

Das KiezDeutsch-Korpus (KiDKo)

(Universität Potsdam, SFB 632, Projekt B6: Analysen an der Peripherie)

Transkriptionsrichtlinien

Stand: November 2014

Das vorliegende Dokument beschreibt die Transkriptions- und Normalisierungsrichtlinien, die bei der Erstellung des KiezDeutsch-Korpus (KiDKo) angewendet wurden.

Das KiezDeutsch-Korpus (KiDKo) beinhaltet Audioaufnahmen, die in der zweiten Projektphase des SFB632 im Projekt B6 „Analysen an der Peripherie“ zwischen 2008 und 2011 aufgenommen und im EXMARaLDA Partitur Editor (Schmidt/Wörner 2009) transkribiert wurden. Die Transkriptionen folgen den Regeln des GAT2-Minimaltranskripts (Gesprächsanalytisches Transkriptionssystem, Selting et al. 2009).

Die Daten wurden orthografisch (d.h. nach den orthografischen Regeln des Deutschen) transkribiert. Es herrscht jedoch strikte Kleinschreibung, da Großbuchstaben für die Kennzeichnung von Akzenten reserviert sind. Auch wenn die Transkription in erster Linie orthografisch erfolgt ist, gibt es keine Korrekturen oder Anpassungen an die Standardsprache. Vielmehr wurde möglichst getreu der gesprochene Wortlaut verschriftet, um Aussprachebesonderheiten zu erfassen. Grundsätzlich wurde darauf geachtet, dass eine möglichst genaue Wiedergabe des Gesprächsablaufs erfolgt und dass stets eine eindeutige Sprecher-Äußerung-Zuordnung erkennbar ist. Sprecherwechsel, Überlappungen und gleichzeitiges Sprechen sind durch parallele Notation in den einzelnen Sprecherebenen kenntlich gemacht.

Hinzu kommt die **orthografische Normalisierung** (Normalisierungsebene, im folgenden meist abgekürzt zu "n") der Transkriptionsebene (Transkriptionsebene, im folgenden meist abgekürzt zu "v"). Bei der orthografischen Normalisierung handelt es sich innerhalb der Transkriptionsdateien um eine neue, separate Ebene, die mit den ursprünglichen Transkriptionen aligniert ist. Im Unterschied zur Transkriptionsebene, die keine Satzsegmentierung oder Interpunktion enthält, wurden hier Interpunktion und Großschreibung nach orthografischen Regeln eingesetzt. Normalisiert wird also ausschließlich auf Wort- und Zeichensetzungsebene, nicht aber bspw. auf Ebene der Wortstellung. Die orthografische Normalisierung stellte die Grundlage dar für alle weiteren automatischen, also technisch gestützten Aufbereitungsschritte, wie bspw. die (semi-)automatische Annotation von Wortarten etc.

Auf der Ebene "Macro" wurden Äußerungen markiert, die ungrammatikalisch sind, also z.B. Reperaturen enthalten. Diese Annotationsebene enthält nur ein einziges Tag, KOR. Des weiteren wurde türkisches Sprachmaterial im Korpus transkribiert und übersetzt.

Die folgende Tabelle zeigt die in KiDKo (Release 1.0) vorhandenen Transkriptions- und Annotationsebenen:

Kürzel	Beschreibung
nv	nonverbale Ebene
v	Transkriptionsebene (verbale Ebene)
n	Normalisierungsebene
POS	Wortartenebene (part-of-speech)
macro	markiert Reperaturen etc.
tr	türkische Transkriptionsebene
trnorm	Normalisierung
trdtwwue	deutsche Übersetzung (Wort für Wort)
trdtue	deutsche Übersetzung (frei)

Table of Contents

Das KiezDeutsch-Korpus (KiDKo).....	1
Transkriptionsrichtlinien.....	1
1 Transkription/Normalisierung allgemein	5
1.1 Auslassungen/Wortwiederholungen/Abbrüche	5
1.2 Silbifizierung	5
1.2.1 Kanonische Silbifizierung	5
1.2.2 Markierte Silbifizierung	6
1.3 Glottisverschlusslaut/Apostroph	6
1.4 Aussprachevariation	6
1.4.1 Reduktionsformen/Verschmelzungen	6
1.4.2 Nebensilbenvokale	8
1.4.3 r- bzw. er-Vokalisierung	8
1.4.4 Erste Person Singular	9
1.4.5 Bildung des Imperativs	9
1.4.6 Weitere Varianten	10
1.4.7 Entscheidungshilfe: Normalisierung vs. Fehlerkorrektur	10
1.5 Komplexe Wortformen.....	11
1.5.1 Komposita	11
1.5.2 Andere komplexe Wortformen.....	12
1.6 Zahlen und Datumsangaben	12
1.7 Abkürzungen	13
1.8 Eigennamen/Markennamen	14
1.9 Orthografische Schreibweisen	14
1.10 Pausen	15
1.11 Dehnung	16
Doppelung von vokalischen Silben	16
1.12 Außersprachliche Handlungen/Ereignisse	17
1.13 Unverständliche/schwer verständliche Passagen	17
1.14 Uneindeutiges Material	18
1.15 Untranskribiertes Material	18
1.16 Fremdsprachliches Material	19
1.16.1 Fremdsprachliches Material allgemein	19
1.16.2 Fremdsprachliches Material: Englisch	19
1.17 Zu anonymisierendes Material	20
1.18 Geräusche/non-verbales Material	23
1.18.1 Geräusche/non-verbales Material mit Funktion innerhalb des Segments	23
1.18.2 Geräusche/non-verbales Material ohne Funktion innerhalb des Segments	23
2 Segmentierung und Akzentsetzung.....	24
2.1 Akzentsetzung.....	24
2.2 Segmentierung.....	24
2.2.1 Punkt.....	25
2.2.2 Ausrufezeichen.....	27
2.2.3 Allgemein: Differenzierung.....	29
2.2.4 Fragezeichen.....	31
2.2.5 Komma.....	32
2.2.6 Doppelpunkt und Anführungszeichen.....	36
2.2.7 Unspezifisches Grenzzeichen (#).....	36
2.2.8 Allgemein.....	37

3 Partikeln und Signale.....	38
3.1 Diskursmarker/Diskurspartikeln.....	38
3.2 Fragepartikeln.....	40
3.3 Antwortpartikeln.....	40
3.4 Modalpartikeln/Abtönungspartikeln.....	41
3.5 Gefüllte Pausen.....	42
3.6 Rezeptionssignale.....	42
3.7 Interjektionen.....	43
3.8 Partikeln: allgemeine Hinweise.....	45
4 Phänomene gesprochener Sprache.....	46
4.1 Reihung.....	46
4.2 Wiederholung.....	46
4.3 Reparatur/Korrektur.....	47
4.4 Abbruch.....	48
4.5 Spezialfall: Wortabbruch.....	49
4.6 Korrektur vs. Abbruch.....	50
5 Besonderheiten bei der Normalisierung.....	50
5.1 Normalisierungsbeispiele (Liste).....	50
5.2 Einwortäußerungen.....	50
5.2.1 Normalisierung von Einwortäußerungen.....	50
5.2.2 Einwortäußerungen und Akzent.....	51
5.3 Segmentierung von nebenordnenden/unterordnenden Konjunktionen.....	52
5.4 Direkte Rede/Zitation/Bericht.....	53
5.4.1 Direkte Rede.....	53
5.4.2 Zitation.....	55
5.4.3 Bericht eigener Äußerungen/Selbstreflexion.....	55
5.4.4 Sprecher/in wiederholt Material von Vorredner/in.....	55
5.5 Unterbrechungen durch Pausen.....	56
5.5.1 Unterbrechung von Wörtern.....	56
5.5.2 Pausen innerhalb einer Äußerung.....	56
5.6 Groß-/Kleinschreibung.....	56
Referenzen.....	58

1 Transkription/Normalisierung allgemein

1.1 Auslassungen/Wortwiederholungen/Abbrüche

Fehlende Wörter werden weder auf der Transkriptionsebenen (v), noch auf der Normalisierungsebene (n) ergänzt.

v: ich war gestern MediAmarkt
n: Ich war gestern MediAmarkt .

v: müssen morgen auf ARbeit
n: Müssen morgen auf Arbeit .

Wortwiederholungen auf Wort- oder Phrasenebene werden nicht vereinfacht. Das wiederholte Material wird auf n durch Komma getrennt.

v: wir müssen müssen UNbedingt noch ins kino
n: Wir müssen , müssen unbedingt noch ins Kino .

v: du wolltest doch du wolltest doch mal wieder ins KIno
n: Du wolltest doch , du wolltest doch mal wieder ins Kino .

Onomatopoetika/Echoismen werden als einzelne Token transkribiert.

v: gutschI gutschI GUTschi
n: GutschI , gutschI , gutschI !

Kürzere Formen wie die folgenden können als ein Token transkribiert werden.

v: HAha
n: Haha !

v: EIEiei
n: Eieiei !

Abbrüche während einer Äußerung werden auf n durch # gekennzeichnet.

v: du hast doch er hat doch das AUto geschrottet
n: Du hast doch # Er hat doch das Auto geschrottet .

v: er soll er wollte doch noch zum SPORT
n: Er soll # Er wollte doch noch zum Sport .

1.2 Silbifizierung

1.2.1 Kanonische Silbifizierung

Die Markierung von Akzenten auf v folgt den deutschen Silbifizierungsregeln. Hier einige Beispiele:

v:	KLASsensprecher	SCHWÖRS	LEHrer
	TASche	muSEum	WEIßT
	iTAlien	MÖCkerbrücke	SITzen
	geWONnen	ECKe	

1.2.2 Markierte Silbifizierung

Innerhalb eines Events wird "=" nur bei einer **markierten Silbifizierung innerhalb eines Wortes** verwendet, d.h. wenn eine Silbengrenze an einer Position gezogen wird, an der im unmarkierten Fall keine ist.

v: das glas war LEER
n: Das Glas war leer .

v: das glas war LE=er
n: Das Glas war leer .

v: ich hab geSCHRIEN
n: Ich habe geschrien .

v: ich hab geSCHRI=en
n: Ich habe geschrien .

1.3 Glottisverschlusslaut/Apostroph

Auf der Transkriptionsebene wird das Apostroph verwendet, um bei bestimmten Ausdrücken anzuzeigen, dass diese mit einem Glottisverschlusslaut bzw. Glottisschlag einsetzen. Dies betrifft in erster Linie bejahende bzw. Rezeption signalisierende und verneinende Partikeln.

Verneinung

v: 'hm'hm
n: Mh-mh .

Bejahung/Rezeptionssignal

v: 'hmlhm
n: Hm-Hm .

1.4 Aussprachevariation

Aussprachevarianten, die auch im Standarddeutschen üblich sind, werden **nicht** auf der Transkriptionsebene erfasst (z.B. könich -> könig, wenich -> wenig).

1.4.1 Reduktionsformen/Verschmelzungen

Auslassungen von Lauten oder Silben am Anfang, in der Mitte oder am Ende eines Wortes werden nicht ergänzt. Für die Auslassungen werden keine Apostrophe gesetzt.

nich vs. nicht	kein vs. keinen
jetz vs. jetzt	n vs. ein
is vs. ist	ne vs. eine
ma vs. mal	was vs. etwas
brauch vs. braucht	rum vs. herum

Werden zwei Morpheme nacheinander mit einem **schnellen Anschluss** realisiert *und* ist eine der beiden Formen lautlich **reduziert**, so wird dies mit "=" vor der reduzierten Form angezeigt.

bist du vs. bist =e
ist er vs. is= er
haben sie vs. ham= se usw.

Hier ein weiteres Beispiel mit Normalisierung:

v: weißt =e wie GEIL der aussieht
n: Weißt du , wie geil der aussieht ?

Fehlt eine schnell angeschlossene, reduzierte pronominale Form (Klitikon), wird diese auf Normalisierung auch nicht ergänzt. Es wird nur annotiert, was an der Oberfläche sichtbar ist.

v: weißt wie GEIL der aussieht
n: Weißt , wie geil der aussieht ?

Sind beide Formen bei einem schnellen Anschluss reduziert, wird die Reduktion durch "=" **immer** am zweiten Morphem angezeigt.

(wir) haben es vs. ham =s
(sie) sollen es vs. solln =s usw.

Nicht jede phonetisch reduzierte Form muss zwingend schnell angeschossen sein und daher auch nicht mit "=" markiert werden. Das kann bspw. die reduzierte Form des unbestimmten Artikels betreffen. Diese Entscheidung muss gemäß dem jeweiligen lautlichen Kontext getroffen werden.

v: da hast =e so n GROßes fruchtzweg
n: Da hast du so ein großes Fruchtzweg

v: wenn da so n Kilo
n: wenn da so ein Kilo

Wenn *so ein(e)* eine wohlgeformte Äußerung ergibt, wird *so =ne* zu *so ein(e)* normalisiert.

v: **so =ne** doofe TANTe
n: **so eine** doofe Tante

Ergibt die Konstruktion keine wohlgeformte Äußerung (z.B. mit Pluralformen), wird die Form *ne* beibehalten und mit *so* zu *some* verschmolzen.

v: dann hatten wir **some** SCHOkokekse da mit
n: Dann hatten wir **some** Schokokekse da mit .

Bei der **Verschmelzung von Präposition und Artikel** erfolgt **keine** Markierung des schnellen Anschlusses + reduzierter Form. Sie wird ohne "=" direkt verschriftet.

auf dem vs. aufm
bei dem vs. beim
zu dem vs. zum usw.

Verschmelzungen von Präposition und Artikel werden sowohl auf v als auch auf n als ein Token analysiert und in ein Event geschrieben.

v: zum | beim | aufm | im | mitm | untern | hinterm | vorm | inne | ...
n: zum | beim | aufm | im | mitm | untern | hinterm | vorm | inne | ...

Einen besonderen Fall stellt hier das Aufeinandertreffen der Präposition *in* mit dem indefiniten und dem definiten Artikel dar. Auch wenn sich die phonetische Form häufig gleich anhört, werden hier für die Transkription und die Normalisierung unterschiedliche Formen zugrunde gelegt: Wenn *in* mit dem definiten Artikel verschmilzt, wird es, wie die obigen Fälle, als ein Token analysiert, im Fall des indefiniten Artikels als zwei.

in + indef. Artikel

v: ich bin in =ner schwierigen Lage
n: Ich bin in einer schwierigen Lage .

in + def. Artikel

v: ich bin inner SCHWEIZ
n: Ich bin inner Schweiz .

DAT (wo ist sie?)

fem. inner Schule
masc. im Turm
neut. im Haus

AKK (wohin geht sie?)

inne Schule
inn Turm
ins Haus

1.4.2 Nebensilbenvokale

Die **Abschwächung bzw. Auslassung von unbetonten Vokalen** wird verschriftet, allerdings mit Augenmaß, d.h. ein Wort soll noch erkennbar sein (*mittnanda* für *miteinander* wäre übertrieben).

Das Schwa wird verschriftet, wenn sich dadurch die Silbenanzahl des Wortes nicht verändert:

wolln (2) vs. **wollen** (2)
machn (2) vs. **machen** (2)
müssn (2) vs. **müssen** (2)
sagn (2) vs. **sagen** (2)
erwischn (3) vs. **erwischen** (3)
redn (2) vs. **reden** (2)

Das Schwa wird nicht verschriftet, wenn sich dadurch die Silbenanzahl des Wortes ändert (z.B. wenn durch die Verschriftung des Schwa eine zusätzliche Silbe eingefügt würde):

gehn (1) vs. gehen (2)
stehn (1) vs. stehen (2)
verstehn (2) vs. verstehen (3)
aufhörn (2) vs. aufhören (3)
gesehn (2) vs. gesehen (3)
hörn (1) vs. hören (2)

In Fällen wie dem folgenden empfiehlt sich die Verschriftung des Schwa:

ham vs. **haben**
willkomm vs. **Willkommen** (etc.)

1.4.3 r- bzw. er-Vokalisierung

Die nahezu standardsprachliche **Vokalisierung von <r> bzw. <er>** wird nur im Ausnahmefall, d.h. bei großer Auffälligkeit, notiert.

lehrer, **nicht:** lehra
berlin, **nicht:** balin
wieder, **nicht:** wieda
für, **nicht:** füa
aber, **nicht:** aba
klar, **nicht:** kla:

1.4.4 Erste Person Singular

Das Schwa wird in unbetonten Endsilben in gesprochener Sprache häufig getilgt. Auf Normalisierungsebene wird es jedoch ergänzt!

v: ich HEUL ich ZEIG ich SCHWÖR ich SCHLAF
n: Ich heule . Ich zeige . Ich schwöre . Ich schlafe .

1.4.5 Bildung des Imperativs

Die Formen des Imperativs werden bei regelmäßigen und unregelmäßigen Verben gleich gebildet.

Singular: Stamm (+ e)
Plural: Stamm + t

v: SCHREI nicht so v: SCHREIT nicht so v: WARte auf mich
n: Schrei nicht so ! n: Schreit nicht so ! n: Warte auf mich!

Wenn auf den Verbstamm *t* oder *st* folgt, muss bei gewissen Verben ein *e* eingefügt werden. Diese **e-Erweiterung** geschieht bei den folgenden Verben:

- alle Verben, deren Stamm auf *d* oder *t* enden, z.B.: *retten, reden, laden, weiden*

v: rettet mich nicht
n: Rettet mich nicht !

- alle Verben, deren Stamm auf einen Verschlusslaut oder Reibelaut + *m* oder *n* enden, z.B.: *atmen, rechnen, widmen, wappnen*

v: ATmet alle ein
n: Atmet alle ein !

Einige unregelmäßige Verben bilden die 2. und 3. Person Singular Präsens sowie den Imperativ Singular, indem sie das *e* des Präsensstammes durch *i* oder *ie* ersetzen, der sog. **e/i-Wechsel bei unregelmäßigen Verben**: *befehlen, sehen, empfehlen, bergen, essen, helfen, ...*

v: seh mich nicht so an
n: Sieh mich nicht so an !

Schwankendes e: Die Singularendung *e* kann oft auch weggelassen werden. Ausnahmen sind aber zum Beispiel:

- immer mit *e* bei Verben, deren Stamm auf einen Verschluss- oder Reibelaut + *m* oder *n* endet (vgl. oben "e-Erweiterung"): *atme!, rechne!, widme!, wappne!*

v: ATme nicht so laut
n: Atme nicht so laut !

- meistens mit *e* bei Verben, deren Stamm auf *d* oder *t* endet: *rede!, lade!, rette!, verbiete!*

v: RED nicht so laut
n: Rede nicht so laut !

- Immer ohne *e* bei Verben mit *e/i*-Wechsel im Präsens (vgl. e/i-Wechsel)

v: ESS nicht so laut
n: Iss nicht so laut !

1.4.6 Weitere Varianten

Der vom **Standard abweichende Sprachgebrauch** einiger Sprecher/innen mit regionalen, individuellen oder ethnolektalen Besonderheiten wird nirgends als vermeintlicher "Fehler" korrigiert. Auf eine übertrieben genaue Verschriftlichung dieser Elemente wird jedoch verzichtet.

umgangssprachlich:

jetze vs. jetzt
ebent vs. eben
fümf vs. fünf
tachchen usw.

dialektal:

ick/icke vs. ich
dit/dat vs. das
ölf vs. elf
jut vs. gut
ko::fen vs. kaufen
alle::ne vs. alleine usw.

ethnolektal:

isch vs. ich
dis/des vs. das
sog. "feste Wörter": ischwör, ich schwöre usw.

Wir transkribieren **keine Unterscheidung zwischen stimmhaftem und stimmlosem /s/**.

Exxx [v]: SCHEI:ße
Exxx [n]: Scheiße !

Bei starken Auffälligkeiten zwischen **stimmhaften/stimmlosen Plosiven** wird dies auf der Transkriptionsebene annotiert.

MuH3WT [v]: die SPAnier seh voll KEIL aus
MuH3WT [n]: Die Spanier sehen voll geil aus .

1.4.7 Entscheidungshilfe: Normalisierung vs. Fehlerkorrektur

Die Grenze zwischen der orthografischen Normalisierung eines transkribierten Wortes und der Korrektur eines "Fehlers" ist manchmal nicht ganz klar. Wenn eine phonetische Form **mehrdeutig** ist, wird als Normalisierung die im Kontext kanonische orthografische Variante gewählt.

v: ich hab ein KIND gesehen | ich hab =n KIND gesehen
n: Ich habe ein Kind gesehen . | Ich habe ein Kind gesehen .

v: ich hab ein JUNgen gesehen | ich hab =n JUNgen gesehen
n: Ich habe einen Jungen gesehen . | Ich hab einen Jungen gesehen .

nicht:

n: Ich habe ein Jungen gesehen . | Ich hab ein Jungen gesehen .
→ gesprochenes ein im Akkusativ sowohl vor Neutra als auch vor Maskulina

v: KEIN einzigen (-) EURO
n: keinen einzigen Euro

Wenn eine phonetische Form **eindeutig** ist, wird auf die entsprechende orthografische Form normalisiert, auch wenn sie im Kontext u.U. nicht kanonisch ist.

v: ich hab ein mann ein BUCH gegeben
n: Ich habe ein Mann ein Buch gegeben .

nicht:

n: Ich habe einem Mann ein Buch gegeben .

v: und WIE will hxxx dass du wie viel BEI ihn ABgibst
n: Und wie will Hxxx , dass du wie viel bei ihm abgibst .

nicht:

n: Und wie will Hxxx , dass du wie viel bei ihm abgibst .

v: den du selber von diesn geld verWALtn kannst
n: den du selber von diesen Geld verwalten kannst

Wenn bspw. durch **unverständliches Material** die kanonische Form nicht disambiguiert werden kann, wird im Fall des indefiniten Artikels der Nominativ Singular gewählt.

v: du hast =n (unverständlich)
n: Du hast ein #

→ hier wird immer Nominativ Singular annotiert: *ein*

1.5 Komplexe Wortformen

1.5.1 Komposita

Komposita werden in den meisten Fällen zusammengeschrieben.

v: wir gehn ins LEHrerzimmer
n: Wir gehen ins Lehrerzimmer .

Es gibt einige Ausnahmen, in denen Komposita mit **Bindestrich** getrennt werden. Dazu gehören Komposita, die Abkürzungen, Eigennamen oder Zahlen enthalten. Hier wird innerhalb eines Events auf v mit **Unterstrich** "_" und auf n mit Bindestrich "-" segmentiert. Das gilt generell für Events, in denen eine Trennung durch Bindestrich erfolgt.

v: MAthe_emesa:_prüfung
n: Mathe-MSA-Prüfung

v: weidig_gymnasium
n: Weidig-Gymnasium

v: halt deine FRESse du ein_meter_ZWANzig
n: Halt deine Fresse , du Ein-Meter-Zwanzig !

v: mercedes_eselfka:
n: Mercedes-Sl1K

Bei der **Koordination von Komposita** wird das Kompositionserstglied in Übereinstimmung auf v und n mit Bindestrich am Ende transkribiert.

v: ich entscheide mich zwischen haupt- und REALschule
n: Ich entscheide mich zwischen Haupt- und Realschule .

Ein Beispiel aus dem Korpus:

MuH1WD [v]: das ist dann eigentlich alles (-) für ESsen weil wir müssen uns

MuH1WD [n]: Das ist dann eigentlich alles für Essen weil wir müssen uns

MuH1WD [v]: doch mittag- und abendessen SELber kaufen

MuH1WD [n]: doch Mittag- und Abendessen selber kaufen .

Eine '**Unterbrechung von Komposita**' kann auch durch (gefüllte oder ungefüllte) Pausen (s. Abschnitt 1.10) oder andere Planungszeichen entstehen. Diese Fälle werden wie die Koordination von Komposita behandelt (siehe jedoch auch Abschnitt 1.10 zu Pausen innerhalb von Wortformen).

Korpusbeleg:

v: JA war KLASsen- (-) äh sprecher- tre trä TREFFen

n: Ja . War Klassen- äh Sprecher- Tre Trä Treffen .

1.5.2 Andere komplexe Wortformen

Kompositionen aus **Nomen und Adjektiv** werden als ein Token transkribiert.

arschkalt, saumüde, hundemüde, arschteuer, hammergeil

Kompositionen aus **Adjektiven** und Kompositionen aus **Präposition und Adjektiv** werden ebenfalls als ein Token transkribiert.

supertoll, übergeil

Partikelverben werden ebenfalls als ein Token transkribiert.

leidtun, weggehen

1.6 Zahlen und Datumsangaben

Zahlen werden auf Transkriptionsebene stets ausgeschrieben, d.h. als Wort notiert. Auf Normalisierungsebene werden **Zahlen bis 12** ausgeschrieben, darüber als Ziffern wiedergegeben.

v: drei vierzehn neunzehnhundertdreiundachtzig

n: drei 14 1983

v: um zwölf uhr fümunvierzig

n: um zwölf Uhr 45

Wir treffen uns um zwölf . → Kleinschreibung!

Wir treffen uns um 13 Uhr .

Fußballergebnisse etc. werden ebenfalls auf Transkriptionsebene ausgeschrieben (ein Event; mit oder ohne Präposition). Auf Normalisierungsebene werden die Zahlen als Ziffern und in der Doppelpunkt-Schreibweise wiedergegeben.

v: einnull / einzunull → 1 Event!

n: 1:0 → 1 Event!

Ein Korpusbeleg:

MuP6MD [v]: FÜNFzich also BIS zu fünfzich euro

MuP6MD [n]: 50 . Also bis zu 50 Euro .

Bei Zahlenwerten mit mehr als 4 Stellen wird zur besseren Lesbarkeit ein Punkt eingefügt (aber nicht segmentiert!).

v:	zehntausend		hunderttausend		einemilliondreihunderttausend
n:	10.000		100.000		1.300.000

Handelt es sich um **komplexe Zahlen** (bspw. Dezimalzahlen), werden diese **ohne** eine interne Segmentierung ebenfalls als ein Token in ein Event eingetragen. Auf Normalisierungsebene werden sie unabhängig von ihrer Größe immer als Ziffern wiedergegeben, getrennt durch ein Komma.

v:	zweikommasix		vierzehnkommanueun
n:	2,6		14,9

Eine Ausnahme bei der Segmentierung sind **Brüche**. Diese werden auf Transkriptionsebene zwar ebenfalls ausgeschrieben, aber in mehrere Events gesplittet. Unabhängig von ihrer Größe werden sie auch auf der Normalisierungsebene ausgeschrieben.

v:	ein drittel		sieben achtel
n:	ein Drittel		sieben Achtel

Schulnoten werden auf der Transkriptionsebene und auf der Normalisierungsebene als zwei Events verschriftlicht.

v:	isch hab		ne drei MINus bekomm
n:	Ich habe eine Drei minus bekommen		.

Datumsangaben werden auf der Transkriptionsebene ausgeschrieben. Auf der Normalisierungsebene werden die Zahlwerte verwendet. Endet das Segment auf ein Datum, so wird nur ein Punkt (nämlich das Satzendezeichen) gesetzt.

v:	heut	ist	der	ELFte	sechste
n:	Heute	ist	der	11.	6.

Ein Korpusbeleg:

MuH1WD [v]:	DONnerstag	zwölfter	sechster
MuH1WD [n]:	Donnerstag	12.	6.

1.7 Abkürzungen

Schriftsprachlich gängige **Abkürzungen und Zeichen** werden als Wort geschrieben, z.B.:

v:	oke:	prozent	zum beispiel	zettde:eff	be:emwe:
n:	okay	Prozent	zum Beispiel	ZDF	BMW

Ein Korpusbeleg:

Spk2 [v]:	ich mach	vielleicht	SPÄter	erst	emesa:
Spk2 [n]:	Ich mache	vielleicht	später	erst	MSA

Ähnlich wird auch mit Firmen- und Produktbezeichnungen verfahren. Diese werden als ein Event verschriftlicht.

v:	tse:elfka:
n:	Cl1K

Wenn Firmen- oder Produktbezeichnungen innerhalb eines Kompositums auftreten, steht dieses ebenfalls in nur einem Event, es wird aber mit Unterstrich (v) bzw. Bindestrich (n) getrennt.

v: mercedes_eselfka:
n: Mercedes-S11K

1.8 Eigennamen/Markennamen

Die orthografische Form von feststehenden Eigennamen/Produktnamen wird auf der Normalisierungsebene beibehalten, auch wenn dies möglicherweise bestimmten Normalisierungsregeln oder Rechtschreibregeln widerspricht. Enthalten feststehende Markennamen ein Apostroph (bspw. im Genitiv), so wird dies auf der Normalisierungsebene übernommen.

v: mein BRU:der war bei kaisers
n: Mein Bruder war bei Kaiser's .

v: lass mal zu mcdONALDS
n: Lass mal zu McDonald's !

Das Genitiv-Apostroph wird wirklich nur in den genannten Fällen gesetzt.

Marias Fahrrad → v: m3xxxxs
n: M3xxxxs

KEIN Apostroph!

1.9 Orthografische Schreibweisen

Die Schreibweise in folgenden Fällen richtet sich nach den **orthografischen Regeln**:

- <ä> wird gemäß den orthografischen Regeln verwendet (wird nicht als <e> oder <eh> notiert)
- gesprochenes <sch> in den Verbindungen <st> und <sp> wird nicht notiert
- Vereinfachung der Affrikate <pf> wird nicht notiert (<pf> nicht als <f>)
- Verwendung von <v> und <f> gemäß den Regeln der Orthografie
- Verwendung von <ei>, <ai>, <ey>, <ay>, <äu> und <eu> gemäß den Regeln der Orthografie

Groß- und Kleinschreibung: Als Nomen verwendete Wortformen anderer Wortklassen

Wörter aus anderen Wortklassen werden großgeschrieben, wenn sie als Nomen verwendet werden:

- Als Nomen verwendete Adjektive und adjektivische Partizipien werden großgeschrieben.

das **Gute** im Menschen
Süßes und **Salziges**
das einzig **Richtige**

Aber: Adjektive und adjektivische Partizipien werden kleingeschrieben, wenn sie sich auf ein vorangehendes oder ein folgendes Nomen beziehen, auch wenn sie mit einem Artikel(wort) stehen.

Das **grüne** ist mein Fahrrad.
Den fünften Witz fand ich noch lustig, den **zehnten** nicht mehr.
Die **guten** ins Töpfchen, die **schlechten** ins Kröpfchen.

- Superlativformen, die mit *am* gebildet werden und nach denen mit *wie?* gefragt werden kann, werden kleingeschrieben.

So ist es am **besten** (wie?).
Am **nötigsten** (wie?) brauchen sie Zelte und warme Decken.
Sie hat die Aufgabe am **schnellsten** (wie?) gelöst.

Aber: wenn nicht mit *wie?* gefragt werden kann, muss großgeschrieben werden. In der Regel kann dann *am* durch *an dem* ersetzt werden.

Es fehlt ihnen am **Nötigsten** (woran? an dem Nötigsten).

- Viele Sprachbezeichnungen sind aus Adjektiven entstanden. Wenn sie die Sprache selber bezeichnen, werden sie großgeschrieben.

Mein **Spanisch** ist leider nicht so gut.
etwas auf **Deutsch** sagen

- Als Nomen verwendete Infinitive werden großgeschrieben.

Das **Reden** fiel ihm schwer.
Wir werden euch beim **Umziehen** helfen.
Das ist gar nicht zum **Lachen**.

- Pronomen, die als Stellvertreter für ein Nomen verwendet werden, werden kleingeschrieben.

Das weiß **jeder** . Das haben **alle** gewusst.
Das ist **alles** nicht so wichtig.
Ich kenne hier **jeden/niemanden/keinen** .
Das hat mich **einiges** gekostet.

- Als Nomen verwendete Adverbien, Präpositionen, Konjunktionen und Interjektionen werden großgeschrieben.

In diesem **Durcheinander** kann man nichts finden. (Adverb)
Leben Sie im **Hier** und **Jetzt** ! (Adverbien)
Überleg dir das **Für** und **Wider** ganz genau! (Präpositionen)
Bei der Begrüßung sollte man seinem **Gegenüber** freundlich und offen entgegentreten. (Präposition)
Sein ständiges **Wenn** und **Aber** geht mir auf die Nerven. (Konjunktionen)

1.10 Pausen

Wir unterscheiden bei der Annotation von **Pausen** drei verschiedene Längen. Markiert werden Pausen nur auf der Transkriptionsebene. Auf Normalisierungsebene haben sie als Entsprechung ein leeres Event.

Pause		Zeichen	Beispiel
kurz:	≤ 1 Sekunde	(-)	v: Plötzlich (-) is sie RAUSgerannt n: Plötzlich ist sie rausgerannt .
lang:	1-3 Sekunden	(--)	v: Plötzlich (--) is sie RAUSgerannt n: Plötzlich ist sie rausgerannt .
gemessen:	> 3 Sekunden (5.5)		v: Plötzlich (5.5) is sie RAUSgerannt n: Plötzlich ist sie rausgerannt .

Ein besonderer Fall sind **Pausen, die innerhalb von Wortformen auftreten** und das Wort daher in mehrere Bestandteile aufteilen. In diesen Fällen bekommt der erste Wortteil auf der Transkriptionsebene einen Bindestrich, auf den dann ein leeres Event und der zweite Bestandteil im nächsten Event folgen. Auf der Normalisierungsebene wird die gesamte Wortform in das Event des ersten Wortteils gesetzt, im Event unter dem zweiten Wortteil wird ein großgeschriebenes LEER gesetzt.

```
v:   ich hab die tür aufge-      (-) SCHLOSsen
n:   Ich habe die Tür aufgeschlossen   LEER   .

v:   warum hast du den schlüssel ver-      (-) LOren
n:   Warum hast du den Schlüssel verloren   LEER   ?
```

1.11 Dehnung

Bei der **Quantität von Vokalen** wird zwischen zwei Dehnungsstufen unterschieden. Längen werden mit Doppelpunkt markiert, keinesfalls mittels Buchstabendopplung oder Dehnungs-*h*.

Dehnung	Zeichen	Beispiel
länger als "normal"	:	v: so LA:NG war die stunde auch nich n: So lang war die Stunde auch nicht .
auffällig lang	::	v: DOCH sie war WA::HNSinnig lang n: Doch . Sie war wahnsinnig lang .

Im Rahmen der Orthografie werden lange/gespannte Vokale (abhängig vom Silbenaufbau) meist durch Verdoppelung des Vokals oder durch Dehnungs-*h* angezeigt. Diese Prinzipien bleiben im Rahmen der Transkriptionskonvention der Dehnung unberührt.

```
v:   der see gefällt mir SEHR
n:   Der See gefällt mir sehr .
```

Werden diese Vokale mit einer wahrnehmbar längeren Dehnung ausgesprochen, werden sie in der Transkription zusätzlich (zur orthografischen Kennzeichnung) mit : oder :: markiert. Der Doppelpunkt steht nach dem Vokal, nicht nach dem "h".

```
v:   der see: gefällt mir SE:HR
n:   Der See gefällt mir sehr .
```

Bei Wortformen/Konstruktionen, die primär in der gesprochenen Sprache vorkommen und für die im Rahmen dieser Konventionen die Verschriftung geregelt wurde (Interjektionen, gefüllte Pausen, Fragepartikeln und andere Signale), werden ebenfalls die orthografischen Prinzipien gewählt, um den Unterschied zwischen kurz/ungespannt und lang/gespannt anzuzeigen (ggf. dann noch mit zusätzlichem Dehnungszeichen).

```
v:   NEE:   das WILL ich nich machen
n:   Nein . Das will ich nicht machen .

v:   das muss ich jetz machen   ne::
n:   Das muss ich jetzt machen . Ne   ?
```

Doppelung von vokalischen Silben

Die Doppelung vokalischer Silben wird mit "=" (ggf. in Kombination mit dem Doppelpunkt als Dehnungszeichen) gekennzeichnet.

v: TSCHÜ:=üß
n: Tschüß

1.12 Außersprachliche Handlungen/Ereignisse

Außersprachliche Handlungen werden nur angegeben, wenn sie für die Äußerung relevant sind: lachen, seufzen, schreien, husten usw. Nicht relevant sind bspw. Hintergrundgeräusche ohne Bedeutung für das Gespräch. Notiert werden außersprachliche Handlungen auf einer unabhängigen globalen non-verbalen Ebene ([nv]).

```
[nv]                hustet
Spk1 [v] gib mir ma ne Zigarette
Spk1 [n] Gib mir mal eine Zigarette .
Spk2 [v]
Spk2 [n]
```

1.13 Unverständliche/schwer verständliche Passagen

Unverständliche Passagen einer Äußerung/eines Gesprächs werden mit "(unverständlich)" in einem Event direkt auf Transkriptionsebene eingetragen. Wenn der unverständliche Abschnitt länger als zwei Sekunden ist, wird die Zeit (in Sekunden) im folgenden Format mit eingetragen: "(unverständlich, 2.0)". Auf Normalisierungsebene hat dieser Eintrag ein leeres Event als Entsprechung. Entsteht durch den unverständlichen Abschnitt eine unvollständige/abgebrochene Äußerung, so wird **auf der Normalisierungsebene** mit # segmentiert. Die Raute wird nicht mit (unverständlich) aligniert, sondern bekommt ein neues Event und hat auf der Transkriptionsebene eine leere Entsprechung.

```
v:   ich war doch noch (unverständlich)   kannst du mir nächstes mal Sagen
n:   Ich war doch noch                       # Kannst du mir nächstes mal sagen .
```

```
v:   mein (unverständlich, 5.0)
n:   Mein                                     #
```

Korpusbelege:

```
Txxx [v]: (unverständlich)
Txxx [n]:                                     #
```

```
Spk1 [v]: (unverständlich, 3.5)
Spk1 [n]:                                     #
```

Lässt sich vor dem unverständlichen Abschnitt eine vollständige Äußerung identifizieren, so wird statt der Raute das entsprechende Satzzeichen gewählt (siehe Kap. 2 Segmentierung); (*unverständlich*) bleibt in die Äußerung mit eingeschlossen.

```
v:   ich war doch noch EINKaufen (unverständlich)   kannst du mir
n:   Ich war doch noch einkaufen                       . Kannst du mir
v:   nächstes mal Sagen
n:   nächstes mal sagen .
```

Das entsprechende Satzzeichen wird auch in Fällen vergeben, in denen sich trotz einer (kurzen) unverständlichen Passage eine kanonische syntaktische Struktur erschließen lässt.

```
v:   du WEIßT doch   wo gxxx (unverständlich) hat
n:   Du weißt doch , wo Gxxx                       hat .
```

v: na da sind deine (unverständlich) sehr sehr UNgeplant
n: Na da sind deine sehr , sehr ungeplant .

v: was hast DU (unverständlich) wieder gemacht
n: Was hast du wieder gemacht ?

Material, das nicht unverständlich ist, bei dem aber Unsicherheit besteht, weil das Gehörte **schwer zu verstehen** ist, wird auf der Transkriptionsebene in Klammern gesetzt (jedes Token einzeln). Diese Klammern werden jedoch nicht auf der Normalisierungsebene gesetzt.

[v]: (NXXX) (kommt) wieder
[n]: Nxxx kommt wieder .

Unverständliche Ortsangaben werden als **X0xxx** wiedergegeben.

v: X0XXX(straße)
n: X0xxx(Straße)

1.14 Uneindeutiges Material

Es kann bei der Transkription auch vorkommen, dass Sprachmaterial nicht eindeutig normalisiert werden kann, z.B. weil wenn es sich um ein unbekanntes Wort handelt, das aber auch nicht als fremdsprachlich zu identifizieren ist.

Ein Beispiel:

v: of VIESCH ey
n: Oh ! ??? . Ey !

Solche Fälle können immer nur im jeweiligen Kontext entschieden werden. Wenn der Kontext es zulässt, ist eine Normalisierung auf *Vieh* möglich.

v: of VIESCH ey
n: Oh ! Vieh . Ey !

Hilft der Kontext jedoch nicht bei der Disambiguierung, so gilt v = n mit angepasster Groß- und Kleinschreibung auf n: Wenn die Wortart nicht klar ist, dann Kleinschreibung.

v: of VIESCH ey
n: Oh ! Viesch . Ey !

→ hier Großschreibung, da satzinitial, ansonsten Kleinschreibung, da Wortart uneindeutig

1.15 Untranskribiertes Material

Wenn ein nicht transkribierter Abschnitt **länger als 1 Minute** ist, dann wird dies auf der Transkriptionsebene vermerkt mit: (*untranskribiert, ZEITANGABE*). Auf der Normalisierungsebene wird kein Segmentierungszeichen gesetzt.

v: dann ist er geGANgen (untranskribiert, 93.2) wir fliegen in Urlaub
n: Dann ist er gegangen . Wir fliegen in Urlaub .

→ 1 Minute, 33.2 Sekunden nicht transkribiertes Material

1.16 Fremdsprachliches Material

1.16.1 Fremdsprachliches Material allgemein

Fremdsprachliche Teile von Äußerungen/Gesprächspassagen werden auf der Transkriptionsebene durch (*fremdsprachlich*) markiert. Ist bekannt, um welche Sprache es sich handelt, so wird diese in die Klammern hinzugefügt, z.B.: (*fremdsprachlich, türkisch*). Ist nicht eindeutig klar, dass es sich hier um fremdsprachliches Material handelt, wird es als (*unverständlich*) markiert.

Auf der Normalisierungsebene hat (*fremdsprachlich*) ein leeres Event als Entsprechung. Wird eine Äußerung dadurch abgebrochen/unvollständig, so wird (wie bei (*unverständlich*)) mit # segmentiert, ansonsten mit dem dem Äußerungstypen entsprechenden Segmentierungszeichen.

```
v:    und er sagt zu MIR so (fremdsprachlich)    hab ich überHAUPT nicht
n:    Und er sagt zu mir so                        # Hab ich überhaupt nicht

v:    verstanden
n:    verstanden .

v:    und er sagt zu MIR den Satz (fremdsprachlich)    hab ich überHAUPT
n:    Und er sagt zu mir den Satz                      . Hab ich überhaupt

v:    nicht verstanden
n:    nicht verstanden .
```

Informationen zu fremdsprachlichem Material auf der non-verbalen Ebene und auf der Transkriptionsebene sind, um ihre Auffindbarkeit für Transkription und Übersetzung zu gewährleisten, zu folgenden Kommentarformen vereinheitlicht worden:

--> **auf nv:**

```
fremdsprachlich oder
fremdsprachlich, türkisch oder
fremdsprachlich, 8.6 oder
fremdsprachlich, türkisch, 8.6
```

--> **auf v:**

```
(fremdsprachlich) oder
(fremdsprachlich, türkisch) oder
(fremdsprachlich, 8.6) oder
(fremdsprachlich, türkisch, 8.6)
```

1.16.2 Fremdsprachliches Material: Englisch

Englisches Material, das in eine deutsche Äußerung integriert ist, wird auf der Transkriptionsebene nach englischer Orthografie verschriftet. Nomen werden jedoch in Anlehnung an die deutsche Orthografie großgeschrieben. Akzente werden gemäß der englischen Phonologie gekennzeichnet.

```
v:    ich hab einfach NAME gesagt
n:    Ich habe einfach Name gesagt .

v:    ich hab einfach BOOK gesagt
n:    Ich habe einfach Book gesagt .

v:    doktor DREI
n:    Doktor Drei .

v:    der heißt doctor DRE
n:    Der heißt Doktor Dre .
```

Wenn englische Fremdwörter (keine Lehnwörter) in einem deutschen Satz eine Position als Verb, Nomen, Adjektiv einnehmen, dann werden sie nach deutscher Orthografie verschriftet:

Ich habe ein Book gekauft . → Großschreibung

Sehr kurze unabhängige Äußerungen, die nicht in einen deutschen Satz integriert sind, werden nach den Regeln der englischen Orthografie verschriftet.

v: fuck YOU
n: Fuck you !

v: i love you
n: I love you .

Englischsprachige Sätze werden nach englischer Orthografie verschriftet.

I bought a book . → Kleinschreibung von Substantiven

Längere englische Passagen werden nicht verschriftlicht, sondern auf der Transkriptionsebene durch (fremdsprachlich, englisch) gekennzeichnet (das Gleiche gilt natürlich auch für andere Fremdsprachen). Auf der Normalisierungsebene markieren wir dies mittels *FOREIGN*, gefolgt von einem #.

MuH17MA [v]: (fremdsprachlich, englisch)
MuH17MA [n]: FOREIGN #

1.17 Zu anonymisierendes Material

Übersicht: Anonymisierung

Alle Personennamen und Namen von Schulen (ggf. auch von anderen Lokalitäten), die in den Aufnahmen vorkommen, wurden anonymisiert. Für die Anonymisierung von Personen, Schulen und anderen Ortsangaben wurde das unten beschriebene Schema verwendet.

1 Sprechersiglen der Ankersprecher/innen

Siglen der InformantInnen: Jede/r Informant/in bekommt eine Sigle, die sich aus den unten erläuterten Kürzeln zusammensetzt. Auch im Transkript wird diese Sigle anstelle des Namens des Informanten / der Informantin benutzt.

MO	01	W	D
Bezirk	fortlaufende Nr. pro Schule	Geschlecht	Familiensprache
Mu (multiethnischer Wohnbezirk)	01	W (weibl.)	T (Türkisch)
Mo (monoethnischer Wohnbezirk)	02	M (männl.)	A (Arabisch)
	...		K (Kurdisch)
			D (Deutsch)

Bei der multiethnischen Gruppe (Mu) gibt es zwei Nummerierungen, unterschieden durch den Zusatz „H“ bzw. „P“, der auf unterschiedliche Durchläufe einer anderen Studie (Wiese et al. 2009) verweist, an der die Informant/inn/en teilgenommen haben. Diese Siglen haben beispielsweise eine Struktur wie MuH9WT. Das Merkmal H/P wird aller Voraussicht nach nicht langfristig Bestandteil der Siglen sein.

2 Begleitsprecher/innen

Die Anonymisierung der Gesprächspartner/innen (Begleitsprecher/innen) der Ankersprecher/innen erfolgt nach folgendem Schema:

Namen wurden durch das Kürzel SPK ersetzt und durchnummeriert, wobei die Annahme getroffen wurde, dass es sich bei Personen gleichen Namens, die jedoch in den Aufnahmen unterschiedlicher Ankersprecher/innen zu hören sind, nicht um die gleiche Person handelt. In solchen Fällen wurden unterschiedliche Kürzel vergeben.

D.h. zum Beispiel, dass eine fiktive Sprecherin Özlem im Subkorpus MuH3WT zu SPK1 anonymisiert wird, während eine Sprecherin gleichen Namens im Subkorpus MuH1WD ein anderes Kürzel erhält, z.B. SPK2).

Diese Kürzel werden sowohl zur Benennung der jeweiligen Sprecherebene als auch im Transkript verwendet. In den folgenden Beispielen steht *v* für die Transkriptionsebene und *n* für die Normalisierungsebene.

Beispiel ohne Anonymisierung:

MuH3WT:v	Özlem		komm	HER			
MuH3WT:n	Özlem	,	komm	her	!		
Özlem:v						oKE:	
Özlem:n						Okay	.

Beispiel mit Anonymisierung:

MuH3WT:v	SPK2		komm	HER			
MuH3WT:n	SPK2	,	komm	her	!		
SPK2:v						oKE:	
SPK2:n						Okay	.

Wird ein **Name im Genitiv** verwendet, wird ohne Apostrophierung ein **-s** angeschlossen:

nicht anonymisiert:	Özlems
v:	SPK13s
n:	SPK13s

3 Ortsangaben

Namen von Schulen und ggf. weiterer Ortsangaben werden wie folgt abgekürzt und anonymisiert.

Anfangsbuchstabe des Namens + xxx z.B. für „Alexander-Puschkin-Schule“: axxx(schule)

Auf der **Transkriptionsebene** wird die anonymisierte Form des Names kleingeschrieben [axxx(schule)], es sei denn, sie trägt einen Akzent. In diesem Fall wird die ganze Form großgeschrieben [AXXX(schule)]. Auf der **Normalisierungsebene** wird der Anfangsbuchstabe großgeschrieben: Axxx(Schule).

Beispiel:

```

| Audio: "Er geht auf die Alexander Puschkin"
|
| v: er geht auf die axxx(schule)                (ohne Akzent)
| v: AXXX(schule)                                (mit Akzent)
| n: Er geht auf die Axxx(Schule) .              (Normalisierungsebene)
|
| Audio: "Er geht auf das Königin-Luise-Gymnasium"
|
| v: er geht auf das kxxx(gymnasium)            (ohne Akzent)
| v: KXXX(gymnasium)                            (mit Akzent)
| n: Er geht auf das Kxxx(Gymnasium) .          (Normalisierungsebene)

```

U- und S-Bahnhöfe sowie Straßen- und Platznamen usw. werden folgendermaßen anonymisiert:

```

| v: c195xxx(straße)
| n: M3xxx(Bezirk)
| v: w10xxx(s-u-bahn)

```

Die Anfangsbuchstaben der zur Anonymisierung verwendeten Kürzel entsprechen nicht den Anfangsbuchstaben der zu anonymisierenden Namen, sondern wurden randomisiert.
Z.B.:

Aachener Straße	Transkriptionsebene:	f114xxx(straße)
Aachener Straße	Normalisierung:	F114xxx(Straße)

4 Unverständliches zu anonymisierendes Material

Bei Sprachmaterial, das anonymisiert werden muss, bei dem aber der genaue Wortlaut nicht verstanden und das dementsprechend bei der Anonymisierung auch nicht auf eine bestimmte Sigle zurückgeführt werden kann, wird ein Default-Kürzel gesetzt (v: x0xxx, n: X0xxx) + Kategorie der Angabe in Klammern:

```

| v: (- Akzent): x0xxx(kategorie)
| v: (+ Akzent): X0XXX(kategorie)
| n:                X0xxx(Kategorie)

```

Beispiele:

```

| v: ich wohne X0XXX(straße)
| n: Ich wohne X0xxx(Straße) .
|
| v: meine HANdnummer ist x0xxx(telefonnummer)
| n: Meine Handnummer ist X0xxx(Telefonnummer)

```

Ausnahmen sind Personennamen, die nicht verständlich sind. Hier entfällt die Angabe der Kategorie in Klammern:

```
-----  
| Akzentuiert:  
| v: wir sin gestern bei X0XXX gewesen  
| n: Wir sind gestern bei X0xxx gewesen .  
|  
| Unakzentuiert:  
| v: wir sin GESTern bei x0xxx gewesen  
| n: Wir sind gestern bei X0xxx gewesen .  
|-----
```

1.18 Geräusche/non-verbales Material

Geräusche oder non-verbales Material wird auf der non-verbale Ebene beschrieben. Es kann mehrere non-verbale Ebenen geben: eine globale Ebene, die z.B. Umweltgeräusche beschreibt, und für jede/n Sprecher/in maximal eine sprecherspezifische non-verbale Ebene. Non-verbales Material wird **nicht** auf der Transkriptionsebene erfasst. Ausnahmen werden im Folgenden beschrieben.

1.18.1 Geräusche/non-verbales Material mit Funktion innerhalb des Segments

Wenn non-verbales Material **innerhalb des Segments eine grammatische Funktion** erfüllt, bspw. als Argument eines Verbs, wird es zusätzlich auch auf der Transkriptionsebene repräsentiert.

```
nv:                               imitiert Autogeräusch  
v:   und denn (-) HÖR  ick nur so (-) wu::  
n:   Und denn       höre ich nur so      Wuh      .
```

→ Geräuschimitation(Wuh) ist hier Argument von *hören*.

```
nv:                               hustet  
v:   der ganze STAUB  ICK so   ICK so   Äche äche äche  
n:   Der ganze Staub . Ich so . Ich so , äche , äche , äche .
```

→ eingeleitete direkte Rede (auch ohne Verb)

1.18.2 Geräusche/non-verbales Material ohne Funktion innerhalb des Segments

Zu unterscheiden von den oben stehenden Fällen sind Fälle, in denen das non-verbale Ereignis diese Funktion nicht erfüllt und daher auch nicht auf v repräsentiert wird.

```
nv:                               imitiert Bremsgeräusche  
v:   und geNAU neben mir fängt er an zu BREMSen  
n:   Und genau neben mir fängt er an zu bremsen .
```

→ hier füllt das Geräusch keine Argumentstelle von *anfangen*

```
nv   : H [ausatmen]  
v:  
n:
```

→ hier wird nur auf der nonverbalen Ebene notiert

Wenn ein non-verbales Ereignis zu einer allgemeinen (sprecher-ungebundenen) Unterbrechung des Redeflusses führt, wird die Zeitangabe mit Komma abgetrennt hinter das Ereignis auf der non-verbale Ebene geschrieben und nicht als sprecherbezogene Pause annotiert.

nv: Straßenlärm, 17.4
nv: Kind brabbelt, 8.7

Im Unterschied dazu werden Pausen, die nicht in Verbindung mit einem non-verbale Ereignis stehen, in runden Klammern () auf der Transkriptionsebene (v) annotiert.

A: wann kommst du heut WIEder (3.3)
B: ich DENke so um (-) NEUN , halb ZEHN

2 Segmentierung und Akzentsetzung

Auf Transkriptionsebene findet keine Segmentierung in Phrasierungseinheiten statt, d.h. es findet sich auch keine Kennzeichnung von Tonhöhenverläufen wie *fallend*, *gleichbleibend* oder *steigend*. Die z.B. für die Akzentpositionierung relevanten prosodischen Einheiten werden lediglich auf der Normalisierungsebene markiert, und zwar mittels der orthografischen Interpunktionszeichen.

2.1 Akzentsetzung

Die Festlegung der **relevanten Einheit für die Akzentplatzierung** erfolgt auf der **Normalisierungsebene**. In der EXMARaLDA-Architektur haben die auf der Normalisierungsebene gesetzten Satzzeichen ein leeres Event auf der Transkriptionsebene als Entsprechung. Annotiert (auf Transkriptionsebene) wird der stärkste Akzent pro Segmentierungseinheit der Normalisierungsebene. Die Silbe, die den Akzent trägt, wird durch Versalien hervorgehoben (β bleibt dabei auch in einer akzentuierten Silbe β und wird nicht als *SS* verschriftet):

Ein Beispiel:

v: ich hab das buch geLEsen es hat mir NICHT gefallen
n: Ich habe das Buch gelesen . Es hat mir nicht gefallen .

Innerhalb eines (normalisierten) Segments kann auch mehr als einen Hauptakzent auftreten:

v: ich hab das buch geLEsen und den fisch geKAUft
n: Ich habe das Buch gelesen und den Fisch gekauft .

Sollte sich ein Unterschied in der Akzentstärke feststellen lassen, so wird nur der am stärksten betonte Akzent als Hauptakzent markiert. Ist das nicht der Fall, so werden beide Akzente markiert.

v: du WEIßT nicht , wie HEIß es ist
n: Du weißt nicht , wie heiß es ist .

2.2 Segmentierung

Im Folgenden werden die **Segmentierungszeichen** und die **Kriterien** gelistet, nach denen sie vergeben werden. Soweit es sinnvoll ist, wird dabei für jedes Segmentierungszeichen unterschieden, ob es nach einem satzwertigen (finites Verb) oder nicht-satzwertigen Segment (kein finites Verb) steht.

2.2.1 Punkt

2.2.1.1 Satzwertig

Der satzbeendende Punkt wird nach kanonischen **deklarativen Hauptsätzen (V2)** gesetzt.
Erläuterung: [v] = Transkriptionsebene, [n] = Normalisierungsebene.

MuH12MD [v]: ich bin ein GANZ lieber netter junge
MuH12MD [n]: Ich bin ein ganz lieber netter Junge .

Der satzbeendende Punkt steht nach nicht-kanonischen **deklarativen Hauptsätzen mit V1-Stellung**.

v: gibt =s einen MANN der total KRASS is
n: Gibt es einen Mann , der total krass ist .

Der satzbeendende Punkt steht nach nicht-kanonischen **deklarativen Hauptsätzen mit V3-Stellung**.

v: dann die sind zur U_bahn gerannt
n: Dann die sind zur U-Bahn gerannt .

2.2.1.2 Nicht-satzwertig

Der satzbeendende Punkt wird nach **Antworten auf Entscheidungsfragen (Ja-Nein-Fragen)** gesetzt, die nicht dem kanonischen Satzmodell (= Satz mit finitem Verb) entsprechen.

v: hast du das BUCH gelesen
n: Hast du das Buch gelesen ?
v: NEIN | NEE | vielLEICHT | wahrSCHEINlich | 'hm'hm | 'hmmh
n: Nein . | Nein . | Vielleicht . | Wahrscheinlich . | Mh-Mh . | Hm-hm .

Diese Entscheidungsfragen müssen nicht dem Satztyp *Interrogativsatz (V1-Stellung)* entsprechen, sondern können auch durch Intonation als solche markiert werden (s. Abschnitt 2.2.4 zu Fragen).
Hier ein Beispiel aus dem Korpus:

MuH1WD [v]: JA
MuH1WD [n]: Ja .
Jxxx [v] : wir laufen grad (-) zu ALdi
Jxxx [n] : Wir laufen gerade zu Aldi ?

Zu unterscheiden sind Fälle, in denen es sich um eine **Interjektion** handelt. Diese zeichnen sich durch eine der Antwort gegenüber markierte Intonation aus und werden durch ein Ausrufezeichen markiert.

v: hast du das BUCH gelesen
n: Hast du das Buch gelesen ?
v: NEE
n: Nein .
v: ich hab das BUCH gelesen
n: Ich habe das Buch gelesen .
v: NEE
n: Nein !

Daneben gibt es noch weitere **elliptische Antworten**, die ebenfalls mit einem Punkt segmentiert werden. Zu beachten sind die folgenden Beispiele:

v: LÄSST du das bitte beim nächsten mal
n: Lässt du das bitte beim nächsten Mal ?
v: Okay
n: Okay .

v: wo WARST du
n: Wo warst du ?
v: im KIno | SCHWIMmen
n: Im Kino . | Schwimmen .

v: was hast du geLEsen
n: Was hast du gelesen ?
v: nich VIEL | irgendWAS
n: Nicht viel . | Irgendwas .

v: wie soll ich das transkriBIERen
n: Wie soll ich das transkribieren ?
v: genau SO
n: Genau so .

Ein Beispiel aus dem Korpus:

MuH1WD [v]: bist du für GRIEchenland oder SCHWEden (--)
MuH1WD [n]: Bist du für Griechenland oder Schweden ?
Jxxx [v]: für GRIEchenland
Jxxx [n]: Für Griechenland .
Spk2 [v]: ich AUCH
Spk2 [n]: Ich auch .

Der satzbeendende Punkt wird nach Rezeptionssignalen gesetzt.

Einsilbige Signale:

v: ah
n: Ah .

v: hm
n: Hm .

Zweiselbige Signale:

v: 'hmhm
n: Hm-hm .

Beispiele:

v: ins muSEum
n: Ins Museum .

v: ah
n: Ah .

v: weißt du WAS ich kauf mir n AUto
n: Weißt du was ? Ich kaufe mir ein Auto .

v: hm
n: Hm .

2.2.2 Ausrufezeichen

2.2.2.1 Satzwertig

Das Ausrufezeichen wird nach **Imperativsätzen mit V1-Stellung** gesetzt:

v: stell dich nicht so AN
n: Stell dich nicht so an !

v: HALT
n: Halt !

MuHl2MD [v]: mach MA
MuHl2MD [n]: Mach mal !

Das Ausrufezeichen wird nach **exklamativen Äußerungen** gesetzt.

v: was DER schon wieder alles erzählt hat
n: Was der schon wieder alles erzählt hat !

v: wie der schon wieder AUSSah
n: Wie der schon wieder aussah !

v: dass das jetzt schon wieder passIERen musste
n: Dass das jetzt schon wieder passieren musste !

Bedingung für diese exklamative Konstruktion: Nicht in den Matrix-Satz eingebettet

Das Ausrufezeichen steht **nicht** nach Deklarativsätzen:

A: ... weil ich mache auf jeden Fall bald wieder Praktikum .
B: Kannst vergessen . (elliptisch: Das kannst du vergessen .)
vgl.: Vergiss es ! (Imperativ)

Wenn die sprachliche Handlung als Imperativ zu interpretieren ist (z.B. durch das Modalverb *sollen*), wird ein V2-Satz mit indikativischer Verbform mit einem Punkt segmentiert (strukturelles Kriterium).

Du sollst aber 30 Stunden aufnehmen .

Nur nach „echten“ Imperativformen steht ein Ausrufezeichen:

Nimm 30 Stunden auf !

2.2.2.2 Nicht-satzwertig

Das Ausrufezeichen wird nach Interjektionen gesetzt.

v: ich hab papas auto gegen die WAND gesetzt
n: Ich habe Papas Auto gegen die Wand gesetzt .

v: ach du MEIne güte | ach du SCHEIße | SCHEIße | na TOLL | verDAMMT
n: Ach du meine Güte ! Ach du Scheiße ! Scheiße ! Na toll ! Verdammt !

v: ich hab gestern ACHT stunden gearbeitet
n: Ich habe gestern acht Stunden gearbeitet .

v: aHA | AU | PFUI
n: Aha ! Au ! Pfui !

Weitere Beispiele:

v: EY | ACH | OH | ach SO | na JA
n: Ey ! Ach ! Oh ! Ach so ! Na ja !

Auch Fälle wie der folgende werden als Interjektionen interpretiert und mit Ausrufezeichen segmentiert:

nv: *lacht*
v: haha
n: Haha !

Wir behandeln *Mann*, wenn es nicht als referentielles Nomen verwendet wird, immer als Interjektion. In diesen Kontexten wird es immer abgetrennt.

v: da müssen wir UNbedingt hin mann
n: Da müssen wir unbedingt hin . Mann !

Das Ausrufezeichen wird nach **Anreden/Vokativen** gesetzt, wenn diese isoliert vorkommen.

v: PEter im Unterschied zu: PEter haben wir noch Äpfel
n: Peter ! Peter , haben wir noch Äpfel ?

Das Ausrufezeichen wird nach **Begrüßungen/Verabschiedungen** gesetzt.

v: HALlo | HI | TSCHÜß | NA
n: Hallo ! Hi ! Tschüß ! Na !

Bei einer **Kombination aus Begrüßung und Anrede** wird die Anrede durch Komma von der Begrüßung getrennt und das gesamte Segment mit Ausrufezeichen versehen. Das Ausrufezeichen richtet sich dabei nach der Begrüßung. Anreden werden, wenn sie nicht isoliert vorkommen, immer mit Komma angeschlossen.

v: NA herr mustermann
n: Na , Herr Mustermann !

v: HALlo exxx
n: Hallo , Exxx !

Das Ausrufezeichen steht nach **Befehlen, die keine Imperativform enthalten**, aber dennoch die Funktion eines Imperativs ausüben.

v: STOP
n: Stop !

v: nochMAL
n: Nochmal !

Dies gilt auch für **indirekte Anordnungen/Befehle** ohne imperativische Verbform.

v: auf DEUTSCH reden | langsam LAUFen | nicht zu LAUT
n: Auf Deutsch reden ! | Langsam laufen ! | Nicht zu laut !

Das Ausrufezeichen steht ebenfalls nach **exklamativen Ausdrücken**, die im Unterschied zu ihren satzwertigen Pendanten kein finites Verb haben müssen.

v: wie BLÖD | wie SCHRECKlich | wie unerhört AUFdringlich |
n: Wie blöd ! | Wie schrecklich ! | Wie unerhört aufdringlich ! |
v: was für ein DUSsel
n: Was für ein Dussel !

Das Ausrufezeichen steht nach **Höflichkeitspartikeln** (mehnteilige Höflichkeitsbekundungen werden auf zwei bzw. mehrere Events aufgeteilt).

v: TSCHULdigung
n: Tschuldigung !

v: DANke
n: Danke !

v: DANke sehr
n: Danke sehr !

Hier ein Beispiel aus dem Korpus:

MuH1WD [v]: DANke | BITte schön | BITte
MuH1WD [n]: Danke ! | Bitte schön ! | Bitte !

Im Unterschied zur Höflichkeitspartikel kann *danke* auch innerhalb eines kanonischen Satzes (möglicherweise mit elliptischem Vorfeld) als finites Verb verwendet werden. Dieses Segment wird am Ende mit . und nicht mit ! annotiert. Das gilt auch für ähnliche Verwendungen von *bitte*.

MuP1MK [v]: dank dir | ich BITte dich
MuP1MK [n]: Danke dir . | Ich bitte dich .

Das Ausrufezeichen steht nicht nach "normalen" Antworten, auch wenn diese emphatisch geäußert werden:

Ja. Nein. Okay.

2.2.3 Allgemein: Differenzierung

Dieses Kapitel gibt eine Zusammenfassung der Unterschiede zwischen Interjektionen, exklamativen Aussagen, Imperativen und deklarativen Aussagen. Je nach Äußerungstyp werden sie unterschiedlich abgetrennt.

2.2.3.1 Interjektionen

Isoliert stehend oder als Mehrwortausdruck

Abtrennung durch **Ausrufezeichen**:

Ach ! Oh ! Ey ! Mann ! Oh mein Gott ! Ach du Scheiße !

2.2.3.2 Exklamative Ausdrücke

Unter exklamativen Ausdrücken verstehen wir emphatische Ausdrücke, die im Gegensatz zu den Interjektionen einen propositionalen Gehalt haben (genau wie deklarative Äußerungen).

Abtrennung durch **Ausrufezeichen**:

Wie blöd! / Wie blöd ist das denn!

hat einen propositionalen Gehalt
→ exklamativer Ausdruck

Hey! hat **keinen** propositionalen Gehalt → Interjektion

Exklamative Ausdrücke beinhalten oft Interjektionen:

Mein Gott, wie heiß!

2.2.3.3 Imperative Ausdrücke

Abtrennung durch **Ausrufezeichen**:

Geh weg !

Guck mal , wer da kommt !

Äh warte mal !

Sag mal deinen Namen richtig (unverständlich) !

Indirekte Anordnungen/Befehle (ohne imperative Verbform) werden ebenfalls als Imperativ behandelt und durch Ausrufezeichen abgetrennt:

(du sollst) Auf Deutsch reden !

(du sollst) Langsam laufen !

(du sollst) Nicht zu laut (sein) !

2.2.3.4 Deklarative Ausdrücke

Abtrennung durch **Punkt**:

Das ist total blöd.

2.2.3.5 Abgrenzung deklarative vs. exklamative Ausdrücke

Sowohl deklarative als auch exklamative Äußerungen haben einen propositionalen Gehalt. Der Unterschied liegt in der Funktion: Erstere haben die Aufgabe, Informationen zu liefern, während letztere eher expressiv sind, d.h. zum Ausdruck von Gefühlen genutzt werden. Exklamative Aussagen haben einen stark emotionalen Gehalt und setzen im Allgemeinen voraus, dass die Proposition beiden, Sprecher/in und Hörer/in, bekannt ist.

Typische exklamative Konstruktionen:

Wie ADJ (es hier ist)/ Wie ADJ (ist das denn) !

Wie heiß es hier ist!

Wie bitter !

Wie geil ist das denn !

Dass X NP (nicht) VERBKOMPLEX !

Dass der sich nicht schämt !

Dass Gabi sich nicht blöd vorkommt !

Dass ich das erdulden muss !

Was für ein/e NP !

Was für ein Schwein !

Aber nicht:

Ich komme mir komisch vor . **Wie so ein Opfer .**

→ Modifikation, neue Information: ich komme mir vor wie so ein Opfer.

Vgl. aber:

Was für ein Opfer !

→ keine neue Information: Du stimmst mir doch zu / wir beide wissen, was für ein Opfer der/die ist!

Nicht alle durch Konjunktion eingeleiteten selbständigen Nebensätze sind exklamativ:

Dass der sich nicht schämt ! → exklamativ

Weil der sich nicht schämt . → deklarativ

2.2.4 Fragezeichen

2.2.4.1 Satzwertig

Das Fragezeichen wird nach **Entscheidungsfragen mit V1-Stellung** gesetzt:

v: hast du das geSEHEN
n: Hast du das gesehen ?

Das Fragezeichen wird nach **W-Fragen** gesetzt:

v: was hast du geSEHEN
n: Was hast du gesehen ?

MuH11MD [v]: und WER spielt
MuH11MD [n]: Und wer spielt?

Das Fragezeichen wird darüber hinaus – unabhängig vom Verbstellungstyp – nach allen **durch Intonation als Frage** gekennzeichneten Äußerungen gesetzt (das Zeichen "/" ist nicht Teil der Transkriptionskonventionen, sondern markiert in den Beispielen einen steigenden Tonhöhenverlauf).

v: hans war schon wieder beTRUNken/
n: Hans war schon wieder betrunken ?
v: das wollte ich schon IMmer mal lesen wieSO/
n: Das wollte ich schon immer mal lesen . Wieso ?

Das Fragezeichen steht nach **Echo-Fragen**.

v: wann KOCHST du
n: Wann kochst du ?
v: wann ich KOche/
n: Wann ich koche ?
v: hat WER gesagt/
n: Hat wer gesagt ?

Liste weiterer subordinierender Konjunktionen, die VL-Stellung auslösen (können):

wenn, ob, wie, weil, damit, bis, so, da, zumal, nachdem, falls, bevor...

Wichtig: vor Konjunktionen mit V2-Stellung setzen wir kein Komma:

v: ich bin Müde weil ich bin gestern spät ins BETT

n: Ich bin müde weil ich bin gestern spät ins Bett .

Vor **uneingeleiteten Nebensätzen mit V2-Stellung** wird ein Komma gesetzt.

v: ich WEIß du KENNST mich

n: Ich weiß , du kennst mich .

Das Komma trennt **Einschübe** ab. Diese können bspw. zur genaueren Spezifizierung des vorher Gesagten genutzt werden oder den Sprechakt selber modifizieren .

MuHlWD [v]: dann (-) die ganze REIse so also alles komPLETT hat

MuHlWD [n]: Dann die ganze Reise so , also alles komplett , hat

MuHlWD [v]: ZWEIhundertsechzig EUro gekostet

MuHlWD [n]: 260 Euro gekostet .

MuHlWD [v]: was ANderes kann ich mir ehrlich gesagt AUCH nicht vorstellen

MuHlWD [n]: Was anderes kann ich mir , ehrlich gesagt , auch nicht vorstellen .

Das Komma trennt **Linksversetzungen** ab. Linksversetzungen sind prosodisch und syntaktisch in den Folgesatz integriert.

Syntax:

- adjazentes D-Pronomen (*der, die, das*) oder D-Adverb (*dadurch, davon*), das das initiale Satzglied wieder aufgreift.

Prosodie:

- gleichbleibender Intonationsverlauf zwischen initialem Satzglied und D-Pronomen/ D-Adverb.

v: die STERne die beGEHRT man nicht

n: die Sterne , die begehrt man nicht .

Ein Korpusbeleg:

MuHl1MD [v]: JA FREItag zum beispiel (--) da warst du LIEB

MuHl1MD [n]: Ja ! Freitag zum Beispiel , da warst du lieb .

Das Komma trennt das **Hanging Topic** (auch "Freies Thema") ab. Ein Freies Thema wird durch syntaktische und prosodische Merkmale markiert.

Syntax:

- wiederaufgreifendes Element (Pronomen *er, sie, es*; D-Pronomen *der, die, das*; DP oder PP)
- resumptives Element nicht notwendigerweise adjazent zum initialen Satzglied

Prosodie:

- vorangestelltes Satzglied prosodisch nicht integriert
- Element häufig durch eine Pause vom Rest der Äußerung abgetrennt

v: die STERne ich habe sie noch NIE leiden können
n: Die Sterne , ich habe sie noch nie leiden können .

v: die FRAU da ich glaub ich hab die schon mal geSEHN
n: Die Frau da , ich glaube , ich habe die schon mal gesehen .

Bei steigender Intonation wird das Freie Thema durch ein Fragezeichen abgetrennt, ansonsten mit einem Komma ("," ist nicht Teil der Transkriptionskonventionen und dient hier nur der Darstellung des steigenden Tonverlaufs).

v: SAbriNa/ ich konnt die tussi noch NIE leiden
n: Sabrina ? Ich konnte die Tussi noch nie leiden .

Beim folgenden Beispiel handelt es sich nicht um ein Freies Thema im eigentlichen Sinne, sondern um die elliptische Antwort auf eine Frage. Deshalb trennen wir hier mit einem Punkt ab.

v: was macht dir am meisten PROBLEme
n: Was macht dir am meisten Probleme ?
v: KAffee wenn ich DEN trink krieg ich leicht KOPFSchmerzen
n: Kaffee . Wenn ich den trinke , kriege ich leicht Kopfschmerzen .

Hanging Topic in Kombination mit *gibs*

Im folgenden Beispiel sind zwei Analysen möglich:

a) als präpositionslose PP

(PP für R) (VVFİN-HD gib=) (NN-SB s) noch diese DINGS da

b) als freies Thema mit ausgelassenem anaphorischen *da* (Auer 2013:33)

(NN R) , da gib= s noch diese DINGS da

Wir transkribieren entsprechend der zweiten Analyse und setzen ein Komma nach dem Hanging Topic:

v: anfangsbuchstabe ER hast du (--) **ER** gib= s noch diese DINGS da (-)
n: Anfangsbuchstabe R hast du . **R** , gibt es noch diese Dings da .
v: OHRringe ER gib =s
n: Ohrringe R gibt es .

Extrapositionen (Versetzung eines Satzglieds in die rechte Peripherie) werden durch Komma abgetrennt, jedoch nur, wenn die rechte Satzklammer (RSK) besetzt ist.

v: ich bin/LSK nach HAUse gefahren/RSK gestern
n: Ich bin/LSK nach Hause gefahren/RSK , gestern .
v: ich hab voll geLACHT so
n: Ich habe voll gelacht , so .

leere RSK:

v: ich fuhr/LSK nach HAUse gestern
n: Ich fuhr/LSK nach Hause gestern .

Wenn die rechte Satzklammer durch ein Verb oder eine Verbpartikel besetzt ist und darauf noch weiteres lexikalisches Material innerhalb des Segments folgt, wird mit Komma segmentiert.

Die hat sich übelst verändert, seitdem die mit dem zusammen ist .

Analog zur Linksversetzung trennt das Komma die **Rechtsversetzung** (pronominale Kopie + Antezedens) ab, auch wenn die rechte Satzklammer nicht besetzt ist.

v: **er** kommt von HINTen **der SCHWUCHtel**
n: Er kommt von hinten , der Schwuchtel .

Bei **mehrfach besetzten Vorfeldern** wie in den folgenden Beispielen wird das erste Element NICHT durch ein Komma abgetrennt.

Heute ich habe Musikarbeit eine Musik jetzt eine Eins geschrieben , eine glatte Eins .

Bei mehrfach besetzten Vorfeldern mit initialem *also* wird dieses nicht durch ein Komma abgetrennt.

Vergleiche:

Denn ich habe ...
Aber ich habe ...
Dann ich habe ...
Also ich habe ...

Begründung: Wir wollen jetzt noch keine theoretische Entscheidung treffen, wie die unterschiedlichen Strukturen behandelt werden sollen, sondern wollen in der Lage sein, alle mehrfachen Vorfelder zu finden.

Das Komma trennt **Reihungen** ab.

v: die vielen vielen BLUmen
n: Die vielen , vielen Blumen .

Das Komma trennt **Wiederholungen**.

v: ich habe habe gestern DOCH noch dran gedacht
n: ich habe , habe gestern doch noch dran gedacht .

Das Komma trennt **Check-Questions/Rückversicherungssignale** innerhalb einer Äußerung ab.

v: wir haben SECHS stunden JA (-) gewARTet
n: Wir haben sechs Stunden , ja , gewartet .

v: der MANN JA der hat einen an der WAFFel
n: Der Mann , ja , der hat einen an der Waffel .

v: den MANN JA ich kann den nicht LEIden
n: Den Mann , ja , ich kann den nicht leiden .

Das Komma trennt **Anreden/Adressierungen innerhalb eines Segments** ab. Dies gilt auch für Anredeformen, die bspw. aus dem Türkischen stammen und auch in Kiezdeutsch gebräuchlich sind

(bspw. *Oğlum* = mein Sohn). Wir behandeln *Oğlum*, *Dicker*, *Süße*, *Alter* etc., die als Anrede verwendet werden, wie Eigennamen und trennen sie durch Komma ab.

v: HXXX wir fahren jetzt LOS
n: Hxxx , wir fahren jetzt los .

v: bei mir kannst du nicht LANden Süßer
n: Bei mir kannst du nicht landen , Süßer .

v: ich bin arCAden oglum
n: Ich bin Arcaden , Oglum .

v: ALter lass das SEIN
n: Alter , lass das sein !

Steht *Alter* isoliert, behandeln wir es wie einen Eigennamen/Vokativ:

Sxxx ! Alter !

2.2.6 Doppelpunkt und Anführungszeichen

Zusätzlich zu den hier angeführten Segmentierungsregeln vgl. Abschnitt 5.4 Direkte Rede/Zitation/Bericht.

2.2.6.1 Satzwertig

Anführungszeichen markieren die **direkte Rede**.

v: mxxx SAGT über sich ich habe IMmer glück
n: Mxxx sagt über sich : " Ich habe immer Glück . " .

MuH3WT [v]: IXXX sagte mir so KU= ma du redest wie ne DEUtsche
MuH3WT [n]: Ixxx sagte mir so : " Guck mal ! Du redest wie eine Deutsche . " .

2.2.6.2 Nicht-satzwertig

Anführungszeichen werden auch gesetzt, wenn die direkte Rede **ohne Verben des Sagens** eingeleitet wird.

v: ich so das musst du mir nächstes mal SAgen er so KLArO
n: Ich so : " Das musst du mir nächstes mal sagen . " Er so : " Klaro .

v: das MACH ich
n: Das mache ich . " .

2.2.7 Unspezifisches Grenzzeichen (#)

Die Raute markiert **Abbrüche**:

v: die MÜHe hättest
n: Die Mühe hättest #

v: jetzt rede ICH erst mal weiter
n: Jetzt rede ich erst mal weiter .

Die Raute steht nach **gefüllten Pausen**, wenn diese nicht innerhalb einer anderen Äußerung stehen.

v: äh
n: Äh #

v: ich äh weiß jetzt AUCH nich weiter
n: Ich äh weiß jetzt auch nicht weiter .

MuH2WT [v]: hä
MuH2WT [n]: Hä #

MuH1WD [v]: hm
MuH1WD [n]: Hm #

S2xxx [v]: yxxx is entTÄUSCHT
S2xxx [n]: Yxxx ist enttäuscht .

2.2.8 Allgemein

Wenn unklar ist, ob lexikalisches Material in zwei Segmente (satzwertig oder nicht-satzwertig) gesplittet werden soll oder eine Einheit bildet, beide Alternativen also möglich sind, verfolgen wir das Prinzip der kleinsten möglichen Einheit. Das bedeutet, dass wir kleinere Segmente größeren vorziehen, wenn die Segmentierung auf Diskurs- oder Sprechaktebene möglich/plausibel ist. Das betrifft häufig die Segmentierung nicht-satzwertiger Äußerungen, kann aber auch bei satzwertigen Äußerungen zur Anwendung kommen.

Beispiele (nicht-satzwertige Äußerungen)

v: red auf DEUTSCH oKE:
n: Rede auf Deutsch ! Okay ?
→ Segmentieren plausibel! Zwei unterschiedliche Sprechakte: Befehl + Frage

v: das ist GROß WArte
n: Das ist groß . Warte !
→ Segmentieren plausibel! Zwei unterschiedliche Sprechakte: Aussage + Befehl

Beispiele (satzwertige Äußerungen)

v: ich WEIß nicht was möchtest du ESsen
n: Ich weiß nicht . Was möchtest du essen ?
→ Segmentieren plausibel! Zwei unterschiedliche Sprachakte: Aussage + Frage

v: ich WEIß nicht was du ESsen möchtest
n: Ich weiß nicht , was du essen möchtest ?
→ keine Segmentierung! Matrixsatz + Relativsatz

3 Partikeln und Signale

3.1 Diskursmarker/Diskurspartikeln

Bei Diskursmarkern oder Diskurspartikeln (im Folgenden *Diskursmarker*) handelt es sich um unflektierbare lexikale Elemente, die fast ausschließlich in gesprochener Sprache vorkommen. Sie dienen der Organisation und Strukturierung von Texten und Diskursen. In der Sprachwissenschaft herrscht eine lange Debatte über den grammatikalischen Status und die adäquaten Beschreibungskriterien für Diskursmarker. Für das Deutsche wird davon ausgegangen, dass sich Diskursmarker durch folgende Merkmale auszeichnen:

Funktion:

- Gliederung des Diskurses
- Aneinanderreihung von Äußerungseinheiten
- Organisation der Interaktion zwischen Sprecher und Hörer, Aufmerksamkeitssteuerung
- Turn-taking

Grammatische Merkmale:

- periphere syntaktische Position: sie stehen vor oder nach unabhängigen Syntagmen (im Unterschied zu Modalpartikeln)
- können eine eigene prosodische Einheit bilden
- homophon mit einigen Adverbien (*jedenfalls*), Fragepronomen (*wobei*) und anderen Wortarten, aber mit einer davon abweichenden Funktion

Im Rahmen unserer Konventionen werden Diskursmarker auf der Normalisierungsebene **nicht** durch Interpunktion von dem folgenden Syntagma (bei linksperipherer Stellung) oder dem vorangehenden Syntagma (bei rechtsperipherer Stellung) abgetrennt. Unsere Richtlinien konzentrieren sich hier auf linksperiphere Diskursmarker. Teilweise wird davon ausgegangen, dass es sich bspw. auch bei einem fragenden *oder*, das einem Satz nachgestellt ist, um einen Diskursmarker handelt. Solche Fälle haben wir als *Fragepartikeln* unter der Segmentierung mit *Fragezeichen* mit eingeschlossen.

Beispiele:

Adverb

v: ICH bin jedenfalls NICHT betrunken auto gefahren
n: Ich bin jedenfalls nicht betrunken Auto gefahren .

Diskursmarker

v: JEdenfalls ICH bin NICHT betrunken auto gefahren
n: Jedenfalls ich bin nicht betrunken Auto gefahren .

Als Adverb ist *jedenfalls* in den Satz integriert und drückt modale Bedeutung aus. Als Diskursmarker ist es weniger stark in den Satz integriert und markiert die Rückkehr zu einem anderen Erzählstrang.

Fragepronomen

v: wobei soll ich dir helfen
n: Wobei soll ich dir helfen ?

Diskursmarker

v: das schaffst du schon alleIne wobei ich soll dir ja HELfen
n: Das schaffst du schon alleine . Wobei ich soll dir ja helfen .

Als Fragepronomen fragt *wobei* in dem Beispielsatz nach dem Präpositionalobjekt von *helfen*. Als Diskursmarker markiert es die Reparatur/Korrektur der vorangehenden Äußerung/Behauptung.

Weitere Beispiele für Diskursmarker:

Vor Syntagma mit V1-Stellung

v: ja bist du denn WAHNSinnig
n: Ja bist du denn wahnsinnig ?

Vor Syntagma mit V2-Stellung

v: ja das ist doch WAHNSinn
n: Ja das ist doch Wahnsinn .

v: ALso ICH wollte gern kekse
v: ALso (-) ICH wollte gern kekse
v: also ICH wollte gern kekse
v: ALso ich wollte gern kekse
v: also ich wollte gern KEKse
n: Also ich wollte gern Kekse .

Unabhängig von den Akzentverhältnissen und möglichen platzierten gefüllten/ungefüllten Pausen zwischen dem Diskursmarker und dem folgenden Syntagma wird der Diskursmarker **nicht** abgetrennt!

MuH3WT [v]: ALSO wir warn grad in der SHIshabar (-) und wir kommen da nicht
MuH3WT [n]: Also wir waren gerade in der Shishabar und wir kommen da nicht

MuH3WT [v]: rein weil wir nicht ACHTzehn sind
MuH3WT [n]: rein , weil wir nicht 18 sind .

Diskursmarker können auch vor nicht-satzwertigen Äußerungen vorkommen. Auch hier werden sie nicht segmentiert. Ein Beispiel aus dem KiezDeutsch-Korpus:

MuP6MD [v]: FÜNfzich also BIS zu fünfzich euro
MuP6MD [n]: 50 . Also bis zu 50 Euro .
Spk1 [v]: JEde
Spk1 [n]: Jede ?

Unter die Konvention von Diskursmarkern fallen nach diesen Richtlinien auch die folgenden Beispiele, unabhängig vom Betonungsmuster und möglichen gesetzten Pausen:

v: dann ich wollte schon immer mal ins KINO
v: DANN ich wollte schon immer mal ins KINO
v: DANN (-) ICH wollte schon immer mal ins kino
n: Dann ich wollte schon immer mal ins Kino .

Temporale Adverbiale im Vorvorfeld eines V2-Satzes (Adv-SVO) sind ein bekanntes und vieldiskutiertes Phänomen im Kiezdeutschen. Dieser Fall von einem nicht-kanonischen Hauptsatz wird auch in den Richtlinien zum Punkt (Abschnitt 2.2.1) behandelt. Auch hier setzen wir kein Komma.

v: danach wir fahren RATHaus .
v: Danach wir fahren Rathaus .

3.2 Fragepartikeln

Vgl. auch die Segmentierungsrichtlinien zu Fragezeichen und Komma (Abschnitt 2.2.4 und 2.2.5) !

Bei Fragepartikeln handelt es sich um Partikeln, die oft der Rückversicherung des Sprechers dienen (*ne, gell, oder, ja* etc.) und die häufig am Anfang oder am Ende einer Äußerung auftreten. Sie werden durch ein Fragezeichen markiert und, wenn sie auf eine andere Äußerung folgen, durch einen Punkt von dieser abgetrennt.

v: wir fahren jetzt nach HAUse NE
n: Wir fahren jetzt nach Hause . Ne ?

v: NE wir fahren jetzt nach HAUse
n: Ne ? Wir fahren jetzt nach Hause .

v: also eigentlich ist das hier ja ne SÜper sache NE
n: Also eigentlich ist das hier ja eine super Sache . Ne ?

v: das essen is SCHEIße
n: Das Essen ist scheiße .

v: JA NE
n: Ja . Ne ?

Es gibt auch komplexe Fragepartikeln, die auf Syntagmen mit bestimmten Verben (meistens kognitiven Verben) zurückgehen; in ihrer Verwendung als Fragepartikel haben sie eine gebleichte Semantik gegenüber dem lexikalischen Vollverb.

v: es fängt bald an zu REGnen WEIßT du
n: Es fängt bald an zu regnen . Weißt du ?

Einen besonderen Fall bildet das Vorkommen dieser Partikeln innerhalb von Äußerungen und in Linksversetzungs- und Hanging Topic-Konstruktionen. In diesen Fällen werden Fragepartikeln mit Komma vom Rest der Äußerung abgetrennt.

Innerhalb einer Äußerung

v: du WEIßT doch dass ich spinAT JA nicht MAG
n: Du weißt doch , dass ich Spinat , ja , nicht mag .

Linksversetzung

v: der typ hier DRÜben NE | JA der malt BILder
n: Der Typ hier drüben, ne | ja , der malt Bilder .

Hanging Topic

v: der typ hier DRÜben NE ich MAG den nich
n: Der Typ hier drüben, ne, ich mag den nicht .

3.3 Antwortpartikeln

Antwortpartikeln (*ja, Hm-hm, Mh-mh,...*) können auf Entscheidungsfragen folgen, um diese zu bejahen oder zu verneinen. Sie können aber auch nach impliziten Fragen gesetzt werden, also solchen, die strukturell oder intonatorisch keinem Frage-Typ entsprechen, sondern sich aus dem Diskurs ergeben. In diesen Fällen wird angenommen, dass jede Assertion eine Antwort auf eine kontextuell relevante (u.U. aber implizit bleibende) Frage ist. Antwortpartikeln werden mit **Punkt** segmentiert (und nicht mit einem Ausrufezeichen). Es gibt Überschneidungen zwischen Antwortpartikeln und Rezeptionssignalen. In KiDKo werden beide mit Punkt segmentiert.

explizit

v: hast du HXXX getroffen
n: Hast du Hxxx getroffen ?
v: JA | NEIN | 'hmHM | 'HM'hm
n: Ja . | Nein . | Hm-hm . | Mh-Mh .

implizit

v: mein kumpel trinkt sein BIER nie aus
n: Mein Kumpel trinkt sein Bier nie aus .
v: JA | 'hmHM | 'HM'hm das STIMMT nicht
n: Ja . | Hm-hm . | Mh-Mh . Das stimmt nicht .

3.4 Modalpartikeln/Abtönungspartikeln

Modalpartikeln (auch Abtönungspartikeln) fügen der Satzbedeutung bestimmte Verwendungsbedingungen hinzu, die sich in der Regel auf Einstellungen der Kommunikationsteilnehmer zu dem durch den Satz ausgedrückten Inhalt beziehen. So fügt *ja* als Modalpartikel der Bedeutung von Aussagesätzen die Verwendungsbedingung hinzu, dass der Satzinhalt für den/die Hörer/in nicht zweifelhaft sein darf: *Hans ist unglücklich* vs. *Hans ist ja unglücklich*.

Zu Modalpartikeln zählen im Deutschen u.a. *ja*, *doch*, *denn*, *aber*, *nur*. Probleme der Klassifikation ergeben sich dadurch, dass viele der Wortformen, die als Modalpartikeln verwendet werden können, auch andere Funktionen tragen können, bspw. *ja* als Fragepartikel oder *nur* als fokusassoziierte Partikel.

Syntaktische Merkmale von Modalpartikeln im Deutschen (jedoch nicht allgemeingültig):

- Modalpartikeln können weder allein noch mit ihrer Bezugskonstituente im Vorfeld stehen: **Ja unglücklich ist Hans*
- Modalpartikeln können nicht betont werden: **Hans ist JA unglücklich*
- Modalpartikeln sind an bestimmte Satztypen gebunden: **Wer ist ja unglücklich*, ABER: *Wer ist denn glücklich*

Weitere Merkmale sind:

- Modalpartikeln sind unflektierbar
- Modalpartikeln ändern nicht den Wahrheitswert einer Äußerung
- Modalpartikeln können nicht negiert werden
-

Modalpartikeln werden auf Normalisierungsebene **nicht** abgetrennt, auch dann nicht, wenn sie am Äußerungsrand stehen!

v: das is ja ne saueREI
n: Das ist ja eine Sauerei .

v: das hab ich halt geLEsen
n: Das habe ich halt gelesen .

v: das hab ich geLEsen halt
n: Das habe ich gelesen halt .

Ein Korpusbeispiel:

MuP6MD [v]: ACH ja STIMMT du bist ja ACH du bist ja MAThe EY WA
MuP6MD [n]: Ach ja ! Stimmt . Du bist ja # Ach ! Du bist ja Mathe . Ey ! Wa ?

3.5 Gefüllte Pausen

Pausen zählen den suprasegmentalen Merkmalen. Wir unterscheiden zwischen ungefüllte Pausen (auch stille Pausen, s. Abschnitt 1.10) und gefüllten Pausen (*äh, ähm, öh...*, s. Tabelle 1 unten).

In der Sprachproduktion sind gefüllte Pausen häufig ein Zeichen für ablaufende Planungsprozesse. Gefüllte Pausen können in jeder Position vorkommen, oft auch in Kombination mit Reparaturen. Sie sind semantisch leer und werden nicht abgetrennt, wenn sie innerhalb eines Segments oder an dessen Rand vorkommen.

v: äh ich war gestern im KIno
n: Äh ich war gestern im Kino .

v: Ich hab gestern äh BLUmenkohl gegessen
n: Ich habe gestern äh Blumenkohl gegessen .

Ein Korpusbeleg (die Raute markiert hier einen Äußerungsabbruch):

Spk2 [v]: und dann ähm (--) äh
Spk2 [n]: Und dann ähm äh #

Isoliert stehende gefüllte Pausen werden durch # abgetrennt! Isoliert sind Pausen, wenn ihnen keine Äußerung direkt vorangeht oder folgt. Wenn sie einer Äußerung vorangehen oder inmitten dieser auftreten, werden sie als integriert betrachtet.

v: erzÄHL doch ma KOMMT da noch was
n: Erzähl doch mal ! Kommt da noch was ?
v: ÄH
n: Äh #

Ähm kennst du noch dieses Schlagerdingsdabumsda ?
und nicht:

Ähm # Kennst du noch diese Schlagerdingsdabumsda ?

3.6 Rezeptionssignale

Rezeptionssignale zeigen die Aufmerksamkeit der zuhörenden Person an. Sie stehen allein oder in Kombination mit anderen Rezeptionssignalen. Wir behandeln hier nur Rezeptionssignale, die keine eigene lexikalische Bedeutung haben (bspw. *hmm, mh-mh, aha, oh, ach*).

Die Verschriftungsformen sind in der folgenden Liste zusammengefasst (s. auch die gesonderte Übersicht "Normalisierung"):

Transkription	Normalisierung	Erläuterung
äh	äh	Filler
ähm	ähm	Filler
hm	hm	Rezeptionssignal
nee	nein	Verneinung
'hm'hm	'hmhm	Bejahung
mh-mh	hm-hm	Rezeptionssignal
'ä'ä	äh-äh	Verneinung

Tab. 2: Normalisierung von gefüllten Pausen (Fillern), Rezeptionssignalen und Antwortpartikeln

Rezeptionssignale werden mit einem Punkt (s. auch Abschnitt 2.2.1) segmentiert.

Beispiele:

v: ins muSEum
n: Ins Museum .

v: ah
n: Ah .

v: WELches denn
n: Welches denn ?

v: weißt du WAS ich kauf mir n AUto
n: Weißt du was ? Ich kaufe mir ein Auto .

v: hm
n: Hm .

3.7 Interjektionen

Interjektionen (auch "Empfindungswörter") können dem Ausdruck von Emotionen, Flüchen, Verwünschungen sowie zur Kontaktaufnahme dienen (*Oh! Au! Verflixt!* etc.). Sie haben im strengeren Sinn keine lexikalische Bedeutung. Oft haben Interjektionen auch einen lautmalerischen Charakter (vgl. *Papperlapapp!*).

Interjektionen werden durch ein Ausrufezeichen (s. Abschnitt 2.2.2) abgetrennt! In Tab. 3 und Tab. 4 sind häufige Interjektionen und Angaben zu Transkription und Normalisierung aufgeführt.

v: ich hab papas auto gegen die WAND gesetzt
n: Ich habe Papas Auto gegen die Wand gesetzt .

v: ach du MEIne güte | ach du SCHEIße | SCHEIße | na TOLL | verDAMMT
n: Ach du meine Güte ! Ach du Scheiße ! Scheiße ! Na toll ! Verdammt !

v: ich hab gestern ACHT stunden gearbeitet
n: Ich habe gestern acht Stunden gearbeitet .

v: aHA | AU | PFUI
n: Aha ! Au ! Pfui !

Weitere Interjektionen (für die Verschriftungsformen auf n siehe Tabellen 3 und 4 sowie die gesonderte Übersicht "Normalisierung"):

v: EY | ACH | OH | ach SO | na JA
n: Ey ! Ach ! Oh ! Ach so ! Na ja !

Interjektionen werden mittels Ausrufezeichen segmentiert.

Transkription	Normalisierung
ach	ach
ah	ah
aha	aha
e:y	ey
ey	ey
grr	grr
hossa	hossa
ih	ih
i:h	ih
lol	LOL
na	na
oh	oh
oah	oah
ohoh	oho
'oh'oh	oh-oh
tja	tja
uh	uh
ui	ui

Tab. 3: Normalisierung von Interjektionen

Mehrwortausdruck
Ja doch
Ach so
Ach ja
Oh nein

Tab. 4: Normalisierung komplexer Interjektionen (Mehrwortausdrücke)

Konstruktionen wie *ach ja*, die als Interjektionen verwendet werden können, können im Diskurs aber auch Funktionen übernehmen, in denen sie intern segmentiert werden müssen, bspw. als Kombination von Interjektion und Antwort.

v: wolltest du nich noch das AUto waschen
n: Wolltest du nicht noch das Auto waschen ?
v: ACH JA
n: Ach ! Ja .

Wenn **Interjektionen in eine Äußerung integriert** auftreten, dann wird dies als Einschub analysiert. Das ist z.B. der Fall, wenn das Verb im zweiten und eines seiner Argumente im ersten Teil der Äußerung stehen. Die Interjektion wird in solchen Fällen zwischen Kommata gesetzt.

v: nachher ALLe bekommen OH die bekommen sogar LEICHte
n: Nachher alle bekommen , oh , die bekommen sogar Leichte .

Deutsche , vallah , sind eigentlich mehr als [...]

v: hat uns zwar ein so oah ein unser elfmeter nich geGEBN aber [...]
n: Hat uns zwar einen so , oah , einen unser Elfmeter nicht gegeben, aber [...]

Ein Korpusbeleg:

MuH1WD [v]: new YORK is doch OH RIChtig cool
MuH1WD [n]: New York ist doch , oh , richtig cool .

Besteht diese Abhängigkeit der beiden Teile der Äußerung nicht, bspw. wenn im zweiten Teil das Material des ersten wiederholt wird, wird die Struktur als Abbruch und die Interjektion als eigenes Segment analysiert.

Deutsche sind # Vallah ! Deutsche sind eigentlich mehr als

Im Gegensatz zur oben beschriebenen Abhängigkeit wird bei gefüllten Pausen nicht davon ausgegangen, dass diese die Struktur der Äußerung unterbrechen. Gefüllte Pausen können an jeder Position in einer Äußerung stehen. Im Beispiel unten wird deshalb von einer Korrektur ausgegangen (KOR auf der Macro-Ebene), bei der die gefüllte Pause als Interregnum zwischen dem Reparandum (dem reparierten Teil der Äußerung) und dem Reparans (der Reparatur, also der eigentlich intendierten Äußerung) fungiert.

v:	nachher ALLe bekommen	äh	die bekommen sogar LEICHte
n:	Nachher alle bekommen	äh	die bekommen sogar Leichte .
macro:		KOR	
	Reparandum	Interregnum	Reparans

Eine der häufigsten Interjektionen in KiDKo ist *Mann*. Wir behandeln *Mann*, wann immer es nicht als referentielles Nomen verwendet wird, als Interjektion. In diesen Kontexten wird es immer abgetrennt und mit Ausrufezeichen ausgezeichnet.

v: da müssen wir UNbedingt hin mann
n: Da müssen wir unbedingt hin . Mann !

Hör auf ! Mann !

Ja . Mann !

Ich bin voll auf seinen Schuh geflogen . Mann !

3.8 Partikeln: allgemeine Hinweise

Partikeln sind nicht immer zu segmentieren. Ein Beispiel für eine ausbleibende Segmentierung ist *ja* in der Funktion als Modalpartikel im Mittelfeld. Im Unterschied dazu wird *ja*, wenn es als Frage- oder Antwortpartikel verwendet wird, segmentiert: *Ja ? Ja* .

Ja ist also polyfunktional. Es kann auch initial vor einer Äußerung stehen. Hier muss auf Basis der kontextuellen Plausibilität, der Pausensetzung, der Akzentsetzung, und eventueller Tonhöhenbewegungen entschieden werden, ob *ja* segmentiert wird oder nicht. Dies betrifft auch andere Partikeln.

v: hast du ne FRage an mich
n: Hast du eine Frage an mich ?

v: JA bist du denn beSCHEUert
n: Ja . Bist du denn bescheuert ?

v: wir wollen auf den BERG klettern
n: Wir wollen auf den Berg klettern .

v: ja bist du denn beSCHEUert
n: Ja bist du denn bescheuert ?

v: NA KENnen wir uns nicht
n: Na ! Kennen wir uns nicht ?

Bei initialem *na* muss es sich nicht immer um eine Interjektion handeln, *na* kann auch als äußerungsinitale Gliederungspartikel verwendet werden. In diesem Fall wird *na* nicht segmentiert.

v: ich hab dein EIS vergessen
n: Ich habe dein Eis vergessen .
v: na dann kauf ich es mir SELber
n: Na dann kaufe ich es mir selber .

4 Phänomene gesprochener Sprache

4.1 Reihung

Reihungen werden durch Komma (s. Abschnitt 2.2.5) abgetrennt.

v: wohnst du hier schon LÄNger
n: Wohnst du hier schon länger ?
v: JA seit VIElen VIElen VIElen jahren
n: Ja. Seit vielen , vielen , vielen Jahren .

4.2 Wiederholung

Als Wiederholung bezeichnen wir ausschließlich die Wiederholung von lexikalisch identischem Material. Das wiederholte Material wird mit Komma getrennt (das gilt nicht für gefüllte Pausen).

v: dann dann müssen wir noch in den ZOO
n: Dann , dann müssen wir noch in den Zoo .
v: im urlaub im urlaub kann man sich SCHÖN entspann
n: Im Urlaub , im Urlaub kann man sich schön entspannen .
v: WAS isst du gern
n: Was isst du gern ?
v: weiße weiße schokoLade
n: Weiße , weiße Schokolade .

Unter diese Definition fällt auch lexikalisch identisches Material, zwischen dem eine gefüllte Pause steht.

n: Aber ich , ich glaube nicht ...
n: Aber ich , äh ich glaube nicht ...

Ein Korpusbeleg:

MuP6MD [v]: JA für den für den genau Abendschule oder sowas
MuP6MD [n]: Ja . Für den , für den genau Abendschule oder sowas .

Nicht als Wiederholung annotiert wird bspw. die folgende Äußerung, da das lexikalische Material nicht identisch ist. Hierbei handelt es sich um eine Reparatur bzw. Korrektur (s. den folgenden Abschnitt 4.3). Diese wird mit KOR auf der Macroebene gekennzeichnet.

v: nachher ALLe bekomm die bekomm sogar LEICHte
n: Nachher alle bekommen die bekommen sogar leichte.

4.3 Reparatur/Korrektur

Reparaturen werden auf der Transkriptionsebene so verschriftet wie gehört. Auf der Normalisierungsebene werden sie von der Transkriptionsebene übernommen, es wird keine Zielhypothese formuliert, wie die grammatikalische Version dieser Äußerung lauten könnte. Auf der Makroebene wird die gesamte Äußerung als Reparatur mit dem Tag *KOR* gekennzeichnet, unabhängig davon, wie groß der korrigierte Bereich ist (Wort, Phrase).

```
v:    und das hatte ich wollte mir doktor WISpel geben
n:    Und das hatte ich wollte mir Doktor Wispel geben .
macro: |                                KOR                                |
```

Die meisten Korrekturen beinhalten auch einen Abbruch (s. Abschnitt 4.4). Deshalb ist die Abgrenzung zwischen Korrektur und Abbruch nicht trivial. Wir sprechen von einer Korrektur, wenn eine Äußerung

- einen Abbruch enthält
- und die Reparatur zu einer Dopplung von Argumenten (Subjekt, Objekte) und/oder dem Kopf (Prädikat) der Äußerung führt.

Beispiele:

KOR: Prädikatdopplung

```
Er gehst ähm geht jetzt nach Hause .
Er geht ähm gehst jetzt nach Hause .
```

KOR: Subjekt- und Prädikatdopplung

```
Er geht ähm der geht jetzt nach Hause .
Er geht (-) Marlen geht jetzt nach Hause .
```

KOR: Dopplung des direkten Objekts

```
Ich habe Bier äh Wodka gekauft .
```

KOR: Dopplung des indirekten Objekts

```
Ich habe Jana äh Anne neue Dateien geschickt .
```

Es gibt auch Fälle, in denen die Dopplung von Lokal- und Temporalangaben, die meist nicht als Argumente, sondern als Adjunkte analysiert werden, als Korrektur annotiert werden sollten. Bedingung für die Annotation mit KOR ist aber, dass es sich dabei um die Korrektur durch eine Lokal- oder Temporalangabe derselben Art (ähnliche Semantik) handelt, die die erste Lokal- oder Temporalangabe "überschreibt" und nicht ergänzt.

KOR: Dopplung der Lokalangabe

```
Er geht jetzt nach Hause ähm ins Kino .
Sie steht auf dem Berg äh im Tal .
```

KOR: Dopplung der Zeitangabe

```
Ich treffe dich nach der Mittagspause äh vor der Mittagspause .
```

Grundsätzlich sind Orts- und Zeitangaben aber wiederholbar und korrigieren nicht zwangsläufig eine vorhergehende Orts- oder Zeitangabe, sondern modifizieren diese, indem sie weitere Informationen hinzufügen. In solchen Fällen wird kein KOR annotiert.

kein KOR: genauere Bestimmung der 1. Lokalangabe

```
Er geht jetzt nach Hause ähm in sein Bett .
Sie steht auf dem Berg äh neben dem Aussichtsturm .
```

kein KOR: genauere Bestimmung der 1. Zeitangabe

Ich treffe dich **nach der Mittagspause** äh **um 14 Uhr** .

Die Entscheidung, ob eine adverbiale Angabe durch eine folgende Angabe ersetzt/korrigiert oder näher bestimmt wird, ist nicht immer leicht zu treffen. In Fällen, in denen beide Lesarten plausibel sind, wird **kein KOR** annotiert.

Strukturelle Umformulierungen einer Äußerung werden ebenfalls als Korrektur annotiert, wenn sie zu Argument- und/oder Kopfdopplung führen.

Ich habe 100 Mails **habe ich** von ihm bekommen .

Ich habe 100 Mails geschrieben **habe ich** ihm .

Hier zwei Belege für KOR aus dem KiezDeutsch-Korpus:

```
v:      nachher ALLe bekommen die bekommen sogar LEICHte
n:      Nachher alle bekommen die bekommen sogar Leichte .
macro:  |                                KOR                                |
```

```
MuH12MD [v]:  is  meine tie stimme TIEF oder HOCH
MuH12MD [n]:  Ist meine tie Stimme tief oder hoch ?
macro:  |                                KOR                                |
```

In Abgrenzung dazu wird in den folgenden Fällen **nicht von einer Korrektur** ausgegangen, obwohl auch hier die Dopplung von Material vorkommt:

- Intensivierung

Ich bin **müde** , total **müde** .

- postnominale Modifikation

Ich hasse **Studenten** , diese ganz typischen **Studenten** .

Er ist ein **Spinner** , genau genommen ein betrunkenener **Spinner** .

- Hedging

Ich fand das Buch **langweilig** , irgendwie total **langweilig** .

Modifizierung einer Äußerung ohne Argument/Kopf/Adjunkt-Dopplung (bspw. durch Einschub mit genauerer Bestimmung *wenn er klein ist*) oder die Wiederholung von schon geäußertem Material (*nicht mal*)

Das willst du ja jetzt **nicht mal** , **wenn er klein ist** , nicht mal deinen Bruder andauernd dabei haben .

4.4 Abbruch

Ein Abbruch wird annotiert, wenn eine Interpretation als Korrektur/Reparatur nicht möglich ist, d.h. wenn nach dem Abbruch neu angesetzt und die begonnene Äußerung nicht beendet wird. Abbrüche unterscheiden sich von Reparaturen also dadurch, dass das abgebrochene Material nicht wieder aufgegriffen und in modifizierter Form verwendet wird. Sie werden durch die Raute (s. Abschnitt 2.2.7 Unspezifisches Grenzzeichen) abgetrennt.

Als Abbrüche werden nur Fälle annotiert, die keine Analyse als Reparatur/Korrektur (Abschnitt 4.3) zulassen!

v: das kleid von gabi is also ICH würde sowas nie kaufen
n: Das Kleid von Gabi ist # Also ich würde sowas nie kaufen .

Ein Korpusbeleg:

MuH1WD [v]: new YORK is doch OH RICHTig cool
MuH1WD [n]: New York ist doch # Oh ! Richtig cool .

Nicht eindeutig zu bestimmen sind Fälle, in denen nicht klar ist, ob die begonnene Äußerung in einer korrigierten oder umformulierten Version beendet werden soll oder ob es sich um eine komplett neue Äußerung handelt. Dies ist oft der Fall, wenn der Abbruch äüßerungsinitial steht.

v: dann habe ich dann fahre ich nach LONDON
n1: Dann habe ich # Dann fahre ich nach London .
n2: Dann habe ich dann fahre ich nach London .
| KOR |

Wir annotieren hier **keine Wiederholung**, da es sich nicht um identisches Material handelt (*dann habe ich* vs. *dann fahre ich*). In diesen Fällen muss die Transkriptorin entscheiden, welche Lesart im Kontext plausibler ist.

Weitere Beispiele:

Wahrscheinlich **haben sie sich** vielleicht **haben sie sich** auch abends gestritten
→ KOR: beinhalten Abbruch und Dopplung von Kopf und Subjekt

Äh Freitag **war ich** dings **war ich** mit SPK9 in einer Bar .
→ KOR: beinhaltet Abbruch und Dopplung von Kopf und Subjekt

SPK13 **wird halt** auch **erst im wird halt erst im** Juli 15 .
→ KOR: Abbruch und Kopfdopplung

Wo wir dann **mit** die ganze Zeit **mit** diesem Gymnast Gymnastikball waren .
→ KOR: Abbruch innerhalb einer Phrase, aber es wird nicht neu angesetzt, sondern die Äußerung wird weitergeführt

4.5 Spezialfall: Wortabbruch

Ein Abbruch innerhalb eines Wortes, bei dem das abgebrochene Wort im direkten Anschluss neu angesetzt und vollständig ausgesprochen wird, wird weder mit KOR noch mit # annotiert, sondern so verschriftet wie gehört (und zwar sowohl auf Transkriptions-, als auch auf Normalisierungsebene).

v: u und Offen fahren
n: U und offen fahren .

v: ha hat mir IRgendeiner erzählt
n: Ha hat mir irgendeiner erzählt .

v: er hat mich TY TYson genannt .
n: Er hat mich Ty Tyson genannt .

Die Abgrenzung zwischen Wortabbruch und Wiederholung ist schwierig, eine Entscheidung kann

Entscheidung für eine Variante getroffen werden.

Ein Beispiel:

v: äh SCHEIße

Hier besteht die Ambiguität zwischen Nomen, Adjektiv und Interjektion, was Auswirkungen auf die Groß- und Kleinschreibung auf Normalisierungsebene hat.

n: Äh scheiße .

n: Äh Scheiße .

Hier wird nach folgender Regel verfahren: Wenn eine Ambiguität zwischen Nomen, Adjektiv oder Interjektion besteht, die durch den Kontext nicht aufgelöst werden kann, dann wird immer eine Normalisierung als Adjektiv vorgenommen.

n: Äh scheiße .

Entscheidungshilfen zur Disambiguierung:

Nach Artikel und Pronomen: Großschreibung

v: ist das eine SCHEIße

n: Ist das eine Scheiße !

v: aber ich war gar nicht SATT von der scheiße

n: Aber ich war gar nicht satt von der Scheiße .

v: ich werde für JEde scheiße danke sagen

n: Ich werde für jede Scheiße Danke sagen .

Ersetzbar durch andere Adjektive: Kleinschreibung

v: sieht das SCHEIße aus

n: Sieht das scheiße aus !

Ambig: Kleinschreibung (Adjektiv)

v: das ist SCHEIße

n: Das ist scheiße . ←

n: Das ist Scheiße .

v: äh SCHEIße

n: Äh scheiße . ←

n: Äh Scheiße .

5.2.2 Einwortäußerungen und Akzent

Gefüllte Pausen und Rezeptionssignale tragen **nie** Akzent (s. die Abschnitte 3.5 und 3.6).

Gefüllte Pause:

Spk2 [v]: und DANN ähm (--) äh

Spk2 [n]: Und dann ähm äh #

Rezeptionssignal:

MuH11MD [v]: hm

MuH11MD [n]: Hm .

Nxxx [v] : HÖRST du

Nxxx [n] : Hörst du ?

Bei einsilbigen Einwortäußerungen, die einen Akzent tragen **können**, ist nicht klar, welches die Vergleichsgröße für den Primärakzent ist. In solchen Fällen wird ein Akzent nur annotiert, wenn er im Vergleich zum Primärakzent des vorangehenden Segments als solcher wahrnehmbar ist.

Einwortäußerungen, die einen Akzent tragen, werden auf der Normalisierungsebene durch . ! ? abgetrennt.

Korpusbeleg:

MuH1WD [v]: new YORK is doch OH RICHTig cool
MuH1WD [n]: New York ist doch # Oh ! Richtig cool .

Fragepartikeln und Interjektionen tragen immer einen Akzent.

Das weißt du, **NE** ?

5.3 Segmentierung von nebenordnenden/unterordnenden Konjunktionen

Generell wird vor Nebensätzen, die mit einer unterordnenden/subordinierenden Konjunktion eingeleitet werden und VL-Stellung aufweisen, ein Komma (s. 2.2.5) gesetzt.

v: ich bin müde weil ich gestern noch TANzen war
n: Ich bin müde , weil ich gestern noch tanzen war .

Wenn nach der Konjunktion ein V2-Satz folgt, dann analysieren wir die Konjunktion als nebenordnend und setzen kein Komma.

v: ich bin müde weil ich war gestern noch TANzen
n: Ich bin müde weil ich war gestern noch tanzen .

v: ich esse jetzt was oder ich gehe nach HAUse
n: Ich esse jetzt was oder ich gehe nach Hause .

Es gibt jedoch Fälle, in denen diese Regeln schwer anzuwenden sind:

(i) bei nebenordnenden Konjunktionen, wenn der Redefluss von Sprecher/in A durch Sprecher/in B unterbrochen wird:

A: Aber ich glaube , sie stört es auch nicht so richtig
B: Nein .
A: weil bei dir und
Mxxx ist es eher was anderes weil Mxxx ist ja noch deutlich kleiner als , als
Cxxx ist .

(ii) bei Aneinanderreihungen mehrerer nebenordnender Konjunktionen, die in einer extrem langen Äußerung resultieren:

v: na ich verSTEH das nich weil das is ja an sich total BEHINDert weil (-)
wir hat- (-) ten das Voriges jahr schon mal (-) und ganz EHRlich mir hat
das nichts geBRACHT weil was BRINGT einem dis wenn man keinen PLAN hat (-)

(iii) wenn keine Möglichkeit besteht, die Äußerung an die vorige Äußerung anzuschließen (unterschiedliche Sprechakttypen, weder Neben- noch Unterordnung):

n: Ha dann haben wir halt noch viel vom Tag . Weißt du ? **Weil** wir ja da noch halt schon ziemlich früh auch ankommen , weil wir ja schon so früh losfahren .

Es gibt verschiedene **Kriterien**, die Einfluss auf die Entscheidung nehmen, wann segmentiert wird:

- a) eingeschobener Redebeitrag einer anderen Person
- b) Pausen vor der Konjunktion
- c) Prosodie, Intonationsverlauf
- d) Semantik der Äußerung

Die folgenden Beispiele zeigen eine Normalisierung, in der von der Default-Normalisierung abgewichen wurde, und benennen die Gründe dafür. Es ist jedoch nicht möglich, den Einfluss dieser Faktoren hinreichend zu beschreiben und als allgemeingültige Richtlinien zu formulieren. Wir versuchen durch die folgenden Beispiele Entscheidungshilfen zu geben, und verlassen uns auf die Intuition der Transkribentinnen, die dem jeweiligen Kontext entsprechend eine Entscheidung treffen müssen.

n: Würde die doch sowieso nicht machen weil das dürfte sie auch gar nicht .
(-) Und Herr K3xxx , als ob der sie lassen würde . (-) Und als ob sie das machen würde .

→ hier wurden die **Pausen als Indiz für Segmentgrenzen** interpretiert

n: Na ich verstehe das nicht weil das ist ja an sich total behindert weil wir hatten das voriges Jahr schon mal . Und ganz ehrlich , mir hat das nichts gebracht . Weil was bringt einem das , wenn man keinen Plan hat ?

→ **Semantik der Äußerung als Indiz für Segmentgrenzen**

(Welche Teile der Äußerung gehören semantisch enger zusammen?)

n: (*Pausen stehen hier nur zur Verdeutlichung und gehören nicht auf die n-Ebene*)

A: Hat voll Spaß gemacht . (-)

B: Genau .

A: Und dieses Jahr müssen wir es unbedingt machen .

→ **Sprecherwechsel als Indiz für Segmentgrenzen**

5.4 Direkte Rede/Zitation/Bericht

Wir unterscheiden bei der Transkription zwischen

- direkter Rede (Doppelpunkt und Anführungszeichen)
- Zitation (nur Anführungszeichen)
- Bericht eigener Gedanken, Meinungen, von Selbstgesprächen, ... (mit bisherigem Segmentierungszeicheninventar)

5.4.1 Direkte Rede

Bei direkter Rede setzen wir Doppelpunkt und Anführungszeichen. Wir annotieren direkte Rede nur für Verba dicendi (Verben des Sagens) wie *sagen, erzählen, berichten*, bestimmte Verwendungen von *meinen* ... (nie für kognitive Verben wie *glauben, denken*, bestimmte Verwendungen von *meinen, urteilen*, etc.). Zusätzlich sind auch die Segmentierungsrichtlinien zu Doppelpunkt und Anführungszeichen (s. Abschnitt 2.2.6) zu beachten.

Bericht von Aussagen einer zweiten Person:

v: er SAGT so die blöde KUH
n: Er sagt so : " Die blöde Kuh ! " .

v: sie SACHT dann so was bildet der sich EIN
n: Sie sagt dann so : " Was bildet der sich ein ? " .

Bericht von Aussagen einer zweiten Person:

v: und dann has= DU gesagt warum kommt der nich endlich REIN
n: Und dann hast du gesagt : " Warum kommt der nicht endlich rein ? " .

Bei der Einleitung von direkter Rede wird das Verb oft ausgelassen:

v: SIE dann so was bildet der sich EIN
n: Sie dann so : " Was bildet der sich ein ? " .

v: und SIE so was bildet der sich EIN
n: Und sie so : " Was bildet der sich ein ? " .

Wir annotieren direkte Rede auch bei "**hypothetischer**" direkter Rede, bei der die Einleitung des Redebeitrags im Konjunktiv steht.

v: würdest du JETZ zu ihm sagen EY junge du bist total blöde
n: Würdest du jetzt zu ihm sagen : " Ey ! Junge ! Du bist total blöde . " .

Das Verb *meinen* kann sowohl als kognitives Verb als auch als Kommunikationsverb benutzt werden.

v: ich mein auch wir MÜSsen das nicht machen
n: Ich meine auch , wir müssen das nicht machen .

→ **kognitives Verb**

v: und dann meint er lass uns doch noch was TRINKen gehen
n: Und dann meint er : " Lass uns doch noch was trinken gehen . " .

→ **Verb des Sagens**

Direkte Rede kann auch mit **Pronomen der 1. Person (*ich, wir*)** verwendet werden. Die Abgrenzung zwischen direkter Rede und 'Wiedergabe eigener Gedanken/Selbstreflexion' ist hier manchmal schwierig und muss im jeweiligen Kontext entschieden werden.

v: isch sag so halt deine FRESse mann isch FAHR jetzt
n: Ich sage so : " Halt deine Fresse ! Mann ! Ich fahre jetzt . " .

v: ich MEINte zu ihr SCHREI nicht so und so
n: Ich meinte zu ihr : " Schrei nicht so ! " und so .

v: ich meinte nur so ach SO dein ERNST
n: Ich meinte nur so : " Ach so ! Dein Ernst ? " .

Eine **Ausnahme** zur Konvention, einen Doppelpunkt bei direkter Rede zu setzen, bildet der folgende Fall. Hier wird die Einleitung der direkten Rede als "Einschub" analysiert und der erste und zweite Teil der direkten Rede als Wiederholung.

v: SCHICK sag ich zu ihr schick mir mal n foto von dem couSIN
n: " Schick , " sage ich zu ihr , " schick mir mal ein Foto von dem Cousin ! " .

Kombinationen aus **Imperativ und direkter Rede** führen oft dazu, dass zwei unterschiedliche Satztypen in einer Äußerung eingebettet sind:

Ja sag mal : " Wo wohnt der ? " !
→ Imperativ + Interrogativ

Sag mir jetzt nicht : " Aus einer (unv.) # " !
→ Imperativ + abgebrochene Äußerung

Obwohl es sich hier nicht um tatsächliche direkte Rede, sondern um die Aufforderung zur direkten Rede handelt, normalisieren wir hier mit Doppelpunkt und Anführungszeichen wie bei direkter Rede.

5.4.2 Zitation

Bericht von Zitaten/geschriebenem Material:

v: da steht drunter mir geFÄLLT =s
n: Da steht drunter " Mir gefällt es . " .
v: WAS hat er geschrieben man sieht sich aber nur ZWEI mal im leben
n: Was hat er geschrieben ? " Man sieht sich aber nur zwei Mal im Leben . "

5.4.3 Bericht eigener Äußerungen/Selbstreflexion

Werden eigene Äußerungen oder Gedanken berichtet, so annotieren wir keine direkte Rede. Oft werden dazu Kognitionsverben wie *meinen, glauben, denken, urteilen* usw. benutzt.

v: isch SACH mir dann hab geDULD
n: Ich sage mir dann , hab Geduld .
v: isch guck so in mein GÄStebuch äh warum hab isch das so geMACHT
n: Ich gucke so in mein Gästebuch . Äh ? Warum hab ich das so gemacht ?

5.4.4 Sprecher/in wiederholt Material von Vorredner/in

In Fällen, in denen Sprecher/innen Äußerungen wiederholen (z.B. nachäffen), die vorher eine andere Person geäußert hat, transkribieren wir diese in doppelten Anführungszeichen:

A v: ich MAG dis , wenn de mich FERTig machst
A n: Ich mag das , wenn du mich fertig machst
B v: nein
B n: Nein .
A v: SAG mal
A n: Sag mal !
B v: ich MAG dis wenn de mich FERTigmachst
B n: " Ich mag das , wenn du mich fertigmachst . "

5.5 Unterbrechungen durch Pausen

5.5.1 Unterbrechung von Wörtern

Für die Unterbrechung von Komposita gelten gesonderte Richtlinien (s. Abschnitt 1.5.1)!

Auch Wörter, die nicht in die Klasse der Komposita fallen, können durch (gefüllte oder stille) Pausen unterbrochen werden. Diese werden folgendermaßen behandelt:

MuH27WT [v]: heute ist der achtzehn- (-) te
MuH27WT [n]: Heute ist der 18. LEER

MuP1MK [v]: über- (-) treiber
MuP1MK [v]: Übertreiber LEER

Ähnlich wie bei den unterbrochenen Komposita wird der erste Bestandteil des Wortes am Ende mit einem Bindestrich versehen (innerhalb eines Events). Der "Rest" des Wortes, der durch die Pause abgetrennt wird, wird auf der Normalisierungsebene mit dem tag *LEER* versehen und bekommt auf v keine besondere Markierung. Die Pause hat keine Entsprechung auf der Normalisierungsebene.

5.5.2 Pausen innerhalb einer Äußerung

Bei Fällen wie dem folgenden, in denen eine Äußerung unterbrochen und später fortgeführt wird, werden die Events bei Speaker A nicht durch Pausen (-) gefüllt, sondern leer gelassen.

A: isch hab doch hab doch ...
B: JA endlich TO:R

5.6 Groß-/Kleinschreibung

Bei **Farbbezeichnungen** ist oft unklar, ob es sich um ein Adjektiv oder um ein Nomen handelt.

Kopula + Farbbezeichnung (prädikative Zuweisung): Adjektiv

Die Ampel ist rot /ADJD .
Test: Wie ist die Ampel?

Nach Präpositionen: Nomen

Nur bei Rot /NN gehen .
Die Ampel steht auf Rot /NN .

Nach Artikel: Nomen

Das Blau /NN ist mir zu grell .
ein helles Grün / ein zartes Lila

Andere nominale Verwendungen (Farbbezeichnungen):

Rot /NN ist Trumpf .
Die Farbe Rot /NN .
Sie trägt gerne Rot /NN .
Ich habe Rot /NN .

Auch bei den **Verwendungen von *dings*** muss zwischen Groß- und Kleinschreibung unterschieden

werden. Wenn der Kontext nicht ausreicht, um die Wortart zu disambiguieren, so schreiben wir die Token auf Normalisierungsebene klein.

v: ich hab äh DINGS hier
n: Ich habe äh Dings hier .

v: der mann wollte DINGs
n: Der Mann wollte dings #

v: erzÄHL doch nichts
n: Erzähl doch nichts !

Wenn vor *Dings* ein Artikel oder eine Präposition steht, so wird *Dings* als Nomen interpretiert und großgeschrieben.

v: ich hab äh das DINGS hier
n: Ich habe äh das Dings hier .

v: der dings kommt heut zum ESsen
n: Der Dings kommt heute zum Essen .

v: beim DINGS ist er hinjefallen
n: Beim Dings ist er hingefallen .

Andere Kontexte, in denen *dings* vorkommen kann:

v: da is dein HÖRgedings .
n: Da ist dein Hörgedings . => Teil eines Kompositums, Großschreibung

v: ich habe äh geDINGST
n: Ich habe äh gedingst . => Partizip, Kleinschreibung

v: wir müssen DINGsen
n: Wir müssen dingsen . => infinites Verb, Kleinschreibung

Attributive Adjektive werden kleingeschrieben, wenn sie sich auf ein vorangehendes oder folgendes Nomen beziehen (*große hat Pullis* als Bezugsnomen).

v: ich hab KLEIne pullis gekauft und GROße
n: Ich habe kleine Pullis gekauft und große .

Dies gilt **nicht über die Satzgrenze hinaus**. Ist kein Bezugsnomen vorhanden, so gehen wir von einer nominalisierten Form aus und verwenden Großschreibung:

v: sind das ALLe farbstifte hast du keine GELben und ROTen mehr
n: Sind das alle Farbstifte ? Hast du keine Gelben und Roten mehr ?

v: ich hab KLEIne pullis gekauft
n: Ich habe kleine Pullis gekauft .
v: und ich GROße
n: Und ich Große .

Nominalisierung: Als Nomen verwendete Adjektive und adjektivische Partizipien werden großgeschrieben.

v: heute gibt =s SÜßes und SAUres
n: Heute gibt es Süßes und Saures .

v: der verPRÜgelte ist nach hause gerannt
n: Der Verprügelte ist nach Hause gerannt .

Referenzen

Schmidt, Thomas & Kai Wörner (2009). EXMARaLDA – Creating, analyzing and sharing spoken language corpora for pragmatic research. *Pragmatics*, 19, pp. 565–582.

Selting, Margaret et al. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion* 10, 2009: 353–402.

Wiese, Heike (2009). Grammatical innovation in multiethnic urban europe: New linguistic practices among adolescents. *Lingua*, 119:782–806.