

Fälle	Geschlecht	Erstsprache	Sprachaufnahme	Token	Lemma-Types
So86	w	bosnisch	02:54:24 (174 min)	223	85
So94	w	griechisch/ albanisch	02:29:08 (149 min)	694	145
So99	w	italienisch	02:20:13 (140 min)	50	18
Si100	w	italienisch	02:40:38 (160 min)	53	34
Si116	m	kurdisch	02:42:08 (162 min)	1 010	144
Si25	w	arabisch	01:55:45 (175 min)	317	104
N = 22	w = 17 m = 4 k.A. = 1	verschiede- ne Sprachen n = 9	53:54:10 Std. (3 276 min) (M = 2:27:00 Std.; 149 min)		

Aus dem Datensatz ging hervor, dass der Sprachgebrauch der Untersuchungsgruppe gemessen an den Token und der Lemma-Types als heterogen einzuschätzen war. Um die Variation der Datenmenge transparenter abzubilden und die verschiedenen Lageparameter zu beschreiben, wurden Boxplots erstellt.

Token

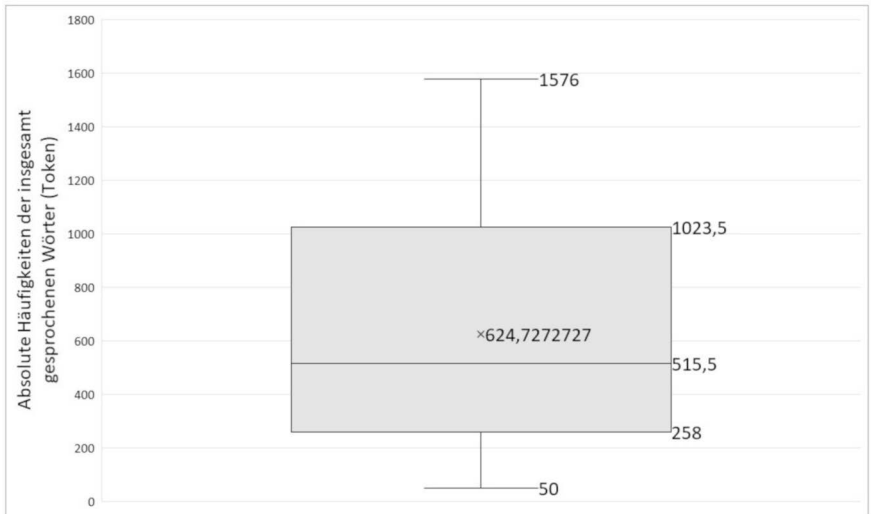


Abb. 21: Verteilung der Token im Primärkorpus (N = 22)

Die Hälfte der Untersuchungsgruppe sprach zwischen dem Interquartilsabstand (IQa) IQA₁ = 258 und IQA₃ = 1 023,5 Token. Der Maximalwert lag bei x_{max} = 1 576 Token.

Für den Minimalwert wurden $x_{\min} = 50$ Token erfasst. Die Spannweite der insgesamt genutzten Token betrug $R = 1\,526$. Der Durchschnitt lag bei $M = 624.72$ Token und der Median lag bei $x_{\text{med}} = 515.5$ Token (Abb. 21).

Lemma-Types

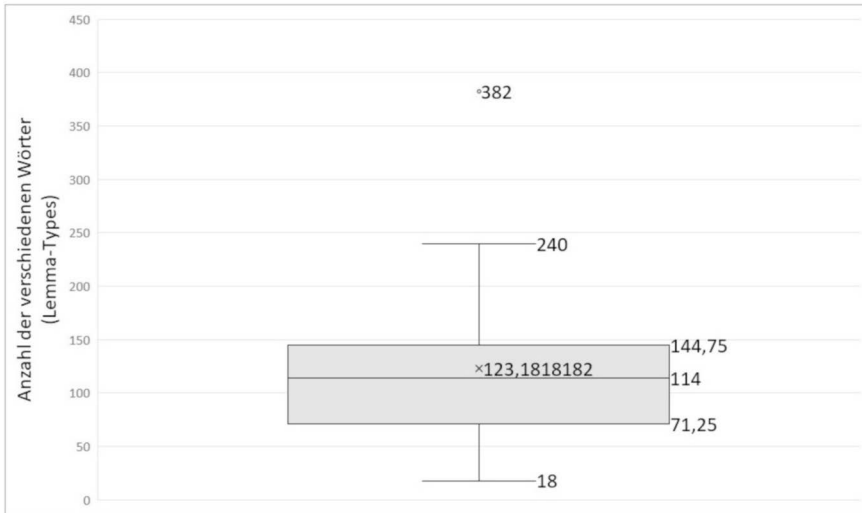


Abb. 22: Verteilung der erfassten Lemma-Types im Primärkorpus ($N = 22$)

Die Hälfte der Untersuchungsgruppe nutzte zwischen dem Interquartilsabstand (IQA) $IQA_1 = 71.25$ und $IQA_3 = 144.75$ Lemma-Types. Der Maximalwert lag bei $x_{\max} = 240$ Lemma-Types. Für den Minimalwert wurden $x_{\min} = 18$ Lemma-Types erfasst. Die Spannweite der Lemma-Types betrug $R = 222$. Der Durchschnitt lag bei $M = 123.18$ Lemma-Types und der Median lag bei $x_{\text{med}} = 114$ Lemma-Types. Der Wert 382 Lemma-Types wurde gemäß der Beschreibung von Boxplots als Ausreißer gewertet. Vermutlich lagen die sprachlich-kommunikativen Fähigkeiten des Schülers/der Schülerin über denen der Untersuchungsgruppe (Abb. 22).

12.1 Kernvokabular

Einzelauswertung Primärkorpus

Das Kernvokabular wurde pro Fall nach der 80 %-Marke von Boenisch (2014b) berechnet. Bei der Betrachtung der 80 %-Marke in Relation zur Häufigkeit pro Lemma-Type konnte eine verhältnismäßig geringe Anzahl an Mindestnennungen (H) identifiziert werden. Mithilfe der Analyse der Lemma-Types Beziehungen zeigte sich, dass bei über der Hälfte