS:Dat.Pl.*}	gegenüber	Lemma: letzterer
jede weitere _{ADJA:Comp.Nom.Sg.Fem}	Offenheit	Lemma: weit
sechs weitere _{PIS:Nom.Pl.*}		Lemma: weiterer

Diese Regelung geht teilweise mit dem STTS konform (*anderer, erster*), teilweise weicht sie vom STTS ab (*ersterer, letzterer* stehen dort unter der Rubrik PIAT oder PIS). Die anderen Fälle (*einzelner, einziger, letzter, weiterer* sowie *reichlich*) werden im STTS nicht erwähnt.

PRELS (substituierendes Relativpronomen)

Morphologische Merkmale: Kasus, Numerus, Genus

Kasus und Numerus werden stets angegeben; Genus, soweit aus dem Kontext erschließbar. In unentscheidbaren Fällen erhält Genus den Wert *.

Lemmatisierung: Als Lemma wird die Zitierform eingetragen, also der Nominativ Maskulin Singular.

Beispiele:

Geschäftsführer, der _{PRELS:Nom.Sg.Masc}		Lemma: der
Unternehmer, die _{PRELS:Nom.Pl.Masc}	meinen	Lemma: der
Bruch... vor dem _{PRELS:Dat.Sg.Masc}	noch jeder	Lemma: der

PRELAT (attribuierendes Relativpronomen)

Morphologische Merkmale: Kasus, Numerus, Genus

Als PRELAT kommen nur die Formen *deren* und *dessen* vor. Diese sind Genitivattribute in der jeweiligen NP - Kasus, Numerus und Genus richten sich nicht nach dem nachfolgenden Nomen! Für Kasus wird immer der Wert Gen vergeben, Numerus und Genus richten sich nach dem Nomen, auf das sich der Relativsatz bezieht.

Lemmatisierung: Als Lemma wird der Nominativ Maskulin Singular eingetragen.

Beispiele:

das Jahr, in dessen_{PRELAT:Gen.Sg.Neut} Verlauf Lemma: der
die Orte, deren_{PRELAT:Gen.Pl.Masc} Hotels weniger Gäste meldeten Lemma: der

PWAT (attribuierendes Interrogativpronomen)

Morphologische Merkmale: Kasus, Numerus, Genus

Kasus und Numerus müssen stets angegeben werden, Genus nur, soweit kontextuell erschließbar. Für unentscheidbare Fälle erhält Genus den Wert *.

Lemmatisierung: Als Lemma steht der Nominativ Singular Maskulin.

Beispiele:

welche_{PWAT:Acc.Pl.Fem} Positionen er einnimmt Lemma: welcher
in welcher_{PWAT:Dat.Sg.Fem} Art Soldaten eingesetzt werden Lemma: welcher

PWS (substituierendes Interrogativpronomen)

Morphologische Merkmale: Kasus, Numerus, Genus

Kasus und Numerus müssen immer angegeben werden. Ist der Numerus nicht entscheidbar, z.B. bei *welche meinst du*, erhält das Attribut Numerus den Wert *. Das Attribut Genus erhält bei Pluralformen den Wert *. Bei *wer* erhält das Attribut Numerus immer den Wert Sg, das Attribut Genus wird hier mit Masc annotiert.

Lemmatisierung: Als Lemma wird die Oberflächenform eingetragen.

Beispiele:

was_{PWS:Acc.Sg.Neut} er eigentlich machen will Lemma: was
wer_{PWS:Nom.Sg.Masc} soll mit wem_{PWS:Dat.Sg.Masc} diskutieren Lemma: wer Lemma: wem

PWAV (adverbiales Interrogativpronomen oder adverbiales Relativpronomen)

Hier erfolgt keine morphologische Annotation.

Lemmatisierung: Als Lemma wird, weil nicht flektierend, die Oberflächenform eingetragen.

Beispiel:

wo_{PWAV:--} Metall verwendet wird Lemma: wo

PROAV (Pronominaladverb)

Hier erfolgt keine morphologische Annotation.

Lemmatisierung: Als Lemma wird, weil nicht flektierend, die Oberflächenform eingetragen.

Beispiel:

man muß davon_{PROAV: --} ausgehen Lemma: davon

9 Adverbien

ADV (Adverb)

Hier erfolgt keine morphologische Annotation.

Lemmatisierung: Als Lemma wird, weil nicht flektierend, die Oberflächenform übernommen.

Beispiele:

Perot wäre vielleicht_{ADV: --} ein prächtiger Diktator Lemma: vielleicht

Ich glaube kaum_{ADV: --} Lemma: kaum

10 Konjunktionen

KOUS, KOU1, KON, KOKOM

Es erfolgt keine morphologische Annotation.

Lemmatisierung: Als Lemma wird, weil nicht flektierend, die Oberflächenform eingetragen.

Beispiele:

Ob_{KOUS: --} diese Fähigkeiten ausreichen Lemma: ob

und_{KON: --} auch die Konzernchefs Lemma: und

11 Adpositionen

Bei Präpositionen, die zwei Kasus regieren können (bei freier Variation des jeweils auftretenden Kasus: *wegen des / dem, trotz des / dem, dank des / dem*), erhält ein regiertes Femininum im Singular für Kasus den Wert *.

wegen der Hitze .Sg .Fem*

Die Präposition *entlang* regiert wie *trotz, dank, wegen* sowohl Genitiv als auch Dativ. Zusätzlich kann sie als Postposition auch den Akkusativ regieren.

APPR (Präposition Zirkumposition links)

Es erfolgt keine morphologische Annotation

Lemmatisierung: Als Lemma wird, weil nicht flektierend, die Oberflächenform eingetragen.

Beispiele:

über_{APPR:--} seine Wirtschaftspolitik Lemma: über
nach_{APPR:--} Meinungsumfragen Lemma: nach

APPO (Postposition)

Es erfolgt keine morphologische Annotation

Lemmatisierung: Als Lemma wird die unflektierte Oberflächenform eingetragen

Beispiele:

dem Magazin zufolge_{APPO:--} Lemma: zufolge
ihrer Ansicht nach_{APPO:--} Lemma: nach

APZR (Zirkumposition rechts) Es erfolgt keine morphologische Annotation

Lemmatisierung: Als Lemma wird die unflektierte Oberflächenform eingetragen.

Beispiele:

aus der deutschen Geschichte heraus_{APZR:--} Lemma: heraus

APPRART (Präposition mit Artikel)

Morphologische Merkmale: Kasus, Numerus, Genus

Kasus und Genus müssen immer angegeben werden. Verschmelzungen kommen nur mit definiten Artikeln im Singular vor. Deshalb trägt das Attribut Numerus stets den Wert Sg.

Lemmatisierung: als Lemma wird die Präposition ohne Artikel eingetragen.

Beispiele:

er liegt gut im_{APPRART:Dat.Sg.Neut} Rennen Lemma: in

12 Partikel

Es erfolgt bei den folgenden Partikeln keine morphologische Annotation.

Lemmatisierung: Als Lemma wird die Oberflächenform eingetragen.

PTKZU (zu vor Infinitiv)

Beispiele:

aus der Krise zu_{PTKZU:--} führen Lemma: zu
die Wahl zu_{PTKZU:--} gewinnen Lemma: zu

PTKNEG (Negationspartikel)

Beispiele:

halten nicht_{PTKNEG:--} viel Lemma: nicht

PTKVZ (abgetrennter Verbzusatz)

Beispiele:

lehnen den Milliardär als US-Präsidenten ab_{PTKVZ: --} Lemma: ab

PTKANT (Antwortpartikel)

Beispiele:

ja_{PTKANT: --} und zwar Lemma: ja

leider nein_{PTKANT: --} Lemma: nein

PTKA (Partikel bei Adjektiv oder Adverb)

Beispiele:

am_{PTKA: --} schönsten Lemma: am

zu_{PTKA: --} schnell Lemma: zu

13 Sonstige

ITJ (Interjektionen) erhalten keine morphologische Annotation. Als Lemma wird, da unflektiert, die Oberflächenform eingetragen.

XY erhält keine morphologische Annotation und kein Lemma.

TRUNC (Wortreste)

Wortreste kommen in verschiedenen Kontexten vor, am häufigsten bei der Koordination zweier Wortformen, die teilweise übereinstimmen. Die Annotation wird vom zweiten Kompositionsglied übernommen.

Lemmatisierung: Als Lemma wird die Form ohne Fugenmorphem eingetragen.

Beispiele:

Bildungs- und Arbeitsmarktpolitik Lemma: Bildung

Kindes- und Jugendalter Lemma: Kind

Kräfte- und Gesundheitsverschleiß Lemma: Kraft

Maschinen- und Anlagenbauer Lemma: Maschine

des Arbeiter- und Bauernstaates Lemma: Arbeiter

in- und ausländischen Schriftstellern Lemma: in

um- und ausgebaut Lemma: um

hin- und herirren Lemma: hin

bi- als auch multilaterale Aktionen Lemma: um

Hier wird nicht er- sondern gefunden Lemma: er

be- und geschlagene Ex-Strabag-Chef Lemma: be

Satzzeichen Satzzeichen erhalten keine morphologische Annotation und auch kein Lemma. Im Augenblick wird automatisch die Oberflächenform eingetragen, was vor der Veröffentlichung des Release wieder automatisch gelöscht wird.

Satzanfang Am Satzanfang werden die entsprechenden Lemmata immer klein geschrieben. Es sei denn, es handelt sich hierbei um Nomen, Eigennamen o.Ä.

14 Kurzfassung zum Nachschlagen

ADJA	Grad Kasus Numerus Genus	Lemma: Positiv b. präd. Gebr.
ADJD	Grad	Lemma: Positiv b. präd. Gebr.
ADV	keine Annotation	Lemma: Oberflächenform
APPR	keine Annotation	Lemma: Oberflächenform
APPRART	Kasus Numerus Genus	Lemma: Präposition ohne Art.
APPO	keine Annotation	Lemma: Oberflächenform
APZR	keine Annotation	Lemma: Oberflächenform
ART	Kasus Numerus Genus	Lemma: Nom. Sg. Masc.
CARD	keine Annotation	Lemma: Oberflächenform
FM	keine Annotation	Lemma: --
ITJ	keine Annotation	Lemma: Oberflächenform
KOUS KOUJ KON KOKOM	keine Annotation	Lemma: Oberflächenform
NN	Kasus Numerus Genus	Lemma: Nom.Sg.
NE	Annotation falls möglich aus dem Kontext: Kasus Numerus Genus ohne Vorname und im Plural: Genus erhält den Wert *	Lemma: Oberflächenform bzw. Nom.Sg.
PIAT PIS	Kasus Numerus Genus	Lemma: Nom. Sg. Masc.
PDAT PDS	Kasus Numerus Genus	Lemma: Nom. Sg. Masc.
PPOSAT PPOSS	Kasus Numerus Genus	Lemma: Nom. Sg. Masc.
PPER	Person Kasus Numerus Genus	Lemma: Oberflächenform
PRELAT	Kasus Numerus Genus	Lemma: Nom. Sg. Masc.
PRF	Person Kasus Numerus	Lemma: sich
PRELS	Kasus Numerus Genus	Lemma: Nom. Sg. Masc.
PROAV	keine Annotation	Lemma: Oberflächenform
PTKZU PTKNEG PTKVZ PTKANT PTKA	keine Annotation	Lemma: Oberflächenform

PWAT	Kasus Numerus Genus	Lemma: Nom. Sg. Masc.
PWS	Kasus Numerus Genus	Lemma: Oberflächenform
PWAV	keine Annotation	Lemma: Oberflächenform
TRUNC	Annotation des 2. Kompositums	Lemma: Wort ohne Fugenmorphem
VVFIN	Person Numerus Tempus Modus	Lemma: Infinitiv
VAFIN		
VMFIN		
VVIMP	Person Numerus Modus	Lemma: Infinitiv
VAIMP		
VVINF	nur Wert Inf bzw Infz zu eintragen	Lemma: Infinitiv
VVIZU		
VAINF		
VVPP VMPP	nur Wert Psp eintragen	Lemma: Infinitiv
VAPP		

A Literatur

Thielen, C., Schiller, A., Teufel, S. & Stöckert, C. (1999). *Guidelines für das Tagging deutscher Textkorpora mit STTS* (Tech. Rep.). Universität Stuttgart, Institut für maschinelle Sprachverarbeitung, and Seminar für Sprachwissenschaft, Universität Tübingen.