

Gerd Willée, IKP – Universität Bonn  
Harald Elsen, DELTA International CITS GmbH Bonn  
unter Mitarbeit von Karlheinz Stöber, IKP – Universität Bonn

Funktionsbeschreibung von LEMMA4,  
einem Wortklassentagger und  
Lemmatisierer für unrestringierte  
deutsche Texte

**IKP-Arbeitsbericht NF 16**  
ISSN 1614-6271

Bonn, 20.12.2005

## Funktionsbeschreibung

LEMMA4 ist ein hauptsächlich wortformenbasierter Wortklassentagger und Lemmatisierer für unrestringierte deutsche Texte.

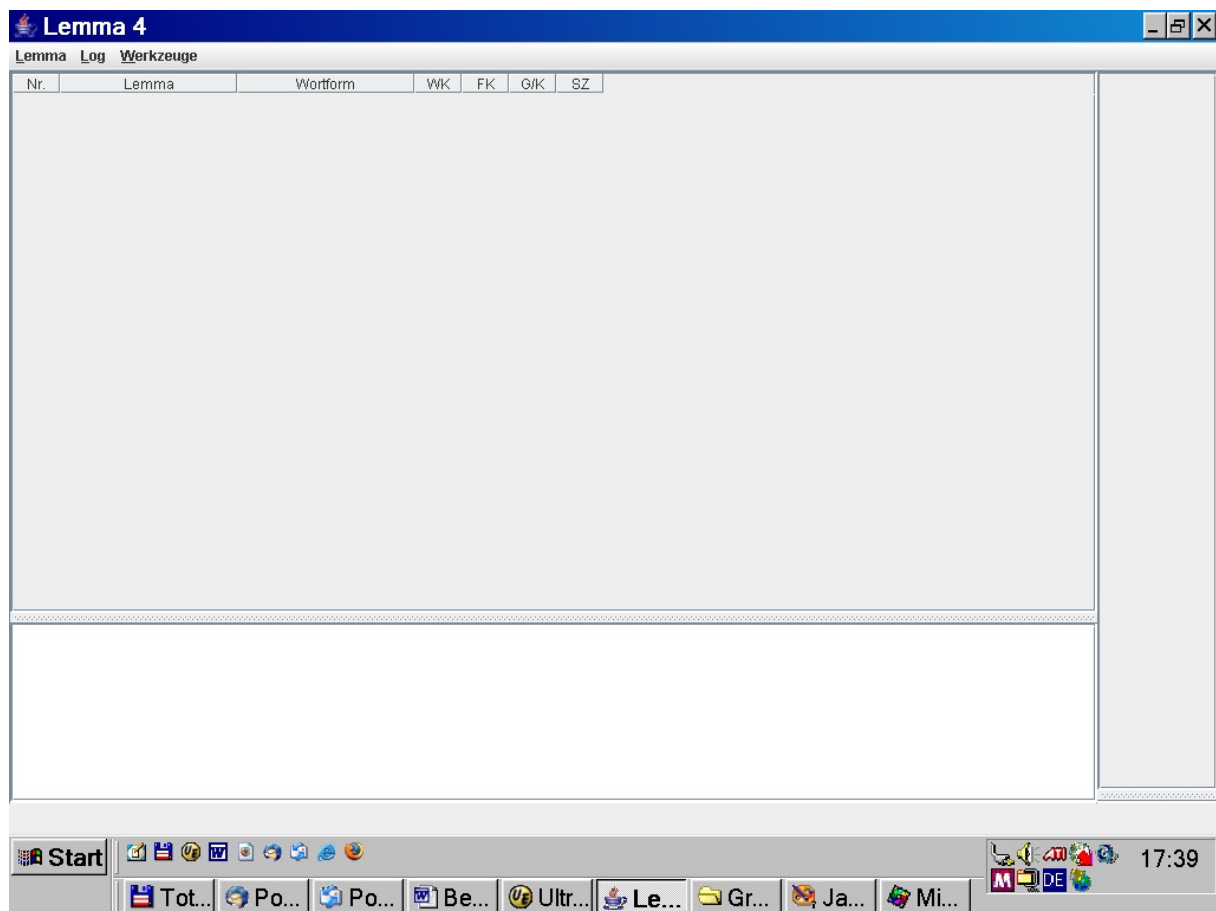
Es wurde zunächst durch Karlheinz Stöber in C++ programmiert. Danach setzte Harald Elsen das Programm in JAVA um. Es wird sowohl als Teilmodul als auch stand-alone einsetzbar sein.

LEMMA4 weist jeder Textwortform die (eine) Wortklassenangabe zu und generiert die zugehörige Grundform. Bei Verben wird zusätzlich noch bei den einfachen Tempora die Flexionsausprägung angegeben.

Die Programmdateien werden als zip-file geliefert und müssen entpackt werden. Durch Aufruf der dann vorliegenden setup.exe wird LEMMA4 installiert und ein Icon auf dem Desktop angelegt.

Der Start von LEMMA4 erfolgt durch Doppelklick auf das Icon.

Es erscheint der folgende Schirm:



Durch Klicken auf 'Lemma' erscheint die Auswahl

- Öffnen
- Sichern
- Beenden

Durch Wählen der Option 'Öffnen' kann die Eingabedatei bestimmt werden.

Mit dem Bestätigen der Auswahl beginnt LEMMA4 zu arbeiten. Nach Abschluss kann über 'Lemma' das Ergebnis gespeichert werden.

Klicken auf 'Log' ermöglicht ein Abspeichern der Log-Datei, in der Fehlermeldungen und Warnungen während des Laufs von LEMMA4 abgelegt werden. Im Fall von Problemen sollte diese Datei zusammen mit dem Textteil, bei dessen Bearbeitung diese aufgetreten sind, an **willee@uni-bonn.de** geschickt werden.

Folgende Wortklassen werden von LEMMA4 erkannt:

a) Verben

- VE** → alle Verbformen außer:
- VP** → Partizip I (Homograph VE-AD)
- VA** → Partizip II (Homograph VE-AD)  
bei erkannter Homographie zu 'VF' mit Angabe der Flexionsausprägung

b) Verbzusätze

- VX** → echte Verbzusätze
- AX** → möglicher Homograph AD-VX bzw. AV-VX

c) Substantive

- SU** → alle Substantiv-Formen

d) Adjektive

- AD** → alle Adjektiv-Formen
- VP** → Partizip I (Homograph VE-AD)
- VA** → Partizip II (Homograph VE-AD)
- AX** → möglicher Homograph AD-VX

e) Adverbien

- AV** → echte Adverbien
- AX** → möglicher Homograph AV-VX

f) Pronomina (incl. Artikel, excl. Personalpronomina)

**PR**

g) Personalpronomina

- P1**(1.Sg) → 'ich'
- P2**(2.Sg) → 'du'
- P3**(3.Sg) → 'er, sie, es'
- P4**(1.Pl) → 'wir'
- P5**(2.Pl) → 'ihr'
- PE** → für alle übrigen

h) Konjunktionen

**KO**

i) Zahlwörter u. Zahlen

**NU**

j) Präpositionen

**PP** → Präpositionen

**PO** → Postpositionen

k) Interjektionen (incl. 'ja', 'nein', 'danke', 'bitte')

**IJ**

l) Abkürzungen (Abkürzungen und Maßbezeichnungen wie 'Baud' oder 'Celsius')

**AK**

m) Homographen

**VP** → Partizip I (Homograph VE-AD)

**VA** → Partizip II (Homograph VE-AD)

**AX** → Verbzusatz-Adverb bzw. Adjektiv

In den meisten Fällen ist eine Differenzierung von 'AX' nur durch semantische Analysen möglich; die Kategorien werden daher nur dort disambiguiert, wo es sich syntaktisch eindeutig um Adverbien handelt.

Im Fall von als 'AD' erkannten Adjektiven wird diese Kategorie beibehalten, auch wenn es sich um einen Verbzusatz handeln könnte.

n) Besonders klassifizierte Satzzeichen und Symbole

"	→	<Q>
'	→	<Q>
(	→	<PaO>
)	→	<PaC>
{	→	<PaO>
}	→	<PaC>
[	→	<PaO>
]	→	<PaC>
<	→	<PaO>
>	→	<PaC>
%	→	<SMB>
\$	→	<SMB>
€	→	<SMB>

'Q' steht für Anführungszeichen (quotationmark);

'PaO' steht für öffnende,

'PaC' steht für schließende Klammer jeder Art;

'SMB' steht für Anführungszeichen (quotationmark);

'Q' steht für Anführungszeichen (quotationmark).

Die Markierung der Flexionsausprägung bei Verben sieht wie folgt aus:

<b>G1-G6</b>	→	Präsens Aktiv
<b>V1-V6</b>	→	Präteritum Aktiv
<b>K1-K6</b>	→	Konjunktiv (I oder II) Aktiv
<b>IS, IP</b>	→	Imperative

## Beispiel einer Bearbeitung mit LEMMA4:

### Input:

(Entnommen dem Roman 'Tonio Kröger' von Thomas Mann)

Was aber den Tanz betraf, so meisterte Herr  
Knaak ihn womöglich in noch höherem Grade.

### Output:

	<i>Wortform</i>	<i>Grundform</i>	<i>Wortklasse</i>	<i>Flexionsausprägung</i>	<i>Satzzeichen</i>
1	Was	wer,was	PR		
2	aber	aber	KO		
3	den	der,die,das	PR		
4	Tanz	tanz	SU		
5	betraf	betreffen	VE	V3	,
6	so	so	AV		
7	meisterte	meistern	VE	V3	
8	Herr	herr	SU		
9	Knaak	knaak	SU		
10	ihn	er,es	PE		
11	womöglich	womöglich	AV		
12	in	in	PP		
13	noch	noch	AV		
14	höherem	hoch	AD		
15	Grade	grad	SU		.

# Lexikonvergleich

## Beschreibung der Elemente des Lexikons

Das Lexikon enthält die Elemente der geschlossenen Wortklassen sowie Abkürzungen.

Jeder Entry hat folgenden Aufbau:

**Wortform – Grundform – <Wortklasse>**

Nur bei den Abkürzungen sowie bei gesteigerten Adverbien ist die Grundform von der Wortform verschieden.

Beispiele aus dem Lexikon:

**AZ. aktenzeichen <SU>**  
**Ab ab <PP>**  
**Abb. abbildung <SU>**  
**Abend abend <S2>**  
**Aber aber <KO>**  
**Abermals abermals <AV>**  
**Abf. abfahrt <SU>**  
**Abseits abseits <pa>**  
**u. u. <AK>**  
**u.a. u.a. <AK>**

Der Vergleich Textwortformen findet mit den Lexikonwortformen unter Berücksichtigung möglicher Groß-/Kleinschreibungen statt.

Soweit Abkürzungen eindeutig auflösbar sind und dabei in einer einteiligen Wortform enden, werden diese für die Grundform aufgelöst und die entsprechende Wortklasse der Vollform angegeben (Beispiele: **AZ.**, **Abb.**, **Abf.**).

Mehrfach auflösbare werden als **AK** klassifiziert (Beispiel: **u.**, das sowohl für 'und' als auch für 'unter' stehen kann).

Abkürzungen, die für einen Mehrwortgruppe stehen (Beispiel: **u. a.**, das für 'unter anderem' steht) bleiben unaufgelöst und werden ebenfalls als **AK** klassifiziert.

Entries wie **ab**, **aber** oder **abermals** erhalten die definitiven eindeutigen Wortklassenzuweisungen, während Entries wie **Abend** oder **Abseits** eine vorläufige homographe Wortklassenzuweisung erhalten, die erst im Modul **UMGEBUNG** entsprechend dem syntaktischen Kontext aufgelöst werden.



# Regeln in MORPH

## Regeln bei der Verwendung der Analysedatei MORPHLIS

Die Entries von MORPHLIS bestehen aus Stämmen bzw. Teilen davon (als zu suchende Strings), der Angabe der Wortklasse, der Grundform und der erlaubten Endungen.

Beispiele für Einträge aus MORPHLIS:

**einzig AD einzig e, em, en, es, er, \$,  
geh VE gehen e,1 \$,11 st,3 est,12 (...)  
liebe SU liebe \$,**

Die Struktur jeden Eintrags sieht aus wie folgt:

**(Teil-)Stamm Wortklasse erlaubte Endungen, ggf. mit Aktionsnr.**

Ein Match ist nur möglich, wenn der (Teil-)Stamm mit einer der erlaubten Endungen kombiniert dem Endstring der zu bearbeitenden Wortform entspricht.

Die zu generierende Grundform wird dann gebildet, indem dieser Endstring durch die Grundform im MORPHLIS-Entry ersetzt wird- Die Wortklasse wird ebenfalls dem Entry entnommen.

'\$' als Endung bedeutet keine Endung.

Substantiv-Entries können nur bearbeitet werden, wenn die zu untersuchende Wortform groß geschrieben ist.

Bei Verben stehen neben den Endungen noch Codes für die Flexionsausprägung (an dieser Stelle meist noch mehrdeutig; die Disambiguierung erfolgt in UMGEBUNG).

Bei bestimmten Endungen wird die Wortklasse dann in 'VP' für Partizip I bzw. 'VA' für Partizip II geändert. Bei Partizipia II wird bei der Herstellung der Grundform ein ggf. vorhandenes infigiertes '-ge-' entfernt, bei Partizipia I ein ggf. vorhandenes infigiertes '-zu-' (letzteres auch bei möglichen Infinitivendungen).

Die Unterscheidung von Adjektiven und Verben mit gleichem Stamm (z.B. 'lieb' vs. 'lieben' kann nur bei unterschiedlichen Endungen vorgenommen werden. Bei gleichen Endungen ('-e' und '-en') muss diese Unterscheidung im Modul UMGEBUNG entsprechend dem syntaktischen Kontext vorgenommen werden.

Die Aktionsnummern werden wie folgt aufgelöst:

Angegeben ist ein numerischer Code, der in den entsprechenden Flexionscode umzusetzen ist.

**(ge)** bedeutet, dass diese Analyse nur vorgenommen werden kann, wenn sich VOR dem in der MORPHLIS gefundenen String kein **'-ge-'** lokalisieren lässt. Falls doch, ist über die Flexivliste zur Bestimmung der Partizipia II neu zu bestimmen.

**(zu)** bedeutet, dass - falls sich ein **'-zu-'** vor dem gefundenen String <wie oben beschrieben> befindet - dieses für die Herstellung der Grundform zu entfernen ist [Beispiel: herauszugebendes -> herausgeben]

Aktionen 23, 24, 25 und 26 stellen keine Bereitstellung eines Flexionscodes dar, vielmehr wird hier die Wortklasse geändert, sei es nach **VP** (Partizip I) oder nach **VA** (Partizip II), wobei im Fall von **VP** ggf. eingeschlossenes **'-zu-'** entfernt werden muss, bzw. nach **NU** oder **AV**.

**0 G1**  
**1 G0#**  
**2 G1#**  
**3 G2**  
**4 G3**  
**5 G3# (ge)**  
**7 G6#**  
**8 G5#**  
**9 G6# (zu)**  
**10 G6# (zu) (ge)**  
**11 IS**  
**12 K2**  
**13 K3#**  
**14 K5**  
**15 K6#**  
**16 V2**  
**17 V3# (ge)**  
**18 V3#**  
**20 V5**  
**21 V6# (ge)**  
**22 V6#**  
**23 VP (zu)**  
**24 VA**  
**25 Wa = PR**  
**26 Wa = NU**  
**27 Wa = AV**  
**28 G6# (zu)**  
**30 G2# (ge)**  
**31 V2#**  
**32 K2#**

Zur Vermeidung falscher Flexivbehandlungen sind die Aktionen wie folgt zu präzisieren (nur da vermerkt, wo sich Änderungen ergeben):

a) generell bei Part. II

die Grundform wird dadurch hergestellt, dass die gefundene Wortform INKLUSIVE eines ggf. vorhandenen 'ge' abgetrennt wird und dann die in der Liste befindliche Grundform angefügt wird. Wortklasse ist dann **'VA'**.

b) bei homographen Flexiven zwischen Präteritum und Part. II

Es handelt sich um folgende Regeln:

**5 G3# (ge)**  
**10 G6# (zu) (ge)**  
**17 V3# (ge)**

**21 V6# (ge)**  
**30 G2# (ge)**

In all diesen Fällen kann der angegebene Flexionscode nur eingefügt werden, wenn KEIN '-ge-' gefunden worden ist. Falls ein '-ge-' da ist, muss die Regel für Part. II angewendet werden, d.h. Wortklasse ist **VA**, Grundformherstellung wie oben.

Bei der Aktionsnummer 10 [**10 G# (zu) (ge)**] ist im Fall, dass ein '-zu-' gefunden wird der Flex\_Code auf '**G6**' zu setzen, das '-zu-' ist zu löschen. Letzteres gilt auch für die erkannten Partizipia I (Aktionsnummer 23).

# Regeln in UMGEBUNG

## 1. Regeln zur Bestimmung der Wortklasse

### 1.1. Regeln für unbestimmte Wortformen auf folgende Endgraphien (und KEINER Großscheidung)

Wenn kleingeschriebene Wortformen folgende Endungen haben:

**a) <-ere, -erem, -eren, -erer, -eres, -este, -estem, -esten, -ester>**

**b) <-ste, -stem, -sten, -ster>**

((b) ist erst zu durchsuchen, wenn (a) erfolglos war)

Die Grundform wird gebildet aus der Wortform ohne die gefundene Endgraphie.

Die Wortart ist jeweils **AD**.

### 1.2 Regeln für unbestimmte kleingeschriebene Wortformen auf -e und -en

Die Regeln sind wie folgt zu lesen, z.B. im Fall der Regel 2:

Steht hinter einer unbestimmten kleingeschriebenen Wortform, die auf '-e' endet das Personalpronomen 'ich' (Wortklasse **P1**), dann wird diese Wortform als Verb mit der Flexionsausprägung <**G1**> (1. Pers. Sing. Präsens Indikativ Aktiv) bestimmt.

#### a) Wf endet auf '-e'

1. 'P1' + ' ' → 'P1' + 'VE' <G1>

2. ' ' + 'P1' → 'VE' + 'P1' <G1>

3. 'P3' + ' ' → 'P3' + 'VE' <K3>

4. ' ' + 'P3' → 'VE' + 'P3' <K3>

**Grundform = Wortform + '-n'**

#### b) Wf endet auf '-en'

5. 'P4' + ' ' → 'P4' + 'VE' <G4>

6. ' ' + 'P4' → 'VE' + 'P4' <G4>

**Grundform = Wortform**

**c) Wf endet auf '-e' oder '-en' oder '-n'**

- 7. ' ' + 'SU' → 'AD' + 'SU'
- 8. 'PR' + ' ' + 'SU' → 'PR' + 'AD' + 'SU'
- 8a. 'AD' + ' ' → 'AD' + 'AD'
- 8b. ' ' + 'AD' → 'AD' + 'AD'
- 9. ' ' + ",," + 'AD' → 'AD' + ",," + 'AD'
- 10. ' ' + "oder" + 'AD' → 'AD' + "oder" + 'AD'
- 11. ' ' + "und" + 'AD' → 'AD' + "und" + 'AD' >> prüfen
- 12. 'AD' + ",," + ' ' → 'AD' + ",," + 'AD'
- 13. 'AD' + "oder" + ' ' → 'AD' + "oder" + 'AD'
- 14. 'AD' + "und" + ' ' → 'AD' + "und" + 'AD'

Wenn Wortform auf '-e', dann  
Grundform = Wortform ohne '-e'.  
Wenn Wortform auf '-en', dann  
Grundform = Wortform ohne '-en'.  
Wenn Wortform auf '-n', dann  
Grundform = Wortform ohne '-n'.

**1.3 Default-Regel, wenn alle anderen Regeln nicht zutreffen:**

**Grundform = Wortform.**  
**Unbestimmte großgeschriebene Wff. → 'SU',**  
**falls Wf auf '-s' endet, wird dieses bei der Gf abge-**  
**trennt;**  
**unbestimmte kleingeschriebene Wff. → 'AD'**

## 2. Regeln zur Vereindeutigung der vorläufigen Flexionscodes

In dem Programmteil UMGEBUNG werden alle vorläufigen Flexionscodes entsprechend dem Vorhandensein bestimmter Personalpronomina vereindeutigt. Folgende Regeln kommen zum Einsatz:

### 1. **G0#**

Bei Vorkommen von P1 ('ich'), endgültiger Code = G1  
sonst Code = K3

Beispiel: aale

### 2. **G1#**

Bei Vorkommen von P1 ('ich'), endgültiger Code = G1  
sonst Code = G3

Beispiel: kann

### 3. **G3#**

Bei Vorkommen von P5 ('ihr'), endgültiger Code = G5  
sonst Code = G3

Beispiel: quirlt

### 4. **G2#**

Bei Vorkommen von P5 ('ihr'), endgültiger Code = G5  
bei Vorkommen von P2 ('du'), endgültiger Code = G2  
sonst Code = G3

Beispiel: schätzt, bremst

### 5. **G5#**

Bei Vorkommen von P5 ('ihr'), endgültiger Code = G5  
sonst Code = IP

Beispiel: schläft

### 6. **G6#**

Bei Vorkommen von P4 ('wir'), endgültiger Code = G4  
sonst Code = G6

Beispiel: schlafen

### 7. **V2#**

Bei Vorkommen von P5 ('ihr'), endgültiger Code = V5  
sonst Code = V2

Beispiel: last

- 8. V3#**  
Bei Vorkommen von P1 ('ich'), endgültiger Code = V1  
sonst Code = V3  
Beispiel: kam, legte
- 9. V6#**  
Bei Vorkommen von P4 ('wir'), endgültiger Code = V4  
sonst Code = V6  
Beispiel: kamen, legten
- 10. K1#**  
Bei Vorkommen von P1 ('ich'), endgültiger Code = K1  
sonst Code = K3  
Beispiel: sei
- 11. K2#**  
Bei Vorkommen von P5 ('ihr'), endgültiger Code = K5  
sonst Code = K2  
Beispiel: läst
- 12. K3#**  
Bei Vorkommen von P1 ('ich'), endgültiger Code = K1  
sonst Code = K3  
Beispiel: miede
- 13. K6#**  
Bei Vorkommen von P4 ('wir'), endgültiger Code = K4  
sonst Code = K6  
Beispiel: könnten

### 3. Regeln zur Markierung von möglichen Homographen zwischen Partizip II – flektierter Verbform

Nach der Vereindeutigung der Flexionscodes ist zu prüfen, ob homografe Formen zu Partizipia II vorliegen bzw. ob Partizipialformen als reine Adjektive gebraucht werden.

1. Wortart ist **'VA'**  
Prüfung auf das Präfix **'un-'** aber nicht **'unter-'**  
falls Präfix vorhanden, dann:  
Wortart ändern in **'AD'**  
Grundform ändern in Wortform,  
ggf. reduziert um **'-e', '-em', '-en', '-er' oder '-es'**.
  
2. Wortart ist **'VE'** und der FlexionsCode ist **'V3'** bzw. **'V1'**  
falls Endung = **'-te'** und Präfix = **'unter-', 'wider-', 'wieder-', 'be-', 'umbe-', 'er-', 'aber-', 'aner-', 'zuer-', 'umer-', 'ver-', 'ent-', 'zer-', 'über-', 'be-', 'um-', 'vorbe-', 'nachbe-',**  
Wortart ändern in **'VA'**
  
3. Wortart ist **'VE'** und der FlexionsCode ist **'V4'** bzw. **'V6'**  
falls Endung = **'-ten'** und Präfix = **'unter-', 'wider-', 'wieder-', 'be-', 'umbe-', 'er-', 'aber-', 'aner-', 'zuer-', 'umer-', 'ver-', 'ent-', 'zer-', 'über-', 'be-', 'um-', 'vorbe-', 'nachbe-'**  
Wortart ändern in **'VA'**



## 4. Regeln zur Vereindeutigung der homographen Wortklassenzuweisungen aus dem Lexikon

**Die geänderten Regeln sind markiert!**

**1. IX** Homograph Interjektion – Verbzusatz – Substantiv  
Beispiel: 'los'

Wenn die Wf, großgeschrieben wird und **nicht** am Satzanfang steht:  
dann Wortklasse = **'SU'**

sonst:

Wenn Wf am Satzanfang steht oder davor ein Satzzeichen steht:

dann Wortklasse = **'IJ'**

sonst:

Wortklasse = **'VX'**

**2. pa** Homograph Präposition - Adverb - Verbzusatz  
Beispiel: 'außerhalb', 'abseits'

Wenn rechts von der Wf entweder **"von"** steht oder eine Wortform mit einer der folgenden Wortklassen:

**<PR, PP, PE, p5, P3, AV, AD, NU, VA, VP>**

dann: Wortklasse = **'PP'**

sonst:

Wenn Wf am Satzanfang steht oder davor ein Satzzeichen steht:

dann Wortklasse = **'AV'**

sonst:

Wortklasse = **'VX'**

**3. ge** Homograph Präposition - Postposition - Adverb - Verbzusatz  
Beispiel: 'gegenüber'

Wenn Wf nicht am Satzanfang steht und wenn links von der Wf eine Wortform mit einer der folgenden Wortklassen steht:

**<PR, PP, PE, p5, P3, SU>, nicht gefolgt von einem Komma**

dann Wortklasse = **'PO'**

sonst:

Wenn rechts von der Wf entweder **"von"** steht

oder eine Wortform mit einer der folgenden Wortklassen:

**<PR, PE, p5, AV, AD, NU, VA, VP>**

dann Wortklasse = **'PP'**

sonst:

Wenn Wf am Satzanfang steht oder davor ein Satzzeichen steht:

dann Wortklasse = **'AV'**

sonst:

Wortklasse = **'VX'**

**4. g1** Homograph Postposition - Präposition  
Beispiel: 'wegen'

Wenn Wf nicht am Satzanfang steht und wenn links von der Wf eine Wortform mit einer der folgenden Wortklassen steht:

**<AV, PR, PP, SU, VA, VP, AD, NU>** , nicht gefolgt von einem Komma  
dann Wortklasse = **'PO'**

sonst:

Wortklasse = **'PP'**

**5. pV** Homograph Pronomen - Verb - Substantiv  
Beispiel: 'sein'

Wenn die Wf, großgeschrieben wird und nicht am Satzanfang steht:  
dann Wortklasse = **'SU'**

sonst:

Wenn rechts von der Wf eine Wf steht mit der Wortklasse **'AV'** und rechts von diesem Adverb eine Wortform mit einer der folgenden Wortklassen:

**<AD, VA, VP>**

dann Wortklasse = **'PR'**

Wenn rechts von der Wf eine Wortform steht mit einer der folgenden Wortklassen:

**<AD, VA, VP, SU>**

dann Wortklasse = **'PR'**

sonst:

Wortklasse = **'VE'**

**6. eV** Homograph Pronomen - Verbzusatz  
Beispiel: 'ein'

Wenn rechts von der Wf eine Wf steht mit der Wortklasse **'AV'** und rechts von diesem Adverb eine Wortform mit einer der folgenden Wortklassen:

**<AD, VA, VP>**

**oder**

rechts von der Wf die Wf **'und'** oder **'oder'** und dahinter eine Wf mit der Wortklasse **'PR'**

dann Wortklasse = **'PR'**

Wenn rechts von der Wf eine Wortform steht mit einer der folgenden Wortklassen:

**<AD, PR, NU, VA, VP, SU>**

dann Wortklasse = **'PR'**

Wenn rechts von der Wf eine Wortform steht mit einer der folgenden Grundformen:

**<'wenig', 'bißchen', 'bisschen'>**

dann Wortklasse = **'PR'**

sonst:

Wortklasse = **'VX'**

**7. p5** Homograph Pronomen - Personalpronomen  
Beispiel: 'ihr'

Wenn rechts von der Wf eine Wf steht mit der Wortklasse **'AV'** und rechts von diesem Adverb eine Wortform mit einer der folgenden Wortklassen:

**<AD, VA, VP>**

dann Wortklasse = **'PR'**

Wenn rechts von der Wf eine Wortform steht mit einer der folgenden Wortklassen:

**<AD, VA, VP, NU, SU>**

dann Wortklasse = **'PR'**

sonst:

Wortklasse = **'P5'**

**8. a7** Homograph Pronomen - Adverb  
Beispiel: 'wenig'

Wenn rechts von der Wf eine Wf steht mit der Wortklasse **'AV'** und rechts von diesem Adverb eine Wortform mit einer der folgenden Wortklassen:

**<AD, VA, VP>**

dann Wortklasse = **'PR'**

Wenn rechts von der Wf eine Wortform steht mit einer der folgenden Wortklassen:

**<AD, VA, VP, SU, PP>**

dann Wortklasse = **'PR'**

sonst:

Wenn Wf nicht am Satzanfang steht und wenn links von der Wf eine Wortform mit einer der folgenden Wortklassen steht:

**<AV, PR>**

dann Wortklasse = **'PR'**

Wenn links von Wf die Wf **<"zu">** steht,

dann Wortklasse = **'PR'**

sonst:

Wortklasse = **'AV'**

**9. S1** Homograph Interjektion - Substantiv  
Beispiel: 'Bitte'

Wenn Wf am Satzanfang steht,

dann Wortklasse = **'IJ'**

sonst:

Wortklasse = **'SU'**

**10. S2** Homograph Interjektion - Substantiv  
Beispiel: 'Vormittag'

Wenn Wf am Satzanfang steht,

dann Wortklasse = **'AV'**

sonst:

Wortklasse = **'SU'**

**11. aK** Homograph Konjunktion - Adverb  
Beispiel: 'da'

Wenn Wf am Satzanfang steht oder davor ein Satzzeichen steht  
und wenn rechts von der Wf **KEINE** Wf steht mit der Wortklasse **'VE'**:  
dann Wortklasse = **'KO'**  
sonst:  
Wortklasse = **'AV'**

**12. pO** mögliche Konjunktion oder Präposition  
Beispiel: 'ohne'

Wenn links von der Wortform <"ohne"> eine Wf. **mit folgendem Komma**  
steht oder **"ohne"** die erste Wortform eines (Teil-)Satzes ist und dahinter  
**keine** Wortform mit der Wortklasse <"PR"> steht,  
oder rechts dahinter vor dem nächsten Komma ein **"zu"** steht,  
dann Wortklasse = **'KO'**  
sonst:  
Wortklasse = **'PP'**

**13. pU** mögliche Konjunktion oder Präposition/Verbzusatz  
Beispiel: 'um'

Wenn links von der Wortform <"um"> eine Wf. **mit folgendem Komma**  
steht oder **"um"** die erste Wortform eines (Teil-)Satzes ist und dahinter **keine**  
Wortform mit der Wortklasse <AD, VA, VP, SU, NU, PE, P5, P3, PR> steht,  
oder rechts dahinter vor dem nächsten Komma ein **"zu"** steht,  
dann Wortklasse = **'KO'**  
sonst:  
Wortklasse = **'PX'**

**14. wK** Homograph Konjunktion - Adverb  
'wie'

Wenn rechts von der Wf Wf steht mit der Wortklasse **'AD' oder 'AV'**  
dann Wortklasse = **'AV'**  
Wenn rechts von der Wf Wf steht mit den Wortklassen  
< 'PR', 'PE', 'P1' .. 'P5', 'PP' >  
dann Wortklasse = **'KO'**  
Wenn Wf am Satzanfang steht oder davor ein Satzzeichen steht  
und wenn rechts von der Wf **KEINE** Wf steht mit der Wortklasse **'VE'**:  
dann Wortklasse = **'KO'**  
sonst:  
Wortklasse = **'AV'**

**15. sK** Homograph Konjunktion - Adverb

'so'

Wenn rechts von der Wf die Wortform

< **'dass' oder 'daß'** > oder einem Komma steht:

dann Wortklasse = **'KO'**

Wenn rechts von der Wf Wf steht mit den Wortklassen

< **'PE', 'P1' .. 'P5', 'PR', 'AV', 'AD', 'SU', 'VP', 'VA', 'NU'** >:

dann Wortklasse = **'KO'**

Wenn Wf am Satzanfang steht oder davor ein Satzzeichen steht

und wenn rechts von der Wf **KEINE** Wf steht mit der Wortklasse **'VE'**:

dann Wortklasse = **'KO'**

sonst:

Wortklasse = **'AV'**

**16. aX** Homograph Adverb - Verbzusatz

Beispiel: 'vorwärts', 'beiseite'

Wenn Wf am Satzanfang steht oder davor ein Satzzeichen steht:

dann Wortklasse = **'AV'**

sonst:

Wortklasse = **'AX'**

**17. Ps** Homograph Adverb -Personalpronomen

Beispiel: 'selbst'

Wenn links von der Wf eine Wf steht mit der Wortklasse

< **PE, P1, P2, P3, P4, P5, p5** >

dann Wortklasse = **'PR'**

sonst Wortklasse = **'AV'**

**18. pK** Homograph Konjunktion - Präposition

Beispiel: 'seit', 'bis'

Wenn Wf am Satzanfang steht oder davor ein Satzzeichen steht

und wenn rechts von der Wf eine Wortform mit **einer anderen** als

einer der folgenden Wortklassen steht:

< **PR, PP, NU, AV, AD, SU, VP, VA, PE, P5, P3** >

und die rechtsstehende Wf **nicht** die Grundform < **'zu'** > hat,

dann Wortklasse = **'KO'**

sonst:

Wortklasse = **'PP'**

**19. Pz** Homograph Präposition – Verbzusatz - Konjunktion

Beispiel: 'zu'

Steht Wf. nach einem Wort mit der Wortklasse **'PR', 'AV'**

dann Wortklasse = **'AV'**

Steht Wf. vor einem Wort mit der Wortklasse **'VE'** und dem FlexCode **'G6'**

oder der Wortform **'sein'**

dann Wortklasse = **'KO'**

Steht Wf. vor einem Wort mit der Grundform '**wenig**' oder '**viel**'

dann Wortklasse = '**AV**'

Steht nach Wf. ein Wort mit der Wortklasse '**NU**',

dann Wortklasse = '**PP**'

sonst:

Wortklasse = '**PX**'

**20. pX** Homograph Präposition - Verbzusatz

Beispiel: 'an', 'vor'

Wenn Wf am Satzanfang steht oder **davor** ein Satzzeichen steht

**oder**

wenn rechts von der Wf eine Wf steht mit der Wortklasse '**AV**'

dann Wortklasse = '**PP**'

Wenn rechts von der Wf eine Wortform steht mit einer der folgenden

Wortklassen:

**<AD, VA, VP, SU, NU, PE, P5, P3, PR>**

dann Wortklasse = '**PP**'

sonst:

Wortklasse = '**VX**'

**21. p8** Homograph Präposition - Adjektiv

Beispiel: 'nahe'

Wenn rechts von der Wf eine Wortform steht mit einer der folgenden  
Wortklassen:

**< NU, PE, P5, PR >** oder '**von**'

dann Wortklasse = '**PP**'

sonst:

Wortklasse = '**AD**'

**22. p9** Homograph Präposition - Adverb

Beispiel: seitlich

Wenn rechts von der Wf '**von**' steht oder eine Wortform mit einer der  
folgenden Wortklassen steht:

**<PR, PP>**

dann Wortklasse = '**PP**'

sonst:

Wortklasse = '**AV**'