

SEITENEINSTEIGER:INNEN-KORPUS

Sprachentwicklung neu zugewanderter Schüler:innen in Intensivklassen

SeiKo

Julia Schlauch
Aylin Braunewell
Jana Gamper

KONTEXT & ZIELE

Das Korpus umfasst mündliche und schriftliche Sprachdaten von neu zugewanderten Kindern und Jugendlichen in hessischen Intensivklassen. Diese sogenannten *Seiteneinsteiger:innen* stehen vor der Herausforderung, innerhalb möglichst kurzer Zeit sowohl basale Sprachkompetenzen als auch spezifische Registerkompetenzen für den schulischen Kontext zu entwickeln. Ziel des Korpus ist es, die systematische Untersuchung dieser Spracherwerbsprozesse zu ermöglichen.

FORSCHUNGSFRAGEN

1. Wie erwerben Seiteneinsteiger:innen grundlegende grammatische Kompetenzen (Verbstellung)?
2. Wie erwerben Seiteneinsteiger:innen nominalgruppeninterne Attribution und alternative Attributionsstrategien?
3. In welchem Zusammenhang stehen grammatische Kernkompetenzen und nominaler Ausbau? Ist nominaler Ausbau von Vorläuferkompetenzen im Bereich des verbalen Ausbaus abhängig oder werden beide sprachlichen Bereiche parallel ausgebaut?

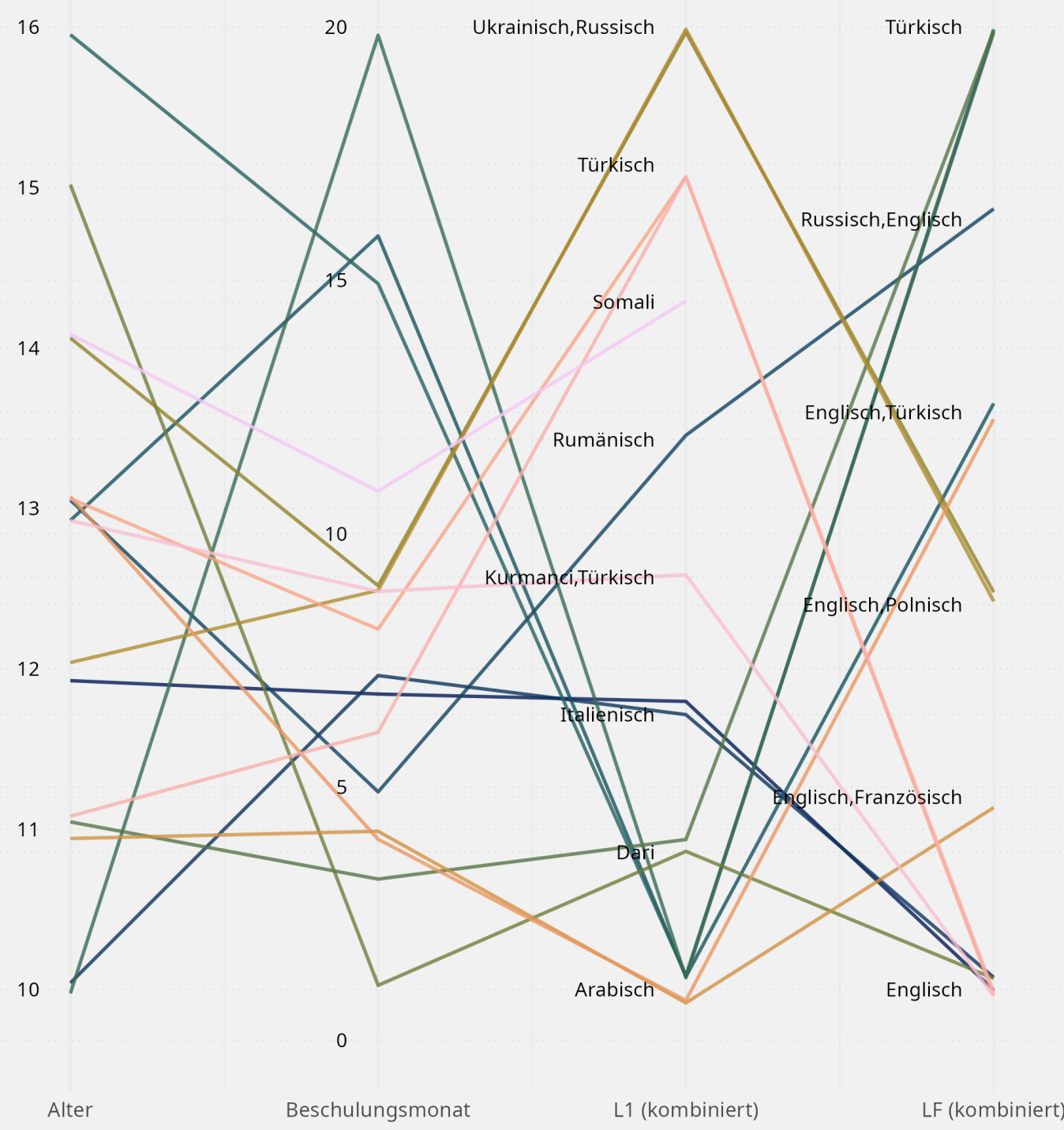


Abb. 1: Heterogenität der Metadaten in SeiKo

ERHEBUNGSDESIGN

- longitudinales Design mit monatlichen Datenerhebungen über bis zu zwei Jahre (initiale Beschulung in Deutschland)
- Elizitationsstimuli = insgesamt sechs monatlich rotierende Bildergeschichten (s. Bsp. Abb. 2)
- drei situativ und registerspezifisch unterschiedliche Aufgaben in zwei Modi:
 1. mündliches Nacherzählen der Bildergeschichte für die anwesende Interviewer:in (N)
 2. anschließende Fragen zur Verständnissicherung (Q)
 3. gemeinsames Verschriftlichen des Geschehens für eine nicht anwesende Person (z.B. Lehrer:in) mit der Interviewer:in (Distanzsprachlichkeit) (D)
 4. eigenständiges, schriftliches Berichten (T)
- ergänzend: freies Interview über den Schulalltag, Ferien-/Freizeitaktivitäten

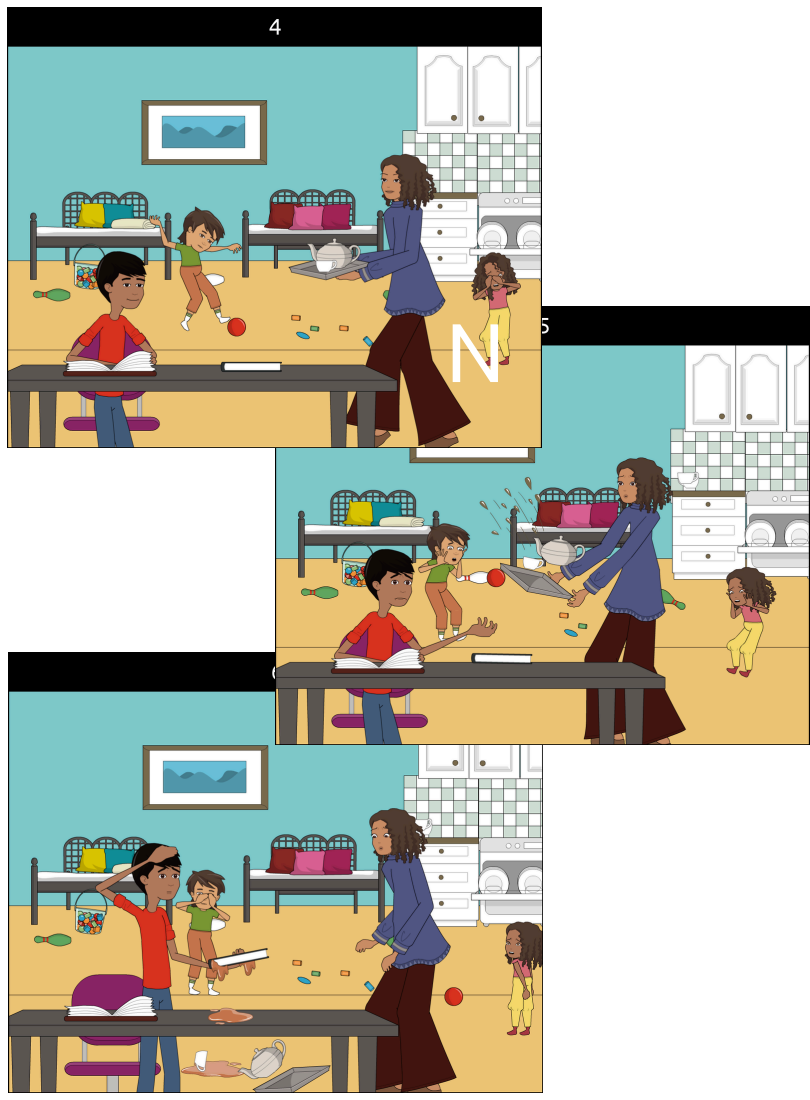


Abb. 2: Beispiel für die Bildstimuli

DATENAUFBEREITUNG

Die Rohdaten werden manuell im EXMARaLDA Partitur-Editor (Schmidt/Wörner 2014) transkribiert, segmentiert und annotiert. Die XML-basierten Dateien werden mithilfe einer Annatto-Pipeline (Krause/Klotz) in ANNIS (Krause 2019) durchsuchbar gemacht.

		72 [01:2]	73 [01:2]	74 [01:2]	75 [01:2]	76 [01:2]	77 [01:2]	78 [01:2]	79 [01:2]	80 [01:2]	81 [01:2]	82 [01:2]	83 [01:2]
Tokenisierte Spur nach manueller Transkription	B09[tok_part]	hier	der	junge	glaub	ich	ha	gesagt	wer	war	das	oder	so
Segmentierung in Äußerungseinheiten, angelehnt an die Analysis of Speech Unit (Foster et al. 2000); Markierung ausgeschlossenes Material (z.B. Korrekturen)	B09[unit]	DEC											
	B09[unit2]				DEC								
	B09[subclause]								MCSUBWINT				
	B09[excl]												
mit [excl] bereinigte tok-Spur	PART_CLEAN[ctok_part]	hier	der	junge	glaub	ich	ha	gesagt	wer	war	das	oder	so
automatisiert generierte Lemmatisierung und POS-Tagging (Treetagger, Schmidt 1995)	PART_CLEAN[auto_lemma]	hier	die	jung	glauben	ich	ha	sagen	wer	sein	die	oder	so
	PART_CLEAN[auto_pos]	ADV	ART	ADJA	NN	PPER	NGHR	VVPP	PWS	VAFIN	PDS	KON	ADV
manuell korrigierte POS- und Lemmaspur	PART_CLEAN[lemma]	hier	die	Junge	glauben	ich	haben	sagen	wer	sein	die	oder	so
	PART_CLEAN[pos]	ADV	ART	NN	VVFIN	PPER	VAFIN	VVPP	PWS	VKFIN	PDS	KON	ADV
manuelle Annotation der Verbstellung (angepasstes DAKODA-Schema, Ruppenhofer et al. 2024)	PART_CLEAN[unit_cons]	X	S		Xcl		V	V	Ocl				
	PART_CLEAN[unit_stage]	SEPanti	ADV										
	PART_CLEAN[unit2_cons]				V	S							
	PART_CLEAN[unit2_stage]				INVin								
	PART_CLEAN[subclause_cons]								S	V	Pred	X	
manuelle Annotation von Nominalgruppen (vgl. Braunewell et al. ang.)	PART_CLEAN[subclause_stage]								INVpseudo				
	PART_CLEAN[ngt]			NGrO									
	PART_CLEAN[kern]		DET	KERN									

Abb. 3: Ansicht im EXMARaLDA Partitur-Editor von MIA, Beschulungsmonat 37

Herausforderungen:
frühe Lerner:innenvarietäten;
Unverständliches, Pausen,
bruchstückhafte Äußerungen;
hohe Variabilität (zwischen Proband:innen und über Erhebungszeitraum hinweg);
verschiedene Modi

>> manuelle Transkription und Korrektur der automatisch erzeugten POS- und Lemmaspur
>> mehrschrittige, feingliedrige Abstimmung und Evaluation des Annotationsvorgehens

KORPUSUMFANG

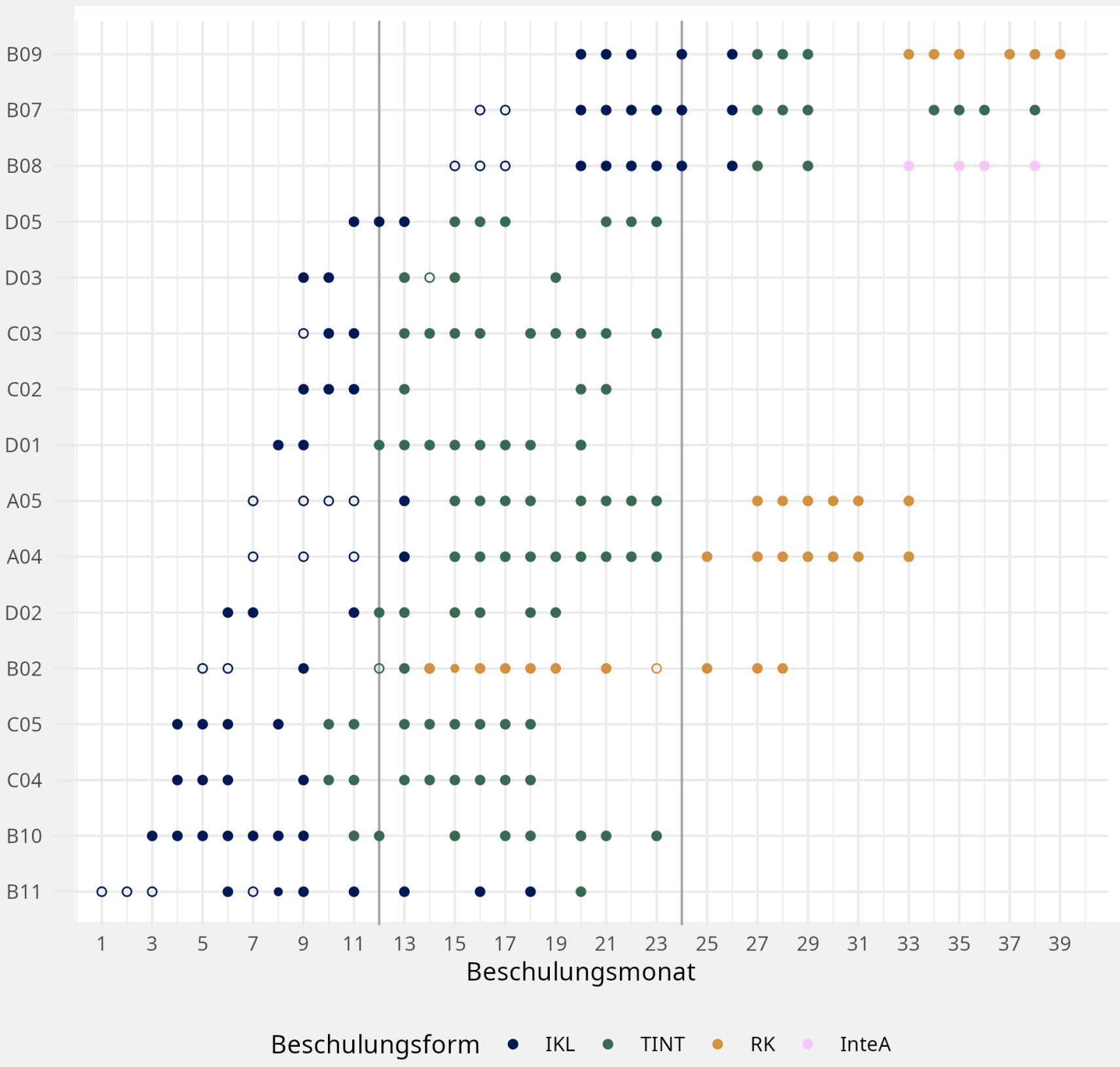


Abb. 4: Datenerhebungszeitpunkte je Schüler:in und Beschulungskontext

SeiKo umfasst insgesamt 778 mündliche und schriftliche Texte von 16 Schüler:innen

Gesamttokenzahl von SeiKo 1 (n=8): 68118

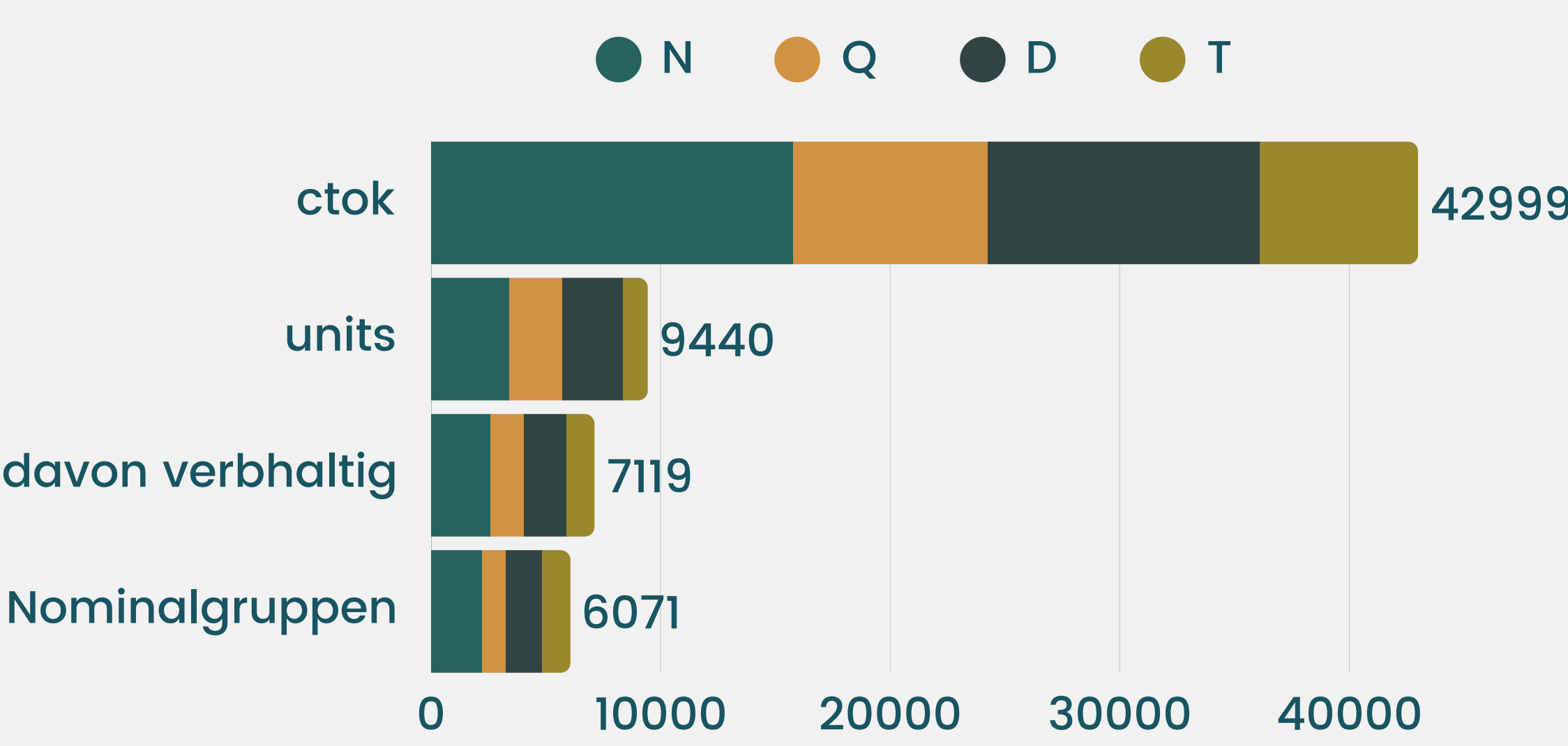


Abb. 5: Datenumfang je Erhebungspart

JUSTUS-LIEBIG-
UNIVERSITÄT
GIESSEN

Gefördert durch
Deutsche
Forschungsgemeinschaft

KORPUSWEBSEITE,
REFERENZEN, KONTAKT
& UPDATES

