

Every Verb in its Right Place? A Roadmap for Operationalizing Developmental Stages in the Acquisition of L2 German

Josef
Ruppenhofer
FernUniversität in
Hagen

Matthias
Schwendemann
Universität Leipzig

Annette
Portmann
Universität Leipzig

Katrin
Wisniewski
Universität Leipzig

Torsten Zesch
FernUniversität in
Hagen

2024-04-22

Table of contents

- Introduction
- Challenges for specification
- Challenges for implementation
- Conclusion
- Appendix



DAKODA

Every verb in its right place?

Introduction

Developmental stages

- Observation: both large inter-individual variance and generally ordered, step-like progress in second language learning
- For German, one central phenomenon: acquisition of verb placement options in finite German clauses

Caution

The German grammar is blistered all over with separable verbs; and the wider the two portions of one of them are spread apart, the better the author of the crime is pleased with his performance. (Mark Twain, "The awful German language")

Verb placement types in L2 acquisition

Constellation Example

SVO

[Ich S] **suche** [eine neue Wohnung O]

lit. I am-looking-for a new apartment

ADV¹

[Darum X] [ich S] **suche** [eine neue Wohnung O]

lit. Therefore, I am-looking-for a new apartment

SEP

[Ich S] **muss** [eine neue Wohnung O] **suchen**

lit. I must a new apartment look-for

INV

[Darum X] **suche** [ich S] [eine neue Wohnung O]

lit. Therefore look-for I a new apartment

VEND

Weil [ich S] [eine neue Wohnung O] **suche**

because I a new apartment look-for

Processability Theory (PT)

- Well-known SLA-theory with significant amount of research on German verb placement
- Focus on explanation of acquisitional order from simpler to more complex structures
- Hypothesis: observed development tracks expanding capabilities of the human language processor (Levelt 1989)
- Theoretical ties to Lexical functional grammar (Kaplan and Bresnan (1982); Dalrymple, Lowe, and Mycock (2019))
- No claim that all language development is driven by the factor processability.

Why developmental stages matter

- Within SLA, assumption of stages not uncontroversial (Larsen-Freeman 2020)
- Curriculum development (Winkler (2014); Schroeder and Gamper (2016); Baten, Kristof and Keßler, Jörg (2019))
- Research on proficiency and language testing (Grießhaber (2012); Heilmann and Grieshaber (2012); Grießhaber (2019); Schulz and Tracy (2011); Schulz and Grimm (2012))
 - real world L2 assessments
- Unclear relation to CEFR-levels (Common European Framework of Reference for Languages) (Wisniewski 2020)

- So far, developmental stages studied on smaller datasets
- To enable large scale automated analysis across multiple datasets, we require
 - a fleshed-out linguistic specification of PT's stages

<https://github.com/dakoda-project/devstages> ✓

- a computational implementation

recommendations, ongoing work ⏺

Challenges for specification

Overview of challenges

- Lack of linguistic precision
- Incomplete empirical coverage
- Analyzing learner language
- Chunks and borrowing of structures
- Emergence criterion

Incomplete empirical coverage

- ADV presented as a non-L1 structure used by learners
 - *Darum ich **suche** eine neue Wohnung
Therefore I am-looking-for a new apartment.
- However, L1 German does feature cases where two constituents precede the finite verb in second position (Müller 2003)

Vermutlich ein technischer Defekt	hatte	dazu	geführt	dass ...
Presumably a technical defect	had	thereto	led	that ...

- Advanced learners producing such tokens are not regressing!

Incomplete empirical coverage

- Some structures perfectly acceptable in spoken German but not in written German

Diese Reaktion finde ich gut

This reaction I find good

a *weil das leider Realität ist* spoken, written

because that is unfortunately reality

b *weil das ist leider Realität* spoken

for that is unfortunately reality

- Variant b is no negative evidence that speaker has not in fact mastered V-END.

Challenges for implementation

- Preprocessing
- Topological + dependency parsing
- Dependency on learner proficiency
- Use of target hypotheses
- Lack of grammatical information

Preprocessing

- Imperfect fit between available tools (L1, written) and our data (L2, incl. spoken)
- Key problem: to look at syntax, we need sentence segmentation
- In spoken language research, sentence concept typically not used => no punctuation

So ehm wir gehen vom Start Start befindet sich in der linken unteren Ecke
so uhm we go from start start is-located in the left lower

(source:
HaMaTaC
corpus)

- Research on spoken L2 data does not use established transcription standards
- In written data, learners' use of punctuation not necessarily L1-like

Ich denke , das Schreit nicht Schön ist , unt wen wen man weis , das einem etwas runter felt , und du
das einfach nimst und aPhaust , nent man das Di@bstal . (source: MultiLit corpus)

→ use or adaptation of segmentation approaches for L1 spoken German necessary

HaMaTac example: two ways to segment

Example contribution: So ehm wir gehen vom Start Start befindet sich in der linken unteren Ecke

Approach of Guhr et al. (2021), plus true-casing Lita et al. (2003), followed by segmentation

punctuation-ful

Speaker	S/SLU
Nat	So ehm, wir gehen vom Start 'So uhm, we go from the start'
	Start befindet sich in der linken unteren Ecke . ('Start is in the lower left corner')

Approach of Rehbein, Ruppenhofer, and Schmidt (2020)
punctuation-less

Speaker	S/SLU
Nat	So
	So
	Ehm
	Uhm
	Wir gehen vom Start
	'We go from the start'
	Start befindet sich in der linken unteren Ecke
	('Start is in the lower left corner')

Segmenting HaMaTaC

	punctless		punctful	
Sentences	5825		4378	
	spacy	syntaxdot	spacy	syntaxdot
Verbs	2274	2295	2276	2291
Verbs, finite	1853	1889	1850	1886

Sentence and verb counts by segmentation approach
for two taggers/parsers

- Sentence counts significantly different
- But number of verbs very close
- Impact on parsing and downstream analyses 
- Results consistent across corpora 



DAKODA

Conclusion

Every verb in its right place?

Wrap-up

- Studying evidence for developmental stages is of theoretical and practical relevance
- To do so on a larger scale, need for automated analyses across many learner corpora
- Our contributions
 - linguistic specification
 - recommendations on implementation

Appendix

Specifications

Lack of linguistic precision

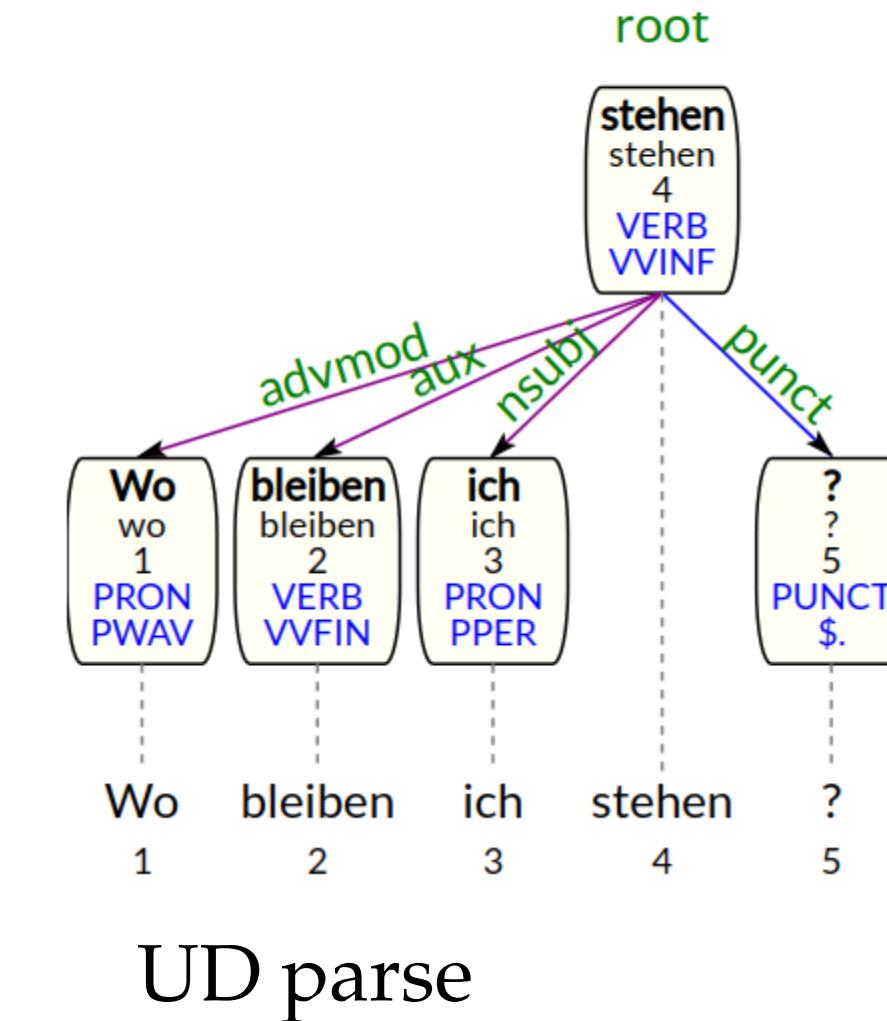
What object types are encompassed by O in PT's structural descriptions?

Potential object types (illustrated in SVO)

Type	Example
Accusative	Hanna isst [Äpfel O] 'Hanna eats apples'
Dative	Peter hilft [mir O] 'Peter is helping me'
Genitive	Eine Reihe von Prominenten gedachte [des Verstorbenen O] 'A series of celebrities remembered the deceased'
Prepositional objects	Wir verlassen uns [auf Zahlen] 'We rely on numbers'

Analyzing learner language

- Learner language often not L1-like
- However, difficult to categorize on its own terms
- Many deviations from L1 amenable to multiple analyses
 - **Wo bleiben ich stehen**
 - agreement error or finiteness error?
 - depending on answer, we have a finite clause for analysis (SEP) or not



Chunks and borrowing of structures

- Difficult distinction between productively formed output and pre-fabricated
 - E.g. multi-word *wissen lassen* ‘let know, inform’ very frequently in SEP
- Similarly problematic: Re-use of material encountered in task prompts, instructions
 - how to identify it
 - really different from e.g. priming?

Emergence criterion

- A stage counts as acquired if some N instances are produced in contexts where the relevant verb constellation is expected by L1 standards
- Typically, a lexical diversity criterion for verbs is employed: verb constellation needs to be found with M different verb lemmas
- Choice of values for N and M may significantly impact results given small size of many text samples

Implementation

Every verb in its right place?

Topological parsing

We cast the analysis task as a combination of topological field and dependency parsing.

Topological analysis: VVF=Vorvorfeld (pre-prefield); VF=Vorfeld(prefield, fore-field); LSK=left sentence bracket; MF=midfield; RSK=right sentence bracket; NF=post-field

	(VVF)	VF	LSK	MF	RSK	NF
a		Ich	suche	eine neue Wohnung		
b	Aber	ich	suche	eine neue Wohnung		
c		Darum ich	suche	eine neue Wohnung		
d		Ich	muss	eine neue Wohnung	suchen	
e	Und	ich	habe	eine neue Wohnung	gefunden	
f		Darum	muss	ich eine neue Wohnung	suchen	
g	Und	darum	muss	ich eine neue Wohnung	suchen	
h			weil	ich eine neue Wohnung	suche	
			Every verb in its right place?			



i
DAKODA

(VVF)

VF

LSK

MF

RSK

NF

weil

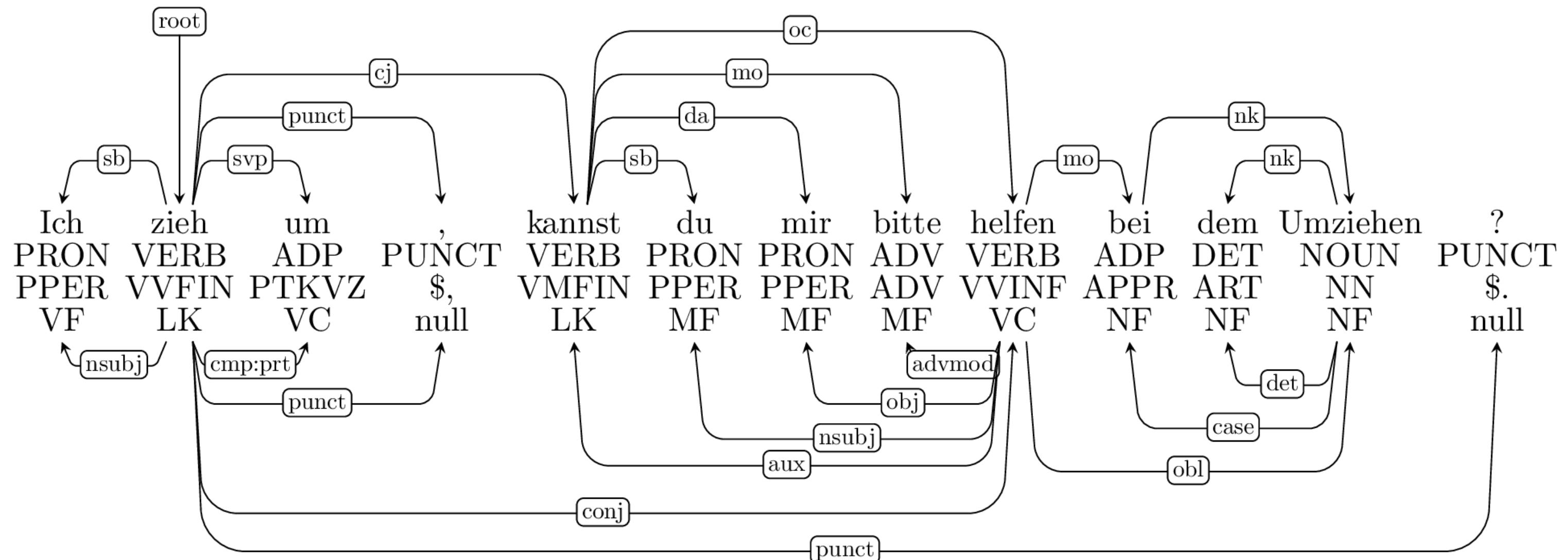
ich eine neue
Wohnung

Suche

eine neue
Wohnung

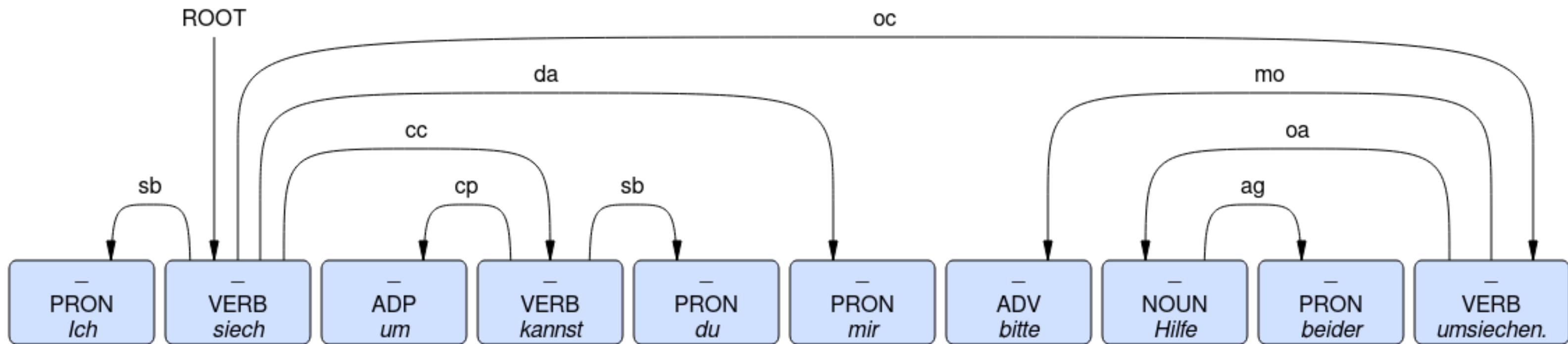
j

Combination of topological and dependency parsing



Dependency on learner proficiency

- Early learners most relevant due to less target language-like state of their grammar
- At the same time, productions harder to analyze for L1-trained tools



Use of target hypotheses

- Explications of the assumed minimally different grammatical structures the learners aim for
- Allow for traceable error annotations (Lüdeling, 2008; Reznicek et al., 2013)
- Idea: parse both target hypothesis and raw productions and project knowledge from former back onto latter
- Problem: THs not available in many corpora

Lack of grammatical information

- We have no access to valence in our parsing approach
- certain spoken language structures that omit an argument appear incorrect

Kann ich lesen .

Can I read

I can read that. (spoken, topic-drop)

*I can read / I know how to read. (written,
habitual)

=> manual post-inspection of such cases

References

- Baten, Kristof and Käßler, Jörg. 2019. "Research timeline : the role of instruction : teachability and processability." In *Teachability and learnability across languages*, edited by Arntzen, Ragnar and Håkansson, Gisela and Hjelde, Arnstein and Käßler, jörg, 6:9–26. Processability Approaches to Language Acquisition Research & Teaching. John Benjamins. {<http://doi.org/10.1075/palart.6.01bat>} ({<http://doi.org/10.1075/palart.6.01bat>}).
- Dalrymple, Mary, J. J. Lowe, and L. Mycock. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford University Press.
- Grieshaber, Wilhelm. 2012. "Die Profilanalyse." In *Einblicke in Die Zweitspracherwerbsforschung Und Ihre Methodischen Verfahren.*, edited by Bernt Ahrenholz, 173–94. DaZ-Forschung. Berlin u.a.: De Gruyter.
- . 2019. "Profilanalysen." In *Sprachdiagnostik Deutsch Als Zweitsprache*, edited by Stefan Jeuk and Julia Settinieri. Berlin, Boston: De Gruyter Mouton.
- Guhr, Oliver, Anne-Kathrin Schumann, Frank Bahrmann, and Hans-Joachim Böhme. 2021. "FullStop: Multilingual Deep Models for Punctuation Prediction." In *Swiss Text Analytics Conference*. <https://api.semanticscholar.org/CorpusID:238232903> (<https://api.semanticscholar.org/CorpusID:238232903>) .
- Heilmann, Beatrix, and Wilhelm Grieshaber. 2012. *Diagnostik & Förderung - Leicht Gemacht : [Das Praxishandbuch]*. Deutsch als Zweitsprache in der Grundschule. Klett Sprachen. <https://books.google.de/books?id=R5kvWNhw3PUC> (<https://books.google.de/books?id=R5kvWNhw3PUC>) .
- Kaplan, Ronald, and Joan Bresnan. 1982. "Lexical-Functional Grammar: A Formal System for Grammatical Representation." In *The Mental Representation of Grammatical Relations*, edited by Joan Bresnan, 173–281. MIT Press.
- Larsen-Freeman, Diane. 2020. "Complex Dynamic Systems Theory." In *Theories in Second Language Acquisition*, edited by Bill VanPatten, Gregory D. Keating, and Stefanie Wulff, 248–70. Taylor & Francis.
- Levelt, Willem J. M. 1989. *Speaking: From Intention to Articulation*. The MIT Press. <https://doi.org/10.7551/mitpress/6393.001.0001> (<https://doi.org/10.7551/mitpress/6393.001.0001>) .
- Lita, Lucian Vlad, Abe Ittycheriah, Salim Roukos, and Nanda Kambhatla. 2003. "TRuEcasIng." In *Proceedings of the 41st Annual Meeting on Discourse and Dialogue*. Association for Computational Linguistics. 1–6. Every verb in its right place?



Meeting of the Association for Computational Linguistics, 152–59. Sapporo, Japan: Association for Computational Linguistics.
<https://doi.org/10.3115/1075096.1075116> (<https://doi.org/10.3115/1075096.1075116>) .

- DAKODA Rehbein, Ines, Josef Ruppenhofer, and Thomas Schmidt. 2020. "Improving Sentence Boundary Detection for Spoken Language Transcripts." In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 7102–11. Marseille, France: European Language Resources Association. <https://aclanthology.org/2020.lrec-1.878> (<https://aclanthology.org/2020.lrec-1.878>) .
- Schroeder, Christoph, and Jana Gamper. 2016. "Sprachliche Bildung für Neuzugewanderte. Ein Plädoyer für einen erwerbssequentiellen Ansatz [Language instruction for new immigrants: A plea for an approach based on order of acquisition]." *Osnabrücker Beiträge Zur Sprachtheorie* 89: 217–30.
- Schulz, Petra, and Angela Grimm. 2012. "Spracherwerb." In *Germanistik. Sprachwissenschaft – Literaturwissenschaft – Schlüsselkompetenzen*, edited by Heinz Drügh, Susanne Komfort-Hein, Andreas Kraß, Cécile Meier, Gabriele Rohowski, Robert Seidel, and Helmut Weiß, 155–72. Stuttgart: Metzler.
- Schulz, Petra, and Rosemarie Tracy. 2011. *Linguistische Sprachstandserhebung - Deutsch als Zweitsprache (LiSe-DaZ): Language Test for Children with German as a Second Language*. Göttingen: Hogrefe. <https://www.biss-sprachbildung.de/btools/linguistische-sprachstandserhebung-deutsch-als-zweitsprache-lise-daz/> (<https://www.biss-sprachbildung.de/btools/linguistische-sprachstandserhebung-deutsch-als-zweitsprache-lise-daz/>) .
- Winkler, Steffi. 2014. "Für eine psycholinguistisch orientierte Fremdsprachendidaktik. Anmerkungen auf Grundlage einer exemplarischen Studie zum Negationserwerb im Deutschen [In favor of a pyscholinguistically oriented foreign language didactics. Remarks based on a sample study on the acquisition of negation in German]." In *Zur Sprache.km*, edited by U. Bredel, I. Ezhova-Heer, and S. Schlickau, 51. Göttingen: Universitätsverlag.
- Wisniewski, Katrin. 2020. "SLA developmental stages in the CEFR-related learner corpus MERLIN: Inversion and verb-end structures in German A2 and B1 learner texts." Journal Article. *International Journal of Learner Corpus Research* 6 (1): 1–37. <https://doi.org/https://doi.org/10.1075/ijlcr.18008.wis> (<https://doi.org/10.1075/ijlcr.18008.wis>) .