

Von der TOP 20-Liste aus dem Primärkorpus konnten die vier fehlenden Lemma-Types (*auch, da, mal, und*) unter den TOP 50 wiedergefunden werden.

Von den 14 abweichenden Lemma-Types in der TOP 50-Liste konnten 10 Lemma-Types in der TOP 100-Liste identifiziert werden (*dann, dürfen, es, in, jetzt, kein, mir, mit, müssen, schon*; fehlend: *den, man, oh, werden*).

Die prozentuale Übereinstimmung zwischen den TOP 100-Listen betrug 59.02 % (61 Lemma-Types). Im Primärkorpus fehlten insgesamt 40 Lemma-Types auf der TOP 100-Liste im Vergleich zum Referenzkorpus. Von diesen tauchten wiederum 35 Lemma-Types in der Gesamtliste auf (fehlend: *ah, aufnehmen, boah, ey, oh*) (Tab. 48).

Tab. 48: Vergleichsanalyse TOP 100: fehlende Lemma-Types im Primärkorpus und Position in der Gesamtliste (H = Häufigkeit; R = Rang)

Lemma-Types	H	R	Lemma-Types	H	R
ach	1	562	ihr	16	115
Affe	9	185	immer	16	115
also	2	429	lang	2	429
brauchen	9	185	man	6	243
dass	2	429	mehr	9	185
dem	12	152	Minute	13	140
den	11	161	ne	7	223
denn	3	355	sehen	4	316
dich	11	161	sollen	13	140
dies	17	108	Stunde	2	429
Ding	2	429	uns	9	185
dir	15	123	viel	11	161
drei	15	123	voll	2	429
ganz	6	243	weil	3	355
gar	7	223	wenn	15	123
gerade	1	562	werden	4	316
gleich	7	223	wieder	9	185
hören	11	161			

12.2 Feste Wortkombinationen

Insgesamt konnten 2 998 *Dreiwortkombinationen* (Token) im Primärkorpus identifiziert werden. Von diesen wurden 1976 *unterschiedliche Dreiwortkombinationen* (Types-Kombinationen) verwendet. Am häufigsten wurde die Dreiwortkombination *Was ist das* (H = 72) verwendet. Der höchste Rang lag bei $R_{\max} = 384$. Von den 1 976 unterschiedlichen Dreiwortkombinationen waren 383 Dreiwortkombinationen mit einem Anteil

von 19.38 % wiederkehrend ($H = >1$). 80.62 % der Dreiwortkombinationen (1 593 Types-Kombinationen) wurden nur einmal ($H = 1$) verwendet.

In Anlehnung an den Verlauf des Liniendiagramms (Abflachen der Kurve) wurden von den insgesamt genutzten Dreiwortkombinationen 30.95 % (928 Token) hochfrequent genutzt (Abb. 39). Die Wortkombination *ruhig mit weiter* wurde nur noch dreimal ($H = 3$) genutzt ($R = 95$). Zu den TOP 95 zählten 149 Types-Kombinationen.

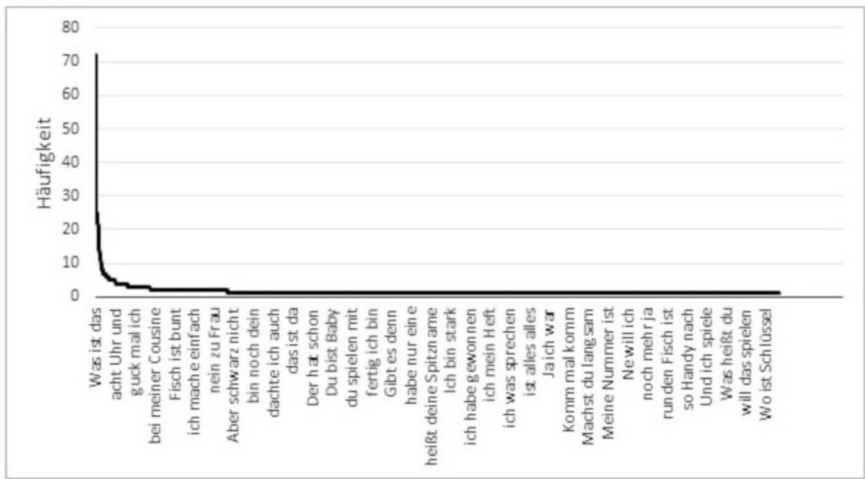


Abb. 39: Absolute Häufigkeiten Dreiwortkombinationen im Primärkorpus ($N = 22$, $\Sigma = 2\,998$ Token)
Anmerkungen: Ränge (R): Was ist das $R = 1$; acht Uhr und $R = 58$; guck mal ich $R = 95$; bei meiner Cousine $R = 159$; Fisch ist bunt $R = 159$; ich mache einfach $R = 159$; nein zu Frau $R = 159$; Aber schwarz nicht $R = 384$; bin noch dein $R = 384$

Ein feineres Bild über den Gebrauch der Dreiwortkombinationen ergab die Untersuchung der TOP 10-, TOP 20-, TOP 30- und TOP 50-Listen. Zu den TOP 10 der Dreiwortkombinationen zählten 292 Token und 10 Types-Kombinationen ($H_{max} = 72$; $H_{min} = 16$). Unter den TOP 20 konnten 416 Token und 22 Types-Kombinationen identifiziert werden ($H_{max} = 72$; $H_{min} = 8$). Zu den TOP 30 zählten 520 Token und 38 Types-Kombinationen ($H_{max} = 72$; $H_{min} = 6$). Innerhalb der TOP 50 wurden 763 Token und 94 Types-Kombinationen erhoben ($H_{max} = 72$; $H_{min} = 4$) (Tab. 49).

Tab. 49: Deskriptive Analyse Dreiwortkombinationen im Primärkorpus: TOP 10, TOP 20 und TOP 50

	TOP 10	TOP 20	TOP 30	TOP 50
R_{max}	9	20	31	58
Token	292	416	520	763

	TOP 10	TOP 20	TOP 30	TOP 50
Types-Kombinationen	10	22	38	94
$H_{\max}$	72	72	72	72
$H_{\min}$	16	8	6	4
Dreiwortkombination	Ich bin fertig	ich will nicht	nein das ist	wo ist meiner

Anmerkung: Dreiwortkombination: Gemeint ist die letzte Dreiwortkombination innerhalb der betrachteten Rangfolge.)

Anhand der Verteilung der Ränge wurde deutlich, dass der Abstand zwischen den Rängen ab Rang 39 (zwischen TOP 30 und TOP 50) mit dem Anstieg der Types-Kombinationen und der Abnahme der Häufigkeit ($H = 5$) sprunghaft zunahm ($M = 86$ Ränge, Tab. 50).

Tab. 50: Rangverteilung innerhalb der Dreiwortkombinationen im Primärkorpus ($R = \text{Rang}$, $H = \text{Häufigkeit}$)

R	H	Types-Kombinationen	R	H	Types-Kombinationen
1	72	1	16	10	1
2	50	1	17	9	3
3	25	2	20	8	3
5	23	1	23	7	8
6	22	2	31	6	8
8	21	8	39	5	19
9	16	2	58	4	37
11	14	1	95	3	64
12	13	2	159	2	224
14	12	1	384	1	1593
15	11	1			

Vor dem Hintergrund dieser Beobachtung und der Analyse des Liniendiagramms (Abb. 39) zur Bestimmung von festen Wortkombinationen ließ sich die Annahme ableiten, dass die Dreiwortkombinationen innerhalb der TOP 31 überwiegend ganzheitlich gebildet wurden. Der cut-off point für den Übergang zu flexibel verwendeten Dreiwortkombinationen wurde nach den TOP 31 identifiziert. Insgesamt wurden 38 *Types-Kombinationen als feste Dreiwortkombinationen* bewertet, die einen Anteil von 17.34 % des untersuchten Korpus ausmachten (Tab. 51).

Tab. 51: Feste Wortkombinationen im Primärkorpus (H = Häufigkeit; S = Streuung; R = Rang)

feste Wortkombinationen	H	S	R
Was ist das	72	14	1
Wie heißt du	50	10	2
Das ist meine	25	5	3
Ich weiß nicht	25	6	3
alt bist du	23	6	5
Das ist nicht	22	7	6
Wie alt bist	22	5	6
ich kann nicht	21	6	8
Das ist ein	16	7	9
Ich bin fertig	16	6	9
Ich hab Nummer	14	3	11
alles gut alles	13	1	12
gut alles gut	13	1	12
Das ist und	12	1	14
Ich bin nicht	11	4	15
Was machst du	10	5	16
bist ein Monster	9	1	17
du bist ein	9	1	17
Ich bin das	9	2	17

feste Wortkombinationen	H	S	R
ein riesiges Schloss	8	1	20
Ich kann das	8	2	20
ich will nicht	8	5	20
Das ist mein	7	4	23
das ist zu	7	1	23
der Drache hatte	7	1	23
Ich bin meine	7	1	23
Ich kann auch	7	2	23
Was hast du	7	4	23
Was ist los	7	4	23
Was kannst du	7	2	23
Fünf Jahre alt	6	2	31
Ich bin da	6	3	31
Ich bin Polizei	6	2	31
Ich habe kein	6	2	31
Ich habe schon	6	3	31
ich will spielen	6	3	31
ist nicht gut	6	3	31
nein das ist	6	3	31

Unter Anwendung des Streuungskriteriums wurde deutlich, dass nur die Wortkombination *Was ist das* von mindestens 50 % der Untersuchungsgruppe verwendet wurde ($n = 14$).

Der Gebrauch des Verbs *sein* und dessen konjugierte Verwendung fiel besonders auf (*bin* $n = 6$, *bist* $n = 3$, *ist* $n = 10$). Demzufolge wurde in 50 % der verwendeten Dreiwortkombinationen eine Wortform von *sein* genutzt. Das Funktionswort *das* wurde in ca. einem Viertel ($n = 10$) der festen Dreiwortkombinationen verwendet. Insgesamt wurde deutlich, dass nur drei Dreiwortkombinationen nicht ausschließlich aus dem Kernvokabular entsprechend der 80 %-Marke (Tab. 40) bestanden (*bist ein **Monster***, *ein **riesiges Schloss***, ***Fünf** Jahre alt*, in **fett** Wörter aus dem Randvokabular).

Vergleichsanalyse

Referenzkorpus: Identifizierung der festen Wortkombinationen

Im Referenzkorpus konnten insgesamt 25 130 *Dreiwortkombinationen* (Token) identifiziert werden sowie 2 058 *Types-Kombinationen*. Gemessen an den Token war der Referenzkorpus ca. acht Mal größer als der Primärkorpus. Am häufigsten wurde die Dreiwortkom-

bination *Guck mal hier* ($H = 43$) verwendet. Der höchste Rangplatz lag bei $R = 2\,610$. Von den insgesamt genutzten Dreiwortkombinationen wurden auf Grundlage der Analyse des Liniendiagramms (Abflachen der Kurve) 8,78 % (3 306 Token) hochfrequent genutzt (Abb. 40). Die Wortkombination *ein bisschen schneller* wurde nur noch $H = 3$ genutzt (Rang 454). Zu den TOP 454 zählten 591 Types-Kombinationen.

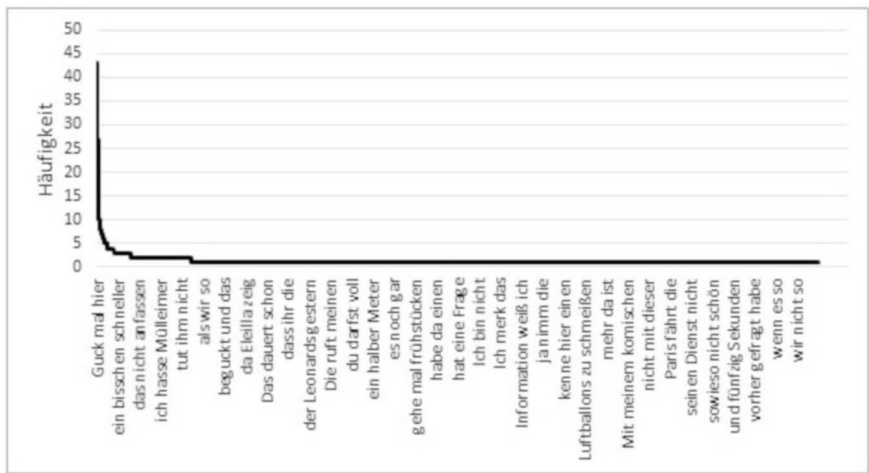


Abb. 40: Absolute Häufigkeiten: Dreiwortkombinationen Referenzkorpus ($N = 28$, $\Sigma = 25\,130$ Token)

Anmerkungen: Ränge (R): Guck mal hier $R = 1$; ein bisschen schneller $R = 454$; das nicht anfassen $R = 931$; ich hasse Mülleimer $R = 931$; tut ihm nicht $R = 931$; als wir so $R = 2\,610$

Zu den TOP 10 der Dreiwortkombinationen zählten 267 Token und 10 Types-Kombinationen ($H_{\max} = 43$; $H_{\min} = 21$). Unter den TOP 20 konnten 513 Token und 26 Types-Kombinationen identifiziert werden ($H_{\max} = 43$; $H_{\min} = 14$). Zu den TOP 30 zählten 599 Token und 33 Types-Kombinationen ($H_{\max} = 43$; $H_{\min} = 12$). Innerhalb der TOP 50 wurden 987 Token und 73 Types-Kombinationen erhoben ($H_{\max} = 43$; $H_{\min} = 9$) (Tab. 52).

Tab. 52: Deskriptive Analyse der Dreiwortkombinationen im Referenzkorpus: TOP 10, TOP 20 und TOP 50

	TOP 10	TOP 20	TOP 30	TOP 50
R	10	20	29	55
Token	267	513	599	987
Types-Kombinationen	10	26	33	73
$H_{\max}$	43	43	43	43
$H_{\min}$	21	14	12	9

	TOP 10	TOP 20	TOP 30	TOP 50
Dreiwortkombination	Ich habe auch	Wo ist der	zehn Jahre alt	Weiß ich nicht

Anmerkung: Dreiwortkombination: Gemeint ist die letzte Dreiwortkombination innerhalb der betrachteten Rangfolge.)

Die Rangfolge nahm nach Rang 34 sprunghaft zu, da immer mehr Types-Kombinationen mit einer geringeren Häufigkeit genutzt wurden ($M = 259$ Ränge, Tab. 53). Dieser Bereich wurde unter Verwendung des Häufigkeitskriteriums als mutmaßlicher cut-off point für den Übergang zu flexibel gebildeten Äußerungen gewertet.

Tab. 53: Rangverteilung innerhalb der Dreiwortkombinationen im Referenzkorpus (R = Rang, H = Häufigkeit)

R	H	Types-Kombinationen
1	43	1
2	31	1
3	27	2
5	26	1
6	24	2
8	22	2
10	21	1
11	19	2
13	17	1
14	16	3
17	15	3
20	14	7
27	13	2

R	H	Types-Kombinationen
29	12	5
34	11	7
41	10	14
55	9	19
74	8	21
95	7	33
128	6	42
170	5	118
288	4	166
454	3	477
931	2	1 679
2 610	1	17 449

Insgesamt wurden 40 Types-Kombinationen als feste Dreiwortkombinationen gezählt (Token-Anteil = 2.69 %).

Die Gegenüberstellung von Primärkorpus und Referenzkorpus ergab hinsichtlich des cut-off points trotz der unterschiedlichen Tokenzahlen (+22 132 Token im Referenzkorpus) einen ähnlichen Trend. Ca. 40 Types-Kombinationen wurden jeweils als feste Wortkombinationen eingestuft.

Qualitative Analyse

In einer qualitativen Analyse wurden die Übereinstimmungen in den TOP 10, TOP 20, TOP 30 und TOP 50 herausgearbeitet (Tab. 54).

Tab. 54: Übereinstimmungen in den festen Dreiwortkombinationen TOP 10 bis TOP 50 (*fett*: neue Dreiwortkombination) zwischen Primärkorpus und Referenzkorpus

	Top 10	Top 20	Top 30	Top 50
Absoluter Anteil	1	2	4	13
Dreiwortkombination	Was ist das	Was ist das Was machst du	Das ist mein/e ich kann nicht Was ist das Was machst du	Das ist meine/e Das ist nicht, Das ist so Ich bin ein Ich/ich habe das Ich/ich habe kein/e Ich habe schon ich kann nicht ich mache das Was hast du Was ist das Was machst du Wo ist meine/r

Anhand der Auswertung ließ sich erkennen, dass die Übereinstimmung an Dreiwortkombinationen mit Anstieg der Ranghöhe zunahm (TOP 10: 1; TOP 20: 2; TOP 30: 4; TOP 50: 13).

Wurde zusätzlich die Gegenüberstellung *auf Sequenzen mit offenem Feld (Slot) reduziert*, ließ sich eine höhere Übereinstimmung im Gebrauch der festen Wortkombinationen bei den TOP 10 (+2), TOP 20 (+3) und den TOP 30 (+2) nachweisen (Tab. 55). Innerhalb der TOP 50 konnte keine höhere Übereinstimmung ermittelt werden, wenngleich weitere und damit differenziertere Gebrauchsweisen erhoben werden konnten (*hast du* __; *kann ich* __; *Wie heißt* __; *Wir haben* __).

Mithilfe der Gegenüberstellung ließ sich eine auffällige Tendenz innerhalb der Dreiwortkombinationen, sowohl mit als auch ohne offenem Feld erkennen: Die Dreiwortkombinationen mit offenem Feld *Ich bin* __ und *Das ist* __ waren bereits unter den TOP 10 zu finden. Als Dreiwortkombinationen ohne offenem Feld konnten die beiden Formulierungen erst unter den TOP 50 identifiziert werden. Auch die Äußerung *Ich hab(e)* __ wurde mit offenem Feld bereits unter den TOP 20 gefunden, wenngleich die Wortkombination ohne offenem Feld erst unter den TOP 50 zu finden war.

Tab. 55: Übereinstimmungen in den festen Dreiwortkombinationen mit offenem Feld TOP 10 bis TOP 50 (**fett**: neue Dreiwortkombination mit offenem Feld)

	Top 10	Top 20	Top 30	Top 50
Absoluter Anteil	3	5	6	13
Dreiwortkombination mit Slot	Das ist __ Ich bin __ Was ist __	Das ist __ Ich bin __ Ich hab/e __ Was ist __ Was machst __	Das ist __ Ich bin __ Ich hab/e __ Ich kann __ Was ist __ Was machst __	Das ist __ hast du __ Ich bin __ Ich hab(e) __ Ich kann __ Ich mache __ kann ich __ Was hast __ Was ist __ Was machst __ Wie heißt __ Wir haben __ Wo ist __

