

Über die Analyse der Häufigkeit-Rang-Beziehung und dem Verlauf der Zipf-Kurve wurde ermittelt, welche Wortkombinationen am häufigsten verwendet wurden und als fest eingestuft werden konnten (heuristischer Zugang). Als cut-off point wurde das rapide Abflachen der Verlaufskurve sowie die Rangverteilung verwendet, um einen möglichen Übergang von festen Wortkombinationen hin zu flexibel gebildeten Wortkombinationen zu identifizieren. Für die gewonnenen und als ganzheitlich eingestuften Dreiwortkombinationen wurde die Streuung berechnet (Wie viele Kinder nutzten die Dreiwortkombination?). Dafür wurde im Editor von MaxDictio zur Berechnung der Wortkombinationen die *Aufgliederung nach Dokumenten* ausgewählt, sodass nachvollzogen werden konnte, in welchem Transkript die Wortkombination genutzt wurde. Auf dieser Grundlage konnte die Anzahl der Streuung ermittelt werden.

Vergleichsanalyse

Bei der Analyse von Lernersprache ist die Hinzunahme eines Referenzkorpus von Personen mit Deutsch als Erstsprache erforderlich (Kap. 10.1). Daher wurde der Referenzdatensatz »Grundschule« aus dem Datensatz von Boenisch (2014b; 2013) herangezogen (Referenzkorpus). Das zur Verfügung gestellte Material bestand aus einem Worddokument mit Transkriptionen von $N = 28$ Schüler:innen sowie einer Excel-Liste mit dem aufbereiteten Wortschatz (Lemma-Types) nach Häufigkeit, Frequenz und Rängen. Die Kernvokabularliste aus dem Primärkorpus wurde mit dem Referenzkorpus über die Funktion S-Verweis in Excel verglichen. Gemeinsamkeiten und Unterschiede im Kern- und Randvokabular konnten auf diesem Weg deskriptiv herausgearbeitet werden. In weiteren Analysen wurden die am häufigsten verwendeten Wörter (TOP-Wörter) verglichen sowie die Wortartenverteilungen gegenübergestellt.

Damit der Referenzdatensatz auch für die Vergleichsanalyse der festen Wortkombinationen verwendet werden konnte, wurden die Transkriptionen nach dem oben beschriebenen Schema in MaxDictio ausgewertet. In der Vergleichsanalyse wurden die festen Wortkombinationen aus Primär- und Referenzkorpus in Excel gegenübergestellt und Gemeinsamkeiten und Unterschiede herausgearbeitet.

11.5 Prüfung der Reliabilität, Objektivität und Validität

Die Interrater-Reliabilität für die Gewinnung der Kernvokabularliste wurde über den Übereinstimmungskoeffizienten (Anzahl der übereinstimmenden Fälle geteilt durch Gesamtheit der analysierten Fälle) berechnet (Hirschmann, 2019, S. 98). Dafür wurden 10 % ($n = 2$) der Sprachaufnahmen aus dem Datensatz ($N = 22$) randomisiert ausgewählt, erneut transkribiert und als Wortlisten aufbereitet. Der Übereinstimmungskoeffizient der Wortschatzlisten entsprach bei den Token $r_{ü} = 0.99$ (99 %) und bei den Lemma-Types $r_{ü} = 1$ (100 %). Damit ließ sich die Reliabilität in Bezug auf die formalsprachlichen Parameter als sehr hoch einstufen. Die Reliabilitätsprüfung sagt jedoch noch nichts über die Übereinstimmung auf Wortebene (qualitative Ebene) aus. Um die Reliabilität auf der qualitativen Ebene einschätzen zu können, wurde die Abweichung der Transkripte im Hinblick

auf die Lemma-Types, Wortartenzuordnung, Rechtschreibung und Datenaufbereitung geprüft. Insgesamt konnten 14 Abweichungen festgestellt werden (Tab. 34).

Tab. 34: Reliabilitätsprüfung (n = 2)

	Reliabilitätsprüfung (qualitative Ebene)	
	Absolut	prozentuale Übereinstimmung
Lemma-Types	9	93.23 %
Wortartenzuordnung	3	97.74 %
Rechtschreibung	1	99.25 %
Datenaufbereitung	1	99.25 %
<i>gesamt</i>	$\Sigma = 14$	$M = 97.37 \%$

Im Hinblick auf das Setting der Datenerhebung (u.a. geräuschvolle Kulisse, keine Videoaufnahme) lässt sich die Reliabilitätsüberprüfung als sehr hoch bewerten (Wendt & Schlosser, 2020, S. 409). Die Reliabilität vergleichbarer Studien lag niedriger (Boenisch & Soto, 2015, S. 79: 94 % Transkriptionen, 93 % Wortartenzuordnung; Marvin et al., 1994, S. 229: 84 %-91 % Transkriptionen; Robillard et al., 2014, S. 6: 70 % Transkriptionen).

Da für die Analyse der Wortkombinationen keine manuellen Veränderungen an den von MaxDictio generierten Daten erforderlich war, wurde auf eine Reliabilitätsprüfung verzichtet.

