ES war Emmad einHundung eine Katze Sie haden Sich die gunze zeit geschle aber nemand wa auf den Spilplaze und dan haden sie haben enschlosen das sie Wider freunde Sin.

NaLeKo – ein Lernerkorpus mit Erzähltexten von Kindern und Jugendlichen



Hagen Hirschmann · HU Berlin hagen.hirschmann@hu-berlin.de

Anja Binanzer · TU Dresden anja.binanzer@tu-dresden.de

Miriam Langlotz · U Kassel m.langlotz@uni-kassel.de

NaLeKo – KEY FACTS

NArratives LErnerKOrpus

- tief annotiertes Korpus zur Untersuchung narrativer Texte in deutscher Sprache von Schüler:innen mit unterschiedlichen Spracherwerbskonstellationen:
 - Erstsprache Deutsch (L1 und 2L1)
 - früher Zweitspracherwerb Deutsch (fL2)
 - später Zweitspracherwerb Deutsch (sL2)
- standardisierte Datenerhebungen 2020–heute
- je Schüler:in zwei narrative Texte, eine Sprachbiographie, ein C-Test Deutsch (allgemeiner Sprachstand)
- Durchführung der Datenerhebungen mit studentischen Mitarbeiter:innen und Studierenden (Anleitung im Rahmen von forschungsorientierten Projektseminaren)

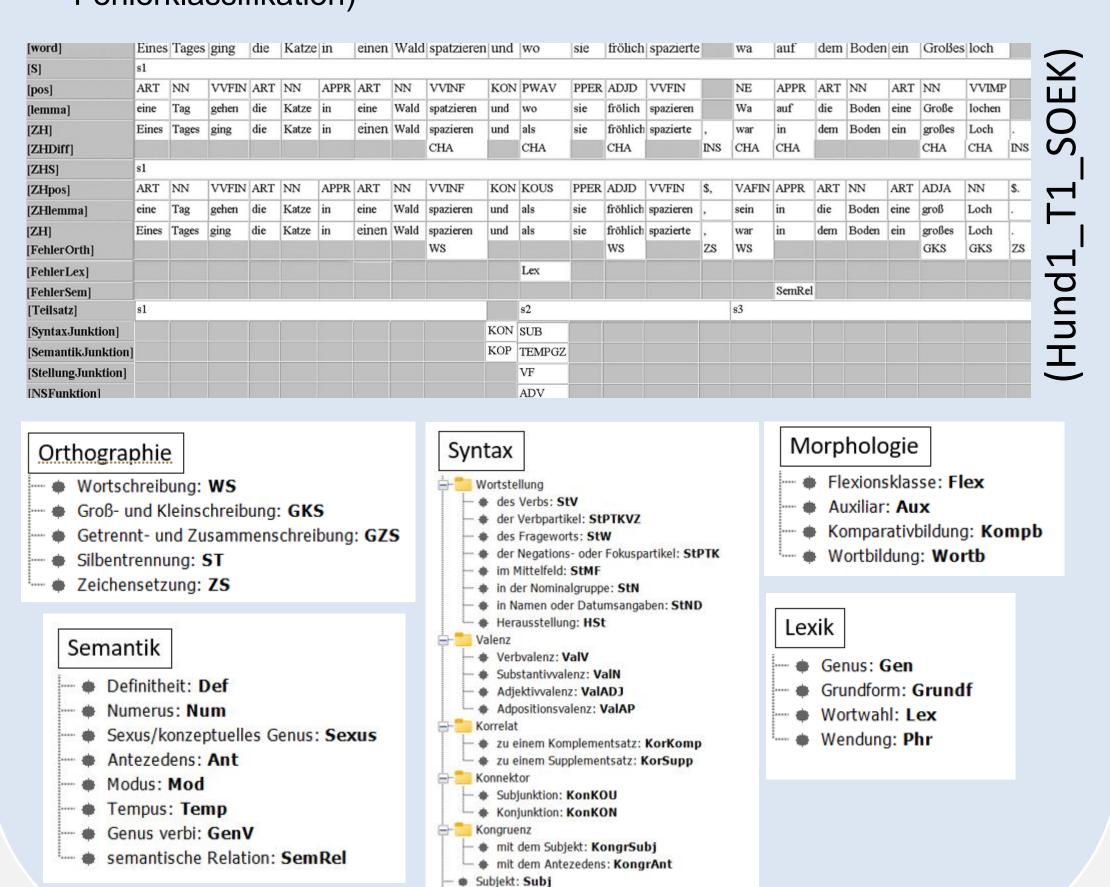
FORSCHUNGSFELDER

- (Entwicklung) narrative(r) Fähigkeiten in der Schriftlichkeit
- Fokus: grammatische Entwicklungen/Lernersprache
- Erwerb / funktionaler Gebrauch von für Erzählungen typischer sprachlicher Mittel
- Vergleiche:
 - Erst- vs. Zweitspracherwerb
 - früher Zweitspracherwerb vs. später Zweitspracherwerb

KORPUSAUFBAU und ANNOTATION

Verarbeitungsschritte:

- Transkription (von handschriftlichen Dokumenten zu .docx-Dateien zu EXMARaLDA)
- Metadatentabellen (.txt-Dateien, später Details zu Variablen)
- Weiterverarbeitung in EXMARaLDA Nutzung der Tranformationspipeline EXMARaLDA (Dulko) (Nolda 2023, Hirschmann & Nolda 2019)
- Tokenisierung
- Formulierung einer Zielhypothese (s.u.) für lernerbedingte Abweichungen
- pos-Tagging und Lemmatisierung
- Satzspannen
- Beschreibung der Abweichungen durch Editierdistanzen (Edit-Tags) sowie einer
- mehrdimensionalen Fehlerklassifikation (erarbeitet im Dulko-Projekt)
- Zusätzlich zu Dulko-Transformationen: Analyse der Verwendung von Junktion (vier Annotationsebenen)
- Teilsätze • Analyse der Verb- und Tempusverwendung (drei Annotationsebenen)
- Zielhypothese und Abweichungskategorien
- Zielhypothesen werden auf einer Beschreibungsebene erstellt (Vervielfältigung der Ebene möglich, falls verschiedene Abweichungsexponenten existieren: Sie wartet fur [ZH: für – ZH:auf] ihn.)
- Abweichungen auf fünf Ebenen explizit kenntlich gemacht (Dulko-Fehlerklassifikation)

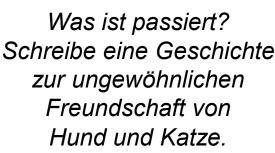


DATENTYPEN

Schreib-/Bildimpulse

Deine Klasse verfasst gemeinsam ein Geschichtenbuch, für das jeder von euch eine Geschichte schreiben darf. Erfinde für das Geschichtenbuch eine lustige / spannende / gruselige Geschichte zu einem Ereignis, das in deiner Schule vorgefallen sein könnte. Langlotz (2014)







von Vater, Mutter

und Baby.

Binanzer (2017)

C-Test

Handys						
Immer mehr Kinder und Jugendliche besitzen heute ein						
eigenes Handy	y. Immer wicht_	wird e	für			
s, da	as rich	_ Handy m	den			
aktu	Klingeltönen u	den cool_				
Logos z	haben. Da_	kann m				
leicht verg	, dass e	Handy en_	viel			
Ge1	kostet. Vi	Jugendliche ge_				
sehr vi	Geld f	ihr Ha	_ aus. Bevor			
man sich also ein Handy anschafft, sollte man sich vorher						
genau über alles informieren.						

Linnemann & Wilbert (2010)

- · allgemeiner Sprachstand im Deutschen • differenzierte C-Tests nach Primarstufe ((Grießhaber o.J.,) und -/Sekundarstufe Linnemann & Wilbert 2010)
- RF-Wert

FraBo Sprachbiographie

Welche Sprachen sprich	•	achen			\
Welche Sprache hast du	in deiner Familie bis zu	deinem 3. Gebu	rtstag gelern	t und gespr	och
				_	
In welchen Sprachen spr	ichst du mit			_	
				_	
deinem Vater _				_	

NaLeKo basierend auf Grießhaber (o.J.)

- Metadaten zu Alter, Geschlecht
- Erst-, Zweit-, Fremdsprache • Sprachgebrauch Familie und Freunde
- Sprachgebrauch Medien

VERÖFFENTLICHTE DATEN (NaLeKo Version 2.0)

Klassenstufe	Texte	Token	Wörter	Teilsätze nach ZH	Junktionen	Abw./Token
2	9	366	315	52	33	141
3	10	887	750	130	59	234
4	37	6693	5337	1077	532	2769
5	26	7115	5752	1108	460	838
7	39	15513	12635	2373	983	997
9	43	14279	12059	2099	975	1154
Σ	164	44853	36848	6839	3042	6133
Innerhalb von l	_1 Deutsch: 2l	_1				
Klassenstufe	Texte	Token	Wörter	Teilsätze nach ZH	Junktionen	Abw./Token
4	5	760	641	113	68	264
9	12	2449	2108	364	209	431
Σ	17	3209	2749	477	277	695
L2 Deutsch (L1	RU, TRK, IT u	ı.a.)				
Klassenstufe	Texte	Token	Wörter	Teilsätze nach ZH	Junktionen	Abw./Token
2	22	879	713	128	64	401
3	36	2401	1978	368	188	892
4	47	6620	5420	1018	538	1937
9	18	3499	2945	499	255	840
Σ	123	13399	11056	2013	1045	4070

- Version NaLeKo 2.0 steht in ANNIS zur Verfügung
- > https://huberlin.de/naleko/



NEUE ERHEBUNGEN 2023–2025

	Sekundarstufe
Textsorte	Hund und Katze Schlaflose Nacht
N S*S	19
N Texte	38
L	sL2 Deutsch (Seiteneinsteiger:innen)
Jahrgang	5, 6, 7, 8, 9, 10, 11

	Primarsture	Sekundarstute
Textsorte	Hund und Katze Schlaflose Nacht	Hund und Katze Schlaflose Nacht
N S*S	20 + 17	24
N Texte	38 + 34	48
L	L1 Deutsch fL2 Deutsch	L1 Deutsch 2L1 Deutsch
Jahrgang	3, 4	7

geplante Erweiterungen

- Seiteneinsteiger:innen mit L1 Russisch
- und/oder Ukrainisch
- fL2-Erwerb: Primar- und Sekundarstufe mit L1 Russisch und Türkisch

METHODISCHE HERAUSFORDERUNGEN

- Händische Annotation lernersprachlicher Variation siehe auch Beispiel links):
- Zeitfaktor, Akkuratheit, Vergleichbarkeit von Fällen Umgang mit mehrdeutigen Fällen
- Frage der Vergleichbarkeit der Texte:
- Narrative Texte gemäß Aufgabe, aber auch nichtnarrative Anteile
- Sprachliche vs. inhaltlich-strukturelle Qualität der Texte (siehe Beispiel rechts)

Es waren einmal verheirate und sie hatten ein Baby. Es war einmal 1. April. An dieser Nacht hat das Baby ein streich gespielt. Das Baby hat die ganze Nachtlang geweint. Die Mutter hat versucht das der Baby ein Bär gegeben. Der Vater hatte das gleiche versucht nur mit einem Ball, aber es hatte nicht geklappt. Das Baby sagte: "April, April." Die Eltern sagten: "Es hat sein erstes Wort gesagt." Jeder konnte jetzt schlafen. (Nacht_T1 VAN)

FORSCHUNGSFRAGEN

Bearbeitet:

- Junktionsentwicklung in Erst- und Zweitsprache Deutsch (Binanzer/Langlotz 2019)
 - im Vergleich von Primar- und Sekundarstufe
 - Berücksichtigung Besonderheiten spezifisch narrativer Themenentfaltung (Zusammenhang von Lexik, Syntax und Semantik)
- Entwicklung temporaler Junktoren im Vergleich zum Gebrauch in Kinder- und Jugendliteratur: (Binanzer/Hirschmann/Langlotz 2022))
 - Welche Ausdrücke werden in der Kinder- und Jugendliteratur zum Herstellen temporaler Relationen verwendet?
 - Inwiefern spiegelt sich diese Verwendungsweise in Texten von Kindern/Jugendlichen wider?

Geplant:

- Entwicklung starke Verbflexion in der Zweitsprache Deutsch
- In welcher Reihenfolge vollzieht sich der Erwerb der starken Flexion in den relevanten Paradigmenpositionen bei L2-
- Kindern? Unterschiede im Iernersprachliche Variation bei Kindern mit fL2 Deutsch und L2 Deutsch

- **Studentische Abschlussarbeiten** • Gumprich, Ricarda (2023): Kohäsion in Schülertexten - eine empirische Untersuchung narrativer Texte. Unveröffentlichte Examensarbeit Universität Kassel
- Baum, Lara (2024): Kasuserwerb von ein- und mehrsprachigen Grundschüler:innen Eine Untersuchung narrativer Texte. Unveröffentlichte Examensarbeit, Universität Kassel. • Schröder, Ruth (2024): Zum Junktorengebrauch in narrativen Texten in der Zweitsprache Deutsch - Eine empirische Untersuchung mit neuzugewanderten ukrainischen
- Schülerinnen und Schülern. Unveröffentlichte Masterarbeit, Leibniz Universität Hannover. • Wiedermann, Leona (2025): Temporale Junktoren im Zweitspracherwerb: Eine korpusbasierte Analyse narrativer Kompetenzen russischer und ukrainischer SeiteneinsteigerInnen
- Unveröffentlichte Bachelorarbeit, Leibniz Universität Hannover, An NaLeKo haben zudem mitgewirkt: