Lemmatisierungsrichtlinien für das EmpiriST-Korpus

Version 2019-08-26

Thomas Proisl Natalie Dykes Philipp Heinrich
Besim Kabashi Stefan Evert

26. November 2019

1 Vorbemerkungen

Die Lemmatisierungsrichtlinien sind als CMC-spezifische Ergänzung zum TIGER Morphologie-Annotationsschema zu lesen.[1] Zusätzlich sind im wortartenspezifischen Teil dieses Dokuments die für die Lemmatisierung wichtigen Punkte aus dem TIGER-Annotationsschema zusammengefasst (Abschnitt 6).

Wir verfolgen zwei verschiedene Lemmatisierungsstrategien – eine oberflächennahe und eine, die auf normalisierten Wortformen basiert. Die beiden Strategien unterscheiden sich in ihrem Umgang mit orthographischen Fehlern, Abkürzungen usw.

Die Tokenisierung der EmpiriST-2015-Daten wird als Goldstandard betrachtet und nicht geändert. Das bedeutet beispielsweise, dass zusammengeschriebene Wörter im Zuge der Lemmatisierung nicht getrennt werden. Das Token *jedenfall* (getaggt als NN) wird demnach als neu gebildetes Substantiv behandelt und zu *Jedenfall* lemmatisiert.

Als Referenzkorpus für Lemmatisierungsfragen kann das TIGER-Korpus[2] verwendet werden. Das Langenscheidt-Online-Wörterbuch[3] dient als Referenzwörterbuch um Zweifelsfälle zu klären; z.B. zur Lexikalisierung umgangssprachlicher Formen.

2 Datenformat

Die Annotationsdateien werden als fünfspaltige Tabelle erstellt. Die Spalten bezeichnen die **Wortform**, den **Part-of-Speech-Tag**, die **normalisierte Wortform**, das **oberflächennahe Lemma** sowie das **normalisierte Lemma**.

[1] `http://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/TIGERCorpus/annotation/tiger_scheme-morph.pdf`
[2] `http://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/tiger.html`
[3] `http://www.woerterbuch.langenscheidt.de/login/ip.html`

3 Normalisierung der Wortformen

Die normalisierte Wortform soll die folgenden Kriterien erfüllen:

- Korrektur offensichtlicher Schreibfehler, z. B. *dass*/*das*, *hinstelt* (*hinstellt*).
- Normalisierung zur neuen deutschen Rechtschreibung (*daß* zu *dass*.
- Normalisierung von Abkürzungen zu ihrer kanonischen Form (*zB* zu *z.B.*).
- Normalisierung der Groß-/Kleinschreibung: Substantive und (in der Regel Eigennamen, siehe Abschnitt 6.1.2) groß, alles andere klein, auch, wenn das Wort am Satzanfang steht (*WICHTIG* wird zu *wichtig*, *traktor* zu *Traktor*).
- Rückführung nichtlexikalisierter umgangssprachlicher Formen auf Standardformen (*ne* wird bspw. je nach Kontext zu *nicht* oder *eine* normalisiert).
- Vervollständigung von Kurzwörtern, sofern diese nicht lexikalisiert sind. Aus *Disku* wird *Diskussion*, aber *Disko* wird nicht zu *Diskothek*.
- Kontraktionen wie *machste* werden nicht aufgelöst.
- Genderspezifische Formen wie *Teilnehmerin* bleiben erhalten.
- Wenn verschiedene Varianten gebräuchlich sind, wird die verwendete Variante beibehalten (*gehts*/*geht's*).
- Bindestriche in Bindestrichkomposita werden **nicht** normalisiert (*Otto-normal-Bürger* bleibt erhalten und wird nicht zu *Otto-Normalbürger* o.ä.).
- Wiederholungsphänomene, bspw. Mehrfachvokale bei Interjektionen, werden auf die Wörterbuchform normalisiert (*uuuh* zu *uh*).
- Rekonstruktion von Flexionsendungen (*ich hab* wird zu *habe*).

4 Oberflächennahe Lemmatisierung

Die oberflächennahe Lemmatisierung ist eine Strategie für den Umgang mit nichtstandardsprachlichen Formen. Die hierbei resultierenden Lemmata erhalten möglichst viel von der ursprünglichen Schreibweise des Wortes. Hauptausschlaggebend für die Zuweisung zu einem Lemma sind der zugewiesene Part-of-Speech-Tag und die Flexionssuffixe.

Abweichungen vom Standard werden, wenn möglich, als kreativer Sprachgebrauch behandelt, aus dem eine Wortneuschöpfung hervorgegangen ist. Demnach wird die Form *Hinstelt* als flektierte Form eines vom Substantiv *Stele* abgeleiteten Präfixverbs verstanden und oberflächennah zu *hinstelen* lemmatisiert.

Wenn Flexionssuffixe nicht ausreichen, bspw. bei Stammänderungen, wird trotz eventueller Fehler standardsprachlich lemmatisiert: so wird *iest* etwa zu *sein*.

Die folgende Tabelle liefert eine Beispielübersicht, wobei das oberflächennahe Lemma als $Lemma_{sur}$ bezeichnet wird.

Beispiel	POS	Normalisiert	$\text{Lemma}_{\text{sur}}$	$\text{Lemma}_{\text{norm}}$
hinstelt	??	hinstellt	hinstelen	hinstellen
fannd	??	fand	finden	finden
weiß	VVFIN	weiß	wissen	wissen
ansosten	ADV	ansonsten	ansosten	ansonsten
Grigfe	NN	Griffe	Grigf	Griff

5 Normalisierte Lemmatisierung

Diese Lemmatisierungsstrategie baut auf den normalisierten Wortformen auf und erzeugt, soweit möglich, standardsprachliche Lemmata gemäß den Richtlinien in Abschnitt 6. Die Unterschiede zwischen den beiden Lemmatisierungsstrategien werden in der folgenden Tabelle dargestellt, wobei das normalisierte Lemma als $Lemma_{norm}$ bezeichnet wird.

Beispiel	POS	Normalisiert	$\text{Lemma}_{\text{sur}}$	$\text{Lemma}_{\text{norm}}$
Mio		Mio.	Mio	Mio.
ne		eine	n	ein
net		nicht	net	nicht
jedenfall	NN	Jedenfall	Jedenfall	Jedenfall
vielen	PIS	vielen	viele	viele
vielen	PIAT	vielen	vieler	vieler
Betreffenden	NN	Betreffenden	betreffend	betreffend
dies		dies	dieser	dieser
liegt's		liegt's	liegen	liegen

6 Richtlinien nach Wortarten

6.1 Substantive (NN, NE))

Das Lemma ist im Normalfall der Nominativ Singular. Ausnahmen sind Pluraliatantum (Nominativ Plural: *Leute* bleibt *Leute*), deadjektivische Substantive (schwache Form des Adjektivs im Nominativ Singular Maskulin), substantivierte Infinitive (verbaler Infinitiv), Substantive aus einem Partizip 1 (unflektierte Form des Partizips 1) und Substantive aus einem Partizip 2 (starke Form des Partizips 2 im Maskulin Singular).

6.1.1 NN (Appellativa)

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: 1) Substantive, die aus Infinitiven abgeleitet sind, 2) solche aus Adjektiven und 3) Substantive,

die aus Partizipien gebildet wurden. Bei den Substantiven aus Partizipien treffen wir zudem die Unterscheidung nach Partizip 1 und Partizip 2. Hier wird als Lemma eine verbale bzw. adjektivische Basis annotiert. Diese Lösung soll einerseits bei der Verwertung des Korpus den Zugriff auf die Nähe einer nominalen Form zum adjektivischen oder verbalen Paradigma ermöglichen, und andererseits wenige Abgrenzungsprobleme verursachen. 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 als Lemma die **schwache Form des Adjektivs im Nominativ Singular Maskulin**. Diese Regel kommt immer dann zur Anwendung, wenn es im gegenwärtigen Deutsch ein Adjektiv gleicher Form gibt. Beispiele:

Beispiel	Lemma
ein Alter	alte
ein Grüner	grüne
der Ältere	alte

2. Substantivierte Infinitive bekommen den **verbalen Infinitiv** als Lemma. Beispiele:

Beispiel	Lemma
das Lesen	lesen
das Fahren	fahren

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

Beispiel	Lemma
der Lesende	lesend
der Fahrende	fahrend
der Vorsitzende	vorsitzend

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

Beispiel	Lemma
ein Betroffener	betroffener
der Angeklagte	angeklagter

TIGER: 6–7

6.1.2 NE (Eigennamen)

Wenn die Standardform des Eigennamens kleingeschrieben wird (bspw. Benutzernamen), dann wird auch das Lemma kleingeschrieben. Hier ist das Weltwissen der Annotatoren gefragt.

Am schwierigsten zu handhaben sind Personennamen oder Benutzernamen in Chats, die oftmals abgekürzt oder verfremdet werden. Hier gelten folgende Regeln: Wenn es

sich bei der Oberflächenform eines (abgekürzten) (Benutzer-)Namens um einen etablierten Namen handelt, wird die Standardschreibung (d.h. Großschreibung) verwendet, unabhängig davon, ob bspw. die Eigenschreibweise des (vollständigen) Benutzernamens klein ist. Bei nicht-etablierten Namen orientiert sich die Groß-/Kleinschreibung an der Oberflächenform.

> Als Lemma wird die Form des **Nominativ Singular** eingetragen. Da die meisten Eigennamen höchstens im Genitiv eine Flexionsendung haben, entspricht das Lemma meist der Oberflächenform. Bei Eigennamen, die keine Singularform haben, wird die Form des **Nominativ Plural** eingetragen.
>
> TIGER: 9

6.2 Adjektive (ADJD, ADJA)

Das Lemma ist in der Regel der Positiv in prädikativem Gebrauch, d.h. die unflektierte Form (*größer* zu *groß*. Bei adjektivisch gebrauchten Partizipien ist das Lemma die Kurzform des Partizips.

6.2.1 ADJD (adverbiales oder prädikatives Adjektiv)

> Lemmatisierung: als Lemma wird der **Positiv** eingetragen. Bei Partizipien aus adjektivisch gebrauchten Verben ist das Lemma die **Kurzform des Partizips**. Beispiele:
>
Beispiel	Lemma
ein Unternehmen erfolgreich leiten	erfolgreich
Probleme sind doch weit größer	groß
Probleme sind gravierender als gemeinhin angenommen	gravierend
wie lange sie bleibt	lang
>
> TIGER: 10

6.2.2 ADJA (attributives Adjektiv)

> Lemmatisierung: Als Lemma wird der **Positiv der prädikativen Form** (= die unflektierte Form) des Adjektivs eingetragen. Falls es diese Kurzform nicht gibt (z.B. *der **zweite** Wagen*), setzen wir die **starke Form des Nominativ Singular Maskulinum** als Lemma ein (*zweiter*). Bei Herkunftsbezeichnungen wie *Berliner Sonntagszeitung* steht als Lemma *Berliner*. Im Zweifelsfall wird davon ausgegangen, dass es sich um einen Komparativ handelt: *früher* oder *weiter* werden als Formen von *früh* bzw. *weit* lemmatisiert.
>
> Für ein als ADJA verwendetes Partizip II wird als Lemma die **Kurzform des Partizips** eingetragen. Beispiele:

Beispiel	Lemma
prächtiger Sonnenschein	prächtig
texanische Milliardär	texanisch
um die größte Volkswirtschaft	groß
ein optimales Konzept	optimal
der zweite Punkt	zweiter
im lila Kleid	lila
mit überzeugten Neonazis	überzeugt
für die Dritte Welt	dritter
ein weiteres Anzeichen sei	weit
während früherer Streiks	früh
das Berliner Haushaltsloch	Berliner
in den siebziger Jahren	siebziger
nächste Woche	nächster

TIGER: 11

6.3 Zahlen (CARD)

Lemmatisierung: Als Lemma wird die **Nennform** eingetragen, die meist der Oberflächenform entspricht. Ausnahmen sind z.B. *zweier*, *dreier* usw. Beispiele:

Beispiel	Lemma
500 Großunternehmen	500
1962	1962
zwei Jahren	zwei

TIGER: 12

6.4 Verben (V.+)

V[VAM]FIN: Finites Verb

V[VA]IMP: Imperativ

V[VAM]INF: Infinitiv

VVIZU: Infinitiv mit *zu*

V[VAM]PP: Partizip Perfekt

Das Lemma ist der **Infinitiv**.

Beispiel	Lemma
Ob diese Fähigkeiten ausreichen	ausreichen
In Indien kann der Mittelstand	können
geh doch nach draußen	gehen
Laßt Honecker in Frieden	lassen
schauen Sie	schauen
es gilt Computer abzusetzen	absetzen
die Wahl zu gewinnen	gewinnen
davon sind selbst seine Kritiker überzeugt	überzeugen
seien gestrichen worden	werden
nicht geäußert hat	äußern
etwas stärker konjunkturstimulierend wirken können	konjunkturstimulierend
mit dem amtierenden Präsidenten	amtierend

TIGER: 12–13

V[VAM]PPER: Kontraktion Verb + Personalpronomen

Das Lemma ist der **Infinitiv des Verbs ohne Pronomen**.

Beispiel	Lemma
schreibste	schreiben
isses	sein
gehts	gehen

6.5 Artikel (ART)

Das Lemma für den bestimmten/definiten Artikel ist *der*, Lemma für den unbestimmten/indefiniten Artikel ist *ein*.

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

Beispiel	Lemma
er wäre ein prächtiger Diktator	ein
die Konzernchefs halten nicht viel von	der
sie lehnen den Milliardär ab	der

TIGER: 13

6.6 Pronomina (P.+)

Tag	Lemma
PAV	Oberflächenform
PDAT	Nominativ Singular Maskulin
PDS	Nominativ Singular Maskulin
PIAT	Nominativ Singular/Plural Maskulin beim artikellosen Gebrauch
PIS	Nominativ Singular/Plural Maskulin beim artikellosen Gebrauch
PPER	Oberflächenform
PPERPPER	Oberflächenform des ersten Pronomens (ichs: ich, dus: du, ers: er)
PPOSAT	Nominativ Singular Maskulin
PPOSS	Nominativ Singular Maskulin
PRELAT	Nominativ Singular Maskulin
PRELS	Nominativ Singular Maskulin
PRF	*sich* (Für alle Formen: mir, sich, einander, uns, euch, deiner, ...)
PWAT	Nominativ Singular Maskulin
PWAV	Oberflächenform
PWS	Oberflächenform

Beispiel	Lemma
ich$_{\mathbf{PPER}}$ glaube kaum	ich
machen ihnen$_{\mathbf{PPER}}$ besonders zu schaffen	ihnen
es$_{\mathbf{PPER}}$ ist wirklich schwer zu sagen	es
mit deinen$_{\mathbf{PPOSAT}}$ Sachen	dein
von seinen$_{\mathbf{PPOSAT}}$ Beschäftigten verlange er	sein
gibt Auskunft über seine$_{\mathbf{PPOSAT}}$ Wirtschaftspolitik	sein
das sei nicht seiner$_{\mathbf{PPOSS}}$	seiner
meinem$_{\mathbf{PPOSS}}$ hat es auch nicht geschadet	meiner
diese$_{\mathbf{PDAT}}$ Fähigkeiten verhelfen ihnen	dieser
dessen$_{\mathbf{PDS}}$ Fähigkeiten verhelfen ihnen	er
aber das$_{\mathbf{PDS}}$ ist nicht unser System	der
sachlich begründete Fragen - die$_{\mathbf{PDS}}$ gibt es durchaus -	der
auf deren$_{\mathbf{PDS}}$ Fragen gibt es keine Antwort	der
die Frau ist verrückt, auf deren$_{\mathbf{PDS}}$ Tour falle ich nicht mehr rein	der
Filter, die zum Schutz der Linse auf diese$_{\mathbf{PDS}}$ gesetzt werden	dieser
Geschäftsführer, der$_{\mathbf{PRELS}}$	der
Unternehmer, die$_{\mathbf{PRELS}}$ meinen	der
Bruch... vor dem$_{\mathbf{PRELS}}$ noch jeder	der
das Jahr, in dessen$_{\mathbf{PRELAT}}$ Verlauf	der
die Orte, deren$_{\mathbf{PRELAT}}$ Hotels weniger Gäste meldeten	der
welche$_{\mathbf{PWAT}}$ Positionen er einnimmt	welcher
in welcher$_{\mathbf{PWAT}}$ Art Soldaten eingesetzt werden	welcher
was$_{\mathbf{PWS}}$ er eigentlich machen will	was
wer$_{\mathbf{PWS}}$ soll mit wem$_{\mathbf{PWS}}$ diskutieren	wer/wem
wo$_{\mathbf{PWAV}}$ Metall verwendet wird	wo
man muß davon$_{\mathbf{PROAV}}$ ausgehen	davon

6.6.1 PIAT und PIS

Flektierbare Formen:

Beispiel	Lemma
alles$_{\mathbf{PIAT}}$ Gute	aller
andere$_{\mathbf{PIAT}}$ Regeln	anderer
beiden$_{\mathbf{PIAT}}$ Kindern	beide
einiges$_{\mathbf{PIAT}}$ Bier	einiger
etlicher$_{\mathbf{PIAT}}$ Unsinn	etlicher
irgendein$_{\mathbf{PIAT}}$ Hund	irgendein
irgendwelcher$_{\mathbf{PIAT}}$ Mist	irgendwelcher
jedem$_{\mathbf{PIAT}}$ Kind	jeder
jedwedes$_{\mathbf{PIAT}}$ Erfolgs	jedweder
jegliches$_{\mathbf{PIAT}}$ Geschäft	jeglicher
kein$_{\mathbf{PIAT}}$ Brot	kein
manchem$_{\mathbf{PIAT}}$ Mann	mancher
mehreren$_{\mathbf{PIAT}}$ Kindern	mehrere
der meiste$_{\mathbf{PIAT}}$ Müll	meister
reichlich$_{\mathbf{PIAT}}$ Fernwärme	reichlich
sämtliches$_{\mathbf{PIAT}}$ Gerümpel	sämtlicher
solcher$_{\mathbf{PIAT}}$ Schmerz	solcher
ebensolcher$_{\mathbf{PIAT}}$ Schmerz	ebensolcher
vieler$_{\mathbf{PIAT}}$ Kaffee	vieler
ebensovieler$_{\mathbf{PIAT}}$ Kaffee	ebensovieler
soviel$_{\mathbf{PIAT}}$ Potenzial	sovieler
zuvieler$_{\mathbf{PIAT}}$ Unsinn	zuvieler
nur weniger$_{\mathbf{PIAT}}$ Wein	weniger
wenigster$_{\mathbf{PIAT}}$ Wein	wenigster
allerwenigster$_{\mathbf{PIAT}}$ Müll	allerwenigster
alles$_{\mathbf{PIS}}$ ist hin	alle
unter anderem$_{\mathbf{PIS}}$	anderer
gegen Erich Honecker und andere$_{\mathbf{PIS}}$	anderer
Die anderen$_{\mathbf{PIS}}$ sind schuld	anderer
Beides$_{\mathbf{PIS}}$ gefällt uns	beide
Der einen$_{\mathbf{PIS}}$ traue ich	einer
Es gefiel so einigen$_{\mathbf{PIS}}$	einige
im einzelnen$_{\mathbf{PIS}}$	einzelner
der einzige$_{\mathbf{PIS}}$, der	einziger
nur wer als erster$_{\mathbf{PIS}}$	erster
was erstere$_{\mathbf{PIS}}$ angeht	ersterer
Etlicher$_{\mathbf{PIS}}$ wird so alt	etlicher
Irgendeiner$_{\mathbf{PIS}}$ kam rein	irgendeiner
Nimm irgendwelchen$_{\mathbf{PIS}}$	irgendwelcher
Irgendwem$_{\mathbf{PIS}}$ kann man immer helfen.	irgendwer
Jede$_{\mathbf{PIS}}$ ist willkommen	jeder
Jedweder$_{\mathbf{PIS}}$ kann bleiben	jedweder
Jegliches$_{\mathbf{PIS}}$ braucht seine Zeit	jeglicher
Jemand$_{\mathbf{PIS}}$ sollte es tun	jemand
Glaube einfach keinem$_{\mathbf{PIS}}$	keiner
das ist das letzte$_{\mathbf{PIS}}$	letzter
letzteren$_{\mathbf{PIS}}$ gegenüber	letzterer

TIGER: 16–17

Beispiel	Lemma
Er hat manche**PIS** gesehen	mancher
Mehrere**PIS** fehlten	mehrere
Die meisten**PIS** sind gut	meister
Sie hat am meisten**PIS**	meister
Ich bin niemandem**PIS** böse	niemand
Solche**PIS** brauchen wir	solcher
Ebensolchen**PIS** will ich	ebensolcher
Erna kauft viele**PIS**	viele
Hilde klaut ebensoviele**PIS**	ebensoviele
Karin kriegt zuviele**PIS**	zuviele
sechs weitere**PIS**	weiterer
Nur weniger**PIS** ist noch da	weniger
Nimm die wenigsten**PIS**	wenigster
Gib das allerwenigste**PIS**	allerwenigster

TIGER: 16–17

Nicht-flektierbare Formen (Lemma gleich Oberflächenform):

PIAT: all, genügend, lauter, manch, solch

PIS: ihresgleichen, jedermann, man

Beide: allerhand, allerlei, beiderlei, (ein) bisschen, derlei, ebensoviel, etwas, genug, irgendetwas, irgendwas, mancherlei, mehr, nichts, nix, (ein) paar, soetwas, solcherlei, sowas, viel, vielerlei, was, wenig, (ein) wenig, weniger, zuviel, zweierlei

TIGER: 17–19

6.7 Adverbien (ADV, ADVART)

ADV: Adverb

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

Beispiel	Lemma
Perot wäre vielleicht ein prächtiger Diktator	vielleicht
Ich glaube kaum	kaum

TIGER: 22

ADVART: Kontraktion: Adverb + Artikel

Lemma ist das **Adverb**. Beispiele:

Beispiel	Lemma
son	so
sone	so

6.8 Konjunktionen (KO.+)

KOUS: unterordnende Konjunktion mit Satz (VL-Stellung)

KOUI: unterordnende Konjunktion mit „zu“ und Infinitiv

KON: nebenordnende Konjunktion

KOKOM: Vergleichspartikel ohne Satz

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

Beispiel	Lemma
Ob diese Fähigkeiten ausreichen	ob
und auch die Konzernchefs	und

TIGER: 22

KOUSPPER: Kontraktion: unterordnende Konjunktion mit Satz (VL-Stellung) + irreflexives Personalpronomen

Lemma ist die **Konjunktion ohne Pronomen**. Beispiele:

Beispiel	Lemma
wenns	wenn
weils	weil
obse	ob

6.9 Adpositionen (APPR, APPO, APZR, APPRART)

APPR: Präposition, Zirkumposition links

APPO: Postposition

APZR: Zirkumposition rechts

Lemma ist die unflektierte **Oberflächenform**.

APPRART: Präposition mit Artikel

Lemmatisierung: als Lemma wird die **Präposition ohne Artikel** eingetragen. Beispiele:

Beispiel	Lemma
er liegt gut im Rennen	in
zum	zu

TIGER: 23

6.10 Partikel (PTK.+)

PTKZU: zu vor Infinitiv

PTKNEG: Negationspartikel

PTKVZ: abgetrennter Verbzusatz

PTKANT: Antwortpartikel

PTKA: Partikel bei Adjektiv oder Adverb

PTKIFG: Intensitäts-, Fokus- oder Gradpartikel

PTKMA: Modal- oder Abtönungspartikel

PTKMWL: Partikel als Teil eines Mehrwort-Lexems

Lemma ist die **Oberflächenform**.

6.11 Emoticons (EMOASC, EMOIMG

EMOASC: Emoticon, als Zeichenfolge dargestellt (Typ „ASCII“)

EMOIMG: Emoticon, als Grafik-Ikon dargestellt (Typ „Image“)

Lemma ist die **Oberflächenform**.

6.12 Online-Phänomene (ADR, AKW, EML, HST, URL)

ADR: Adressierung

AKW: Aktionswort

EML: E-Mail-Adresse

HST: Hashtag

URL: Uniform Resource Locator

Lemma ist die **Oberflächenform**.

6.13 Sonstige

6.13.1 ITJ (Interjektionen)

Lemma ist die **Oberflächenform**.

6.13.2 TRUNC (Wortreste)

Lemmatisierung: Als Lemma wird die **Form ohne Fugenmorphem** eingetragen. Beispiele:

Beispiel	Lemma
Bildungs- und Arbeitsmarktpolitik	Bildung
Kindes- und Jugendalter	Kind
Kräfte- und Gesundheitsverschleiß	Kraft
Maschinen- und Anlagenbauer	Maschine
des Arbeiter- und Bauernstaates	Arbeiter
in- und ausländischen Schriftstellern	in
um- und ausgebaut	um
hin- und herirren	hin
bi- als auch multilaterale Aktionen	um
Hier wird nicht er- sondern gefunden	er
be- und geschlagene Ex-Strabag-Chef	be

TIGER: 24

6.13.3 DM (Diskursmarker)

Lemma ist die **Oberflächenform**.

6.13.4 ONO (Onomatopoetikon)

Lemma ist die **Oberflächenform**.

6.13.5 FM (Fremdsprachliches Material)

Lemma ist die **Oberflächenform**.

6.14 Interpunktion ($., $„ $()

Lemma ist die **Oberflächenform**.