

# Fehlerannotation: halbautomatisch statt manuell?



**Zum Umgang mit schriftlichen Produktionen auf Anfängerniveau  
am Beispiel des Schweizer Lernerkorpus SWIKO**

Leipzig (D)

**Nina Selina HICKS**  
**Thomas STUDER**

*Tagung Grosse Lernerkorpora (?) – Möglichkeiten und Grenzen*  
*25. September 2025*

# Inhalt

- Kontext & Aufbau
- Fehlerannotation
- Ausgewählte Ergebnisse
- Ausblick

# SWIKO – Kontext

- Wandel im schulischen Fremdsprachenunterricht  
(Council of Europe 2001, 2020; EDK 2011, 2017; Bertschy et al. 2015; Lenz & Wiedenkiller 2019; Peyer et al. 2016)
  - Kompetenzbasiert, handlungs- & inhaltsorientiert
  - HarmoS: 2 Fremdsprachen ab Primar (~ 8 & 10 Jahre), Nationalsprache + Englisch, A2.2 (A2.1 Schreiben)
- Fragestellung
  - Wie präsentieren sich im Zuge dieses «neuen» Fremdsprachenunterrichts Wortschatz und Grammatik in den Produktionen von Lernenden am Ende der obligatorischen Schulzeit?

# SWIKO – Kontext

- Ziele
  - Lernalterssprache am Ende der obligatorischen Schulzeit korpuslinguistisch dokumentieren & analysieren
  - Korpusnutzungen für das Sprachenlernen explorieren
- Komplementäre Anlage
  - Zum Bildungsmonitoring
  - Zur internationalen Lernerkorpus-Landschaft

# SWIKO – Workflow & Komponenten



Tasks

# Aufgabenvariation

cf. *tasks* in Ellis et al. 2020

Eigenschaft	Variation	SWI01	SWI02	SWI03	SWI04	SWI05	SWI06	SWI07	SWI08
Texttyp	deskriptiv	x	x			x	x		
	argumentativ			x	x			x	x
Thema	persönlich	x		x		x		x	
	schulisch		x		x		x		x
Strukturiertheit	mehr	x	x	x	x				
	weniger					x	x	x	x

## Produktionsbedingungen:

- 3 Sprachen (Deutsch, Englisch, Französisch; als Schul- & Fremdsprache)
- 2 Medien: auf Papier und Computer
- 2 Modalitäten: mündlich und schriftlich

# SWIKO – Aufgabe SWI04

## SWI04\_df

Du hast im Internet eine Liste mit den 100 wichtigsten Erfindungen gefunden.

**Schau die Liste an.**

**Schreibe einen Text mit deiner Meinung.**

**Warum sind diese Erfindungen auf der Liste?**

**Bist du mit der Reihenfolge der Erfindungen einverstanden? Warum (nicht)?**

Schreibe zu mindestens drei Punkten.

**Welche anderen Erfindungen sollten auf der Liste sein? Wo? Warum?**

Schlage mindestens zwei neue Punkte vor.

Écris en  
français!

Schreibe auf  
Französisch!



Schreibe etwa 60-80 Wörter.

- Schulsprache (LoS): Deutsch
- Zielsprache (FL): Französisch
- Modalität: schriftlich
- Medium: Papier
- Texttyp: argumentativ
- Thema: schulisch
- Struktur: mehr

## 100 Erfindungen, die die Welt verändert haben

1. Computer
2. Elektrizität
3. Flugzeug
4. Internet
5. Telefon

...

96. Brille
97. Armbanduhr
98. Bus
99. Schreibtisch
100. Löffel

# SWIKO – Aufgabe SWI05

Ta classe va faire un échange scolaire avec une autre classe.  
En préparation, votre classe a créé un audioblog où chacun doit se présenter.

**Enregistre un texte pour te présenter à d'autres élèves.**  
Parle au moins 100 secondes.

Sprich  
Deutsch!

Parle en  
allemand !



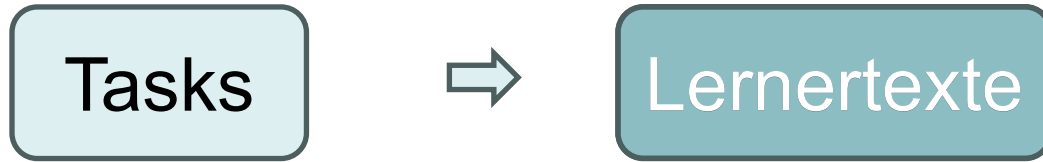
0 secondes

- Schulsprache (LoS): Französisch
- Zielsprache (FL): Deutsch
- Modalität: mündlich
- Medium: Computer
- Texttyp: deskriptiv
- Thema: persönlich
- Struktur: weniger

Suivant



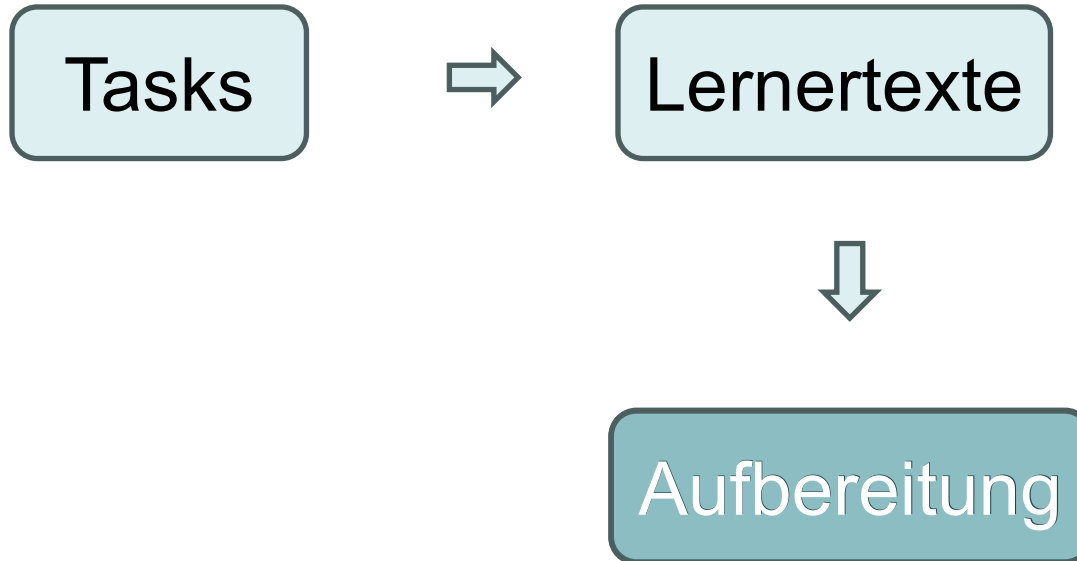
# SWIKO – Komponenten und Workflow



## Datenerhebung

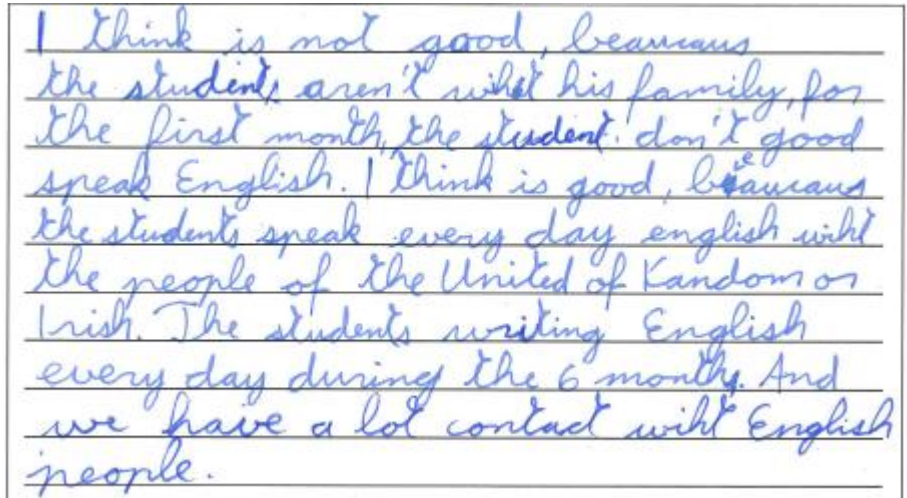
- zwischen 2017 und 2022
- 340 SchülerInnen aus 16 Klassen der Sek I (15-17 Jahre alt)
- Deutsch- und französischsprachige Schweiz

# SWIKO – Komponenten und Workflow



# SWIKO – Aufbereitung

Die Originaltexte wurden...

A photograph of a piece of lined paper with handwritten text in blue ink. The text is written in a cursive script and contains several grammatical errors. The paper is placed on a white background.

I think is not good, beacaus  
the student, aren't with his family, for  
the first month, the student don't good  
speak English. I think is good, beacaus  
the student speak every day english with  
the people of the United of Kandoms or  
Irish. The student's writing English  
every day during the 6 months. And  
we have a lot contact with English  
people.

# SWIKO – Aufbereitung

Die Originaltexte wurden...



transkribiert und  
manuell annotiert...

Copyright 2012 project Merlin, <http://merlin-platform.eu>; adapted for

Transcriber: JBE

Checked by: NMU

Author ID: Mo121

Task ID: SWI08\_fE

Medium: p

Original Text:

I think is not good, beaucas the students aren't whit his family, for the first month, the student don't good speak English. I think is good beaucas the students speak every day english wiht the people of the United of Kandom or Irish. The students writing English every day during the #6# month. And we have a lot contact wiht English people.

Tagged Text:

I think is not good, [beaucas because] the students aren't [wiht with] his family, for the first month, the student don't good speak English. I think is good, [beaucas because] the students speak every day [english English] [wiht with] the people of the United [of Kandom Kingdom] or Irish. The students writing English every day during the #6# months. And we have a lot contact [wiht with] English people.

*In XMLmind with the help of an xml-specification kindly provided by the **MERLIN project**: “MERLIN- Multilingual Platform for European Reference Levels: Interlanguage Exploration in Context” (project number: 518989-LLP-1-2011-1-DE-KA2-KA2MP).*

# SWIKO – Aufbereitung

Die Originaltexte wurden...



transkribiert und  
manuell annotiert...



automatisch POS annotiert...

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	original	doc_id	token	common.fen.POS.tag	lemma	litr	wclass	desc	stop	stem	idx	sntc	
2	I	SWI08_fe_I		PRO:PER	PP	I	1 pronoun	NA	FALSE	NA		1	1
3	think	SWI08_fe_think		VER:PRE	VVP	think	5 verb	NA	FALSE	NA		2	1
4	is	SWI08_fe_is		VER:PRE	VBZ	be	2 verb	NA	FALSE	NA		3	1
5	not	SWI08_fe_not		ADV	RB	not	3 adverb	NA	FALSE	NA		4	1
6	good	SWI08_fe_good		ADJ	JJ	good	4 adjective	NA	FALSE	NA		5	1
7	,	SWI08_fe_		\$,	,	,	1 comma	NA	FALSE	NA		6	1
8	because	SWI08_fe_because		PRP	IN	because	7 prepositio	NA	FALSE	NA		7	1
9	the	SWI08_fe_the		DET	DT	the	3 determine	NA	FALSE	NA		8	1
10	students	SWI08_fe_students		NN	NNS	student	8 noun	NA	FALSE	NA		9	1
11	aren't	SWI08_fe_are		VER:PRE	VBP	be	3 verb	NA	FALSE	NA		10	1
12	aren't((2))	SWI08_fe_n't		ADV	RB	n't	3 adverb	NA	FALSE	NA		11	1
13	whit	SWI08_fe_whit		NN	NN	whit	4 noun	NA	FALSE	NA		12	1
14	his	SWI08_fe_his		PRO:POS	PP\$	his	3 pronoun	NA	FALSE	NA		13	1
15	family	SWI08_fe_family		NN	NN	family	6 noun	NA	FALSE	NA		14	1
16	,	SWI08_fe_		\$,	,	,	1 comma	NA	FALSE	NA		15	1
17	for	SWI08_fe_for		PRP	IN	for	3 prepositio	NA	FALSE	NA		16	1
18	the	SWI08_fe_the		DET	DT	the	3 determine	NA	FALSE	NA		17	1
19	first	SWI08_fe_first		ADJ	JJ	first	5 adjective	NA	FALSE	NA		18	1
20	month	SWI08_fe_month		NN	NN	month	5 noun	NA	FALSE	NA		19	1

by TreeTagger (Schmid 2013), wrapped in the koRpus package (Michalke 2017) in R (R Core Team 2022).

# SWIKO – Aufbereitung

Die Originaltexte wurden...



transkribiert und  
manuell annotiert...



automatisch POS annotiert...



und in EXMARaLDA  
weiterbearbeitet.

	0	1	2	3	4	5	6	7	8	9	10
Mo121 [tok]	I	think	is	not	good	,	beaucaus	the	students	aren't	
Mo121 [ctok]	I	think	is	not	good	,	because	the	students	are	n't
Mo121 [lemma]	I	think	be	not	good	,	because	the	student	be	n't
Mo121 [clemma]											
Mo121 [commonPOS]	PRO-PR	VER-PR	VER-PR	ADV	ADJ	\$,	PRP	DET	NN	VER-PR	ADV
Mo121 [lg-specific POS]	PP	VVP	VBZ	RB	JJ	,	IN	DT	NNS	VBP	RB
Mo121 [cpos]											
Mo121 [tag]							error				
Mo121 [markup]											
[comment]											

Schmidt & Wörner 2009

# SWIKO – Umfang

(Stand Ende Projektphase SWIKO 2024)

Sprache	Deutsch		Französisch		Englisch		TOTAL
	FL	LoS	FL	LoS	FL	LoS	
Klasse	10 & 11	11 & 12	11 & 12	10 & 11	10 - 12	10	
schriftlich							
Original	566	355	396	426	770	103	2'616
Transkript	543	347	322	384	684	102	2'382
Token	23'737	23'667	17'567	27'584	45'173	8'556	146'284
mündlich							
Original	49	72	57	64	140	28	410
Transkript	42	72	7	0	108	8	237
Token	2'174	13'530	177	0	15'118	3'028	34'097

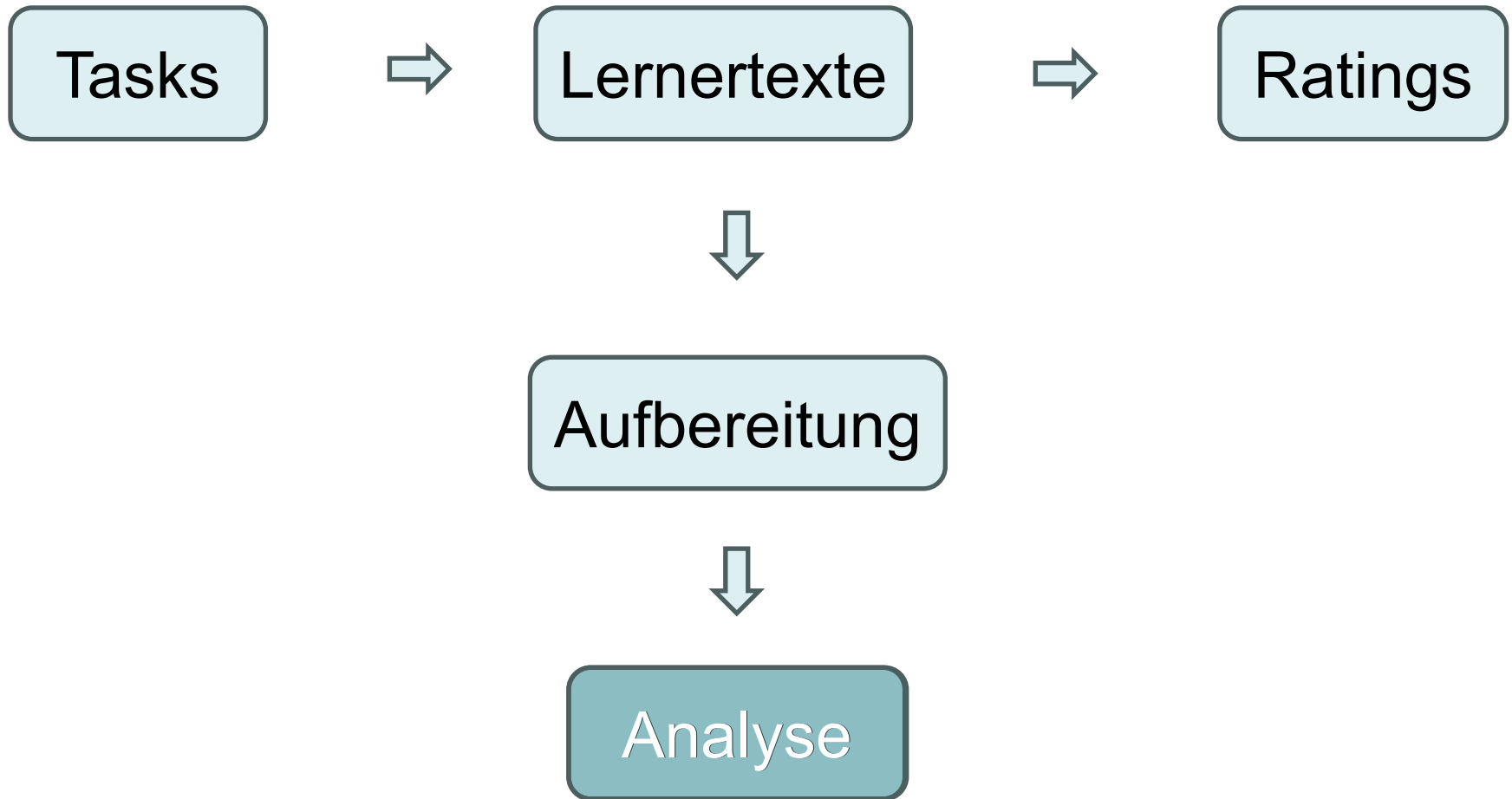
# SWIKO – Komponenten und Workflow



- Analytisch:  
Lexikon, Grammatik, Orthographie, Text  
(Lenz & Studer 2007, Council of Europe 2020)
- 1550 Texte: DaF, FLE & EFL
- 47 geschulte RaterInnen
- Mehrfacetten-Rasch-Analyse  
(Eckes 2015, Linacre 1994, 2022)



# SWIKO – Komponenten und Workflow



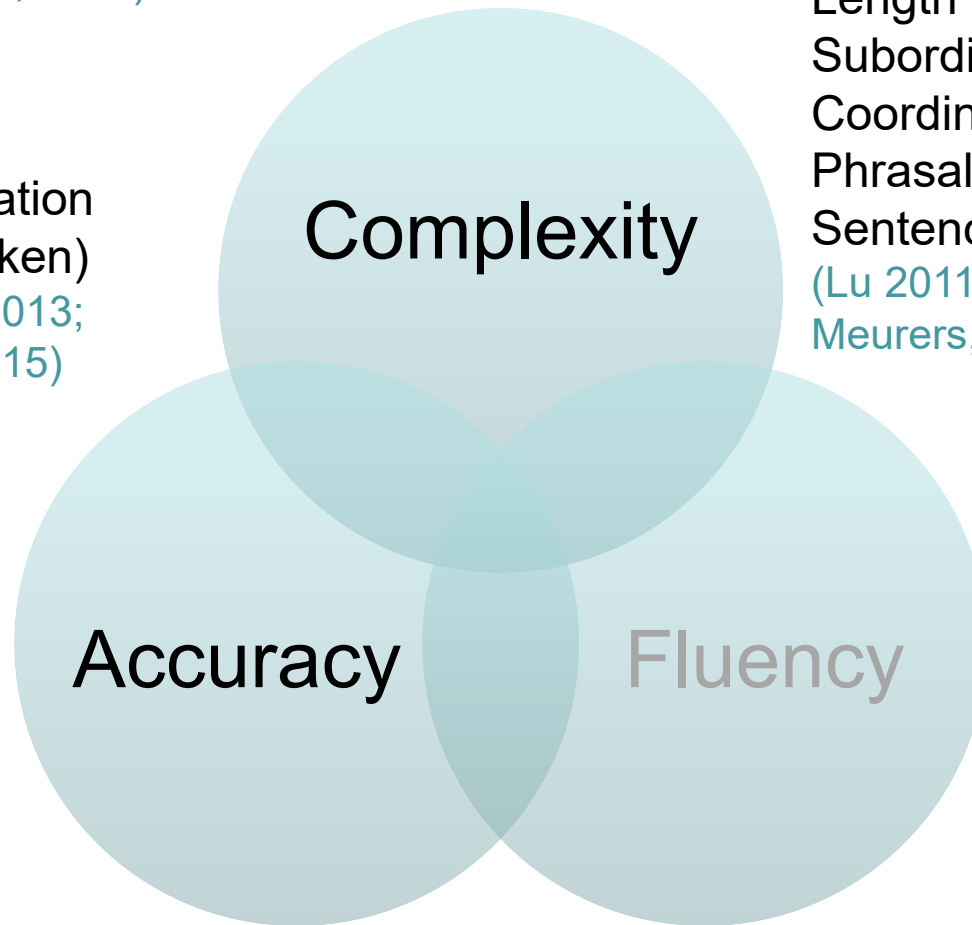
# CAF Framework

(Housen et al. 2012, 2019)

Lexical density  
Lexical diversity  
Lexical sophistication  
Text length (n. token)  
(Ure 1971; Jarvis 2013;  
Kyle & Crossley 2015)

Length of production unit  
Subordination  
Coordination  
Phrasal sophistication  
Sentence complexity  
(Lu 2011; CTAP Chen &  
Meurers, 2016)

?



# Inhalt

- SWIKO
- Fehlerannotation
- Ausgewählte Ergebnisse
- Fazit

# Fehlerannotation

Fehler = Oberflächenreflex der *interlanguage*  
(*interlanguage* nach Selinker 1972)

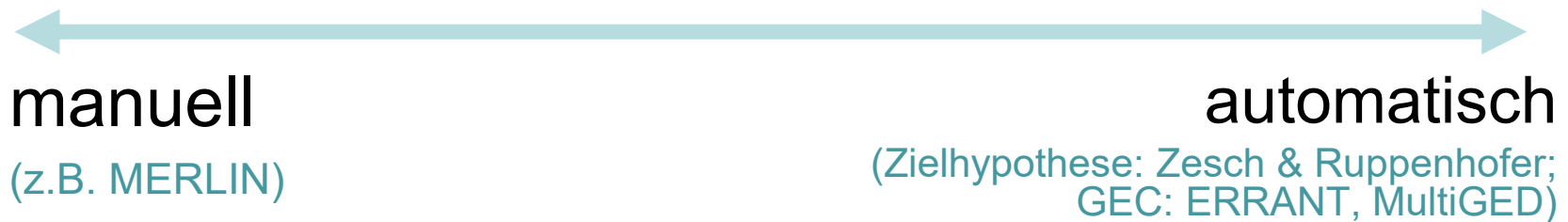
*Comparative fallacy* (e.g., Bley-Vroman 1983) – aber:

- “all the studies that compare learners of different proficiency levels are in fact based on an underlying L1 norm, as proficiency is usually assessed with an L1 target in mind” (Granger 2015:14)
- *accuracy* erforschen & Hypothesen der Lernenden über Zielsprache aufzeigen (Lüdeling & Hirschmann 2015)

# Fehlerannotation

«Wunschliste»:

- transparent, nachvollziehbar, konsistent
- wenig zeitaufwändig
- Grob- und Feintags...?



# Halbautomatische Fehlerannotation

## 1) Manuelle Formulierung einer Zielhypothese

(Corder 1981; Lüdeling et al. 2005, 2008, 2021)

## 2) Automatisches Tagging

(Lüdeling & Hirschmann 2015, Weiss 2020)

# 1) Zielhypothese

*“We identify errors by comparing original utterances with what I shall call reconstructed utterances, that is, correct utterances having the meaning intended by the learner.”* (Corder 1981:37)

→ Zielhypothese (Lüdeling et al. 2005, 2008, 2021)

TH1 = orthographisch & grammatisch korrekt

TH2 = pragmatisch & semantisch korrekt  
(e.g., FALKO, MERLIN)

# 1) Zielhypothese

*«Ich bin dafür for das schule begen 10h.»*

*a) Ich bin dafür, dass die Schule um 10h beginnt.*

*b) Ich bin für einen Schulbeginn um 10h.*



# 1) Zielhypothese

*«Ich ajutire der Ferien in Italy»*

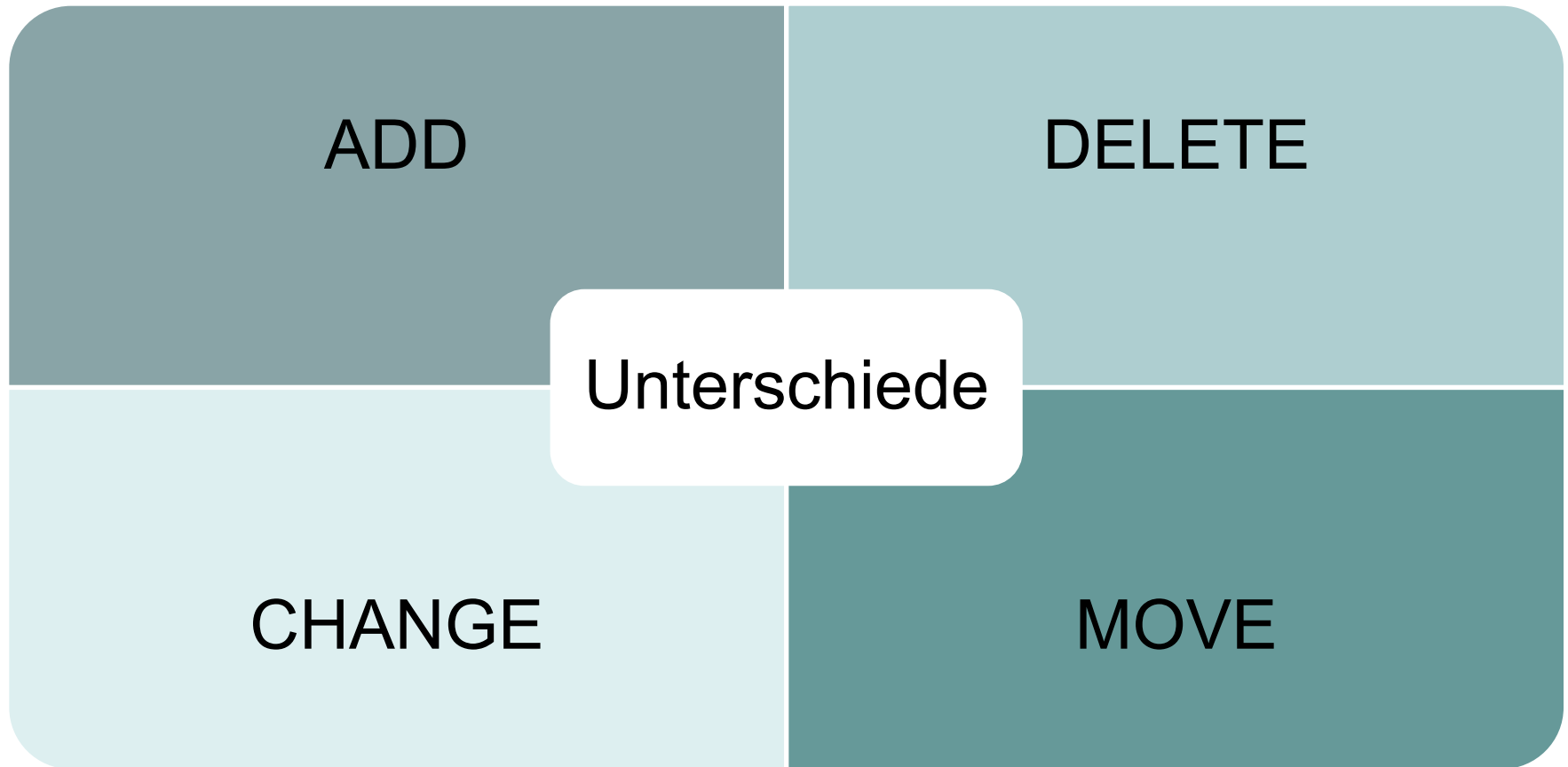
→ ajutire = füge hinzu vs. ergänze

*«wir dependsiren 1000 frs pro jahr»*

→ dependsiren = geben...aus vs. verprassen?

## 2) Automatisches Tagging

(Lüdeling & Hirschmann 2015)



## 2) Automatisches Tagging

*in Schweiz es gipt 8,2 Milion die tiere domestique.*

	0	1	2	3	4	5	6	7	8	9	10	11
<b>Bsp0 [original]</b>	in		Schweitz	es	gipt		8,2	Milion	die	tiere	domestique	.
<b>Bsp0 [TH1]</b>	In	der	Schweiz		gibt	es	8,2	Millionen		Haustiere		.
<b>TH1-DIFF</b>	CHA	ADD	CHA	MOVS	CHA	MOVT		CHA	DEL	CHA		

## 2) Automatisches Tagging

*in Schweiz es gipt 8,2 Milion die tiere domestique.*

	0	1	2	3	4	5	6	7	8	9	10	11
<b>Bsp0 [original]</b>	in		Schweitz	es	gipt		8,2	Milion	die	tiere	domestique	.
<b>Bsp0 [ctok]</b>	In		Schweiz	es	gibt		8,2	Million	die	Haustiere		.
<b>Bsp0 [TH1]</b>	In	der	Schweiz		gibt	es	8,2	Millionen		Haustiere		.

## 2) Automatisches Tagging I: Orthografie

*in Schweiz es gibt 8,2 Milion die tiere domestique.*

	0	1	2	3	4	5	6	7	8	9
<b>Bsp1 [original]</b>	in	Schweitz	es	gibt	8,2	Milion	die	tiere	domestique.	
<b>Bsp1 [ctok]</b>	In	Schweiz	es	gibt	8,2	Million	die	Haustiere		
<b>O_Capit</b>	O_cap_bos							O_cap_noun		
<b>O_graph</b>		O_graph_del		O_graph_cha		O_graph_add		O_graph_mul		
<b>O_wdbd</b>								O_wdbd_mg		

*DELETE      CHANGE      ADD*

*GROSS-/KLEINSCHREIBUNG*

*WORTGRENZEN*

## 2) Automatisches Tagging II: Grammatik

*In Schweiz es gibt 8,2 Million die Haustierte.*

	0	1	2	3	4	5	6	7	8	9	10	11
<b>Bsp2 [original]</b>	in		Schweiz	es	gibt		8,2	Milion	die	tiere	domestique	.
<b>Bsp2 [ctok]</b>	In		Schweiz	es	gibt		8,2	Million	die	Haustiere		.
<b>Bsp2 [TH1]</b>	In	der	Schweiz		gibt	es	8,2	Millionen		Haustiere		.
<b>G_add</b>		G_ART_add										
<b>G_del</b>									G_ART_del			
<b>G_cha</b>								G_NN_cha				
<b>G_mov</b>				G_PPER_movs	G_PPER_movt							
<b>G_wordorder</b>	G_wordorder											

## 2) Automatisches Tagging III

*in Schweiz es gibt 8,2 Milion die tiere domestique.*

	0	1	2	3	4	5	6	7	8	9	10	11
<b>Bsp3 [original]</b>	in		Schweitz	es	gibt		8,2	Milion	die	tiere	domestique	.
<b>Bsp3 [ctok]</b>	In		Schweiz	es	gibt		8,2	Million	die	Haustiere		.
<b>Bsp3 [TH1]</b>	In	der	Schweiz		gibt	es	8,2	Millionen		Haustiere		.
<b>O_Capit</b>	O_cap_bos									O_cap_noun		
<b>O_graph</b>			O_graph_del		O_graph_cha			O_graph_add		O_graph_mul		
<b>O_wdbd</b>										O_wdbd_mg		
<b>G_add</b>		G_ART_add										
<b>G_del</b>									G_ART_del			
<b>G_cha</b>								G_NN_cha				
<b>G_mov</b>				G_PPER_movs		G_PPER_mov						
<b>G_wordorder</b>	G_wordorder											

# Bilanz Fehlerannotation


- Transparent, nachvollziehbar, konsistent ✓
- Wenig(er) zeitaufwändig (✓)
- Grob- und Feintags ✓

## Optimierungsmöglichkeit:

- Orthographiefehler, die auch als Grammatikfehler interpretiert werden könnten (z.B. *habst*)
- Reliabilität?



# Fehlerannotation Potenzial I: SWIKOweb



Menu

INSTITUT FÜR  
INSTITUT DE  
ISTITUTO DI  
INSTITUT DA  
INSTITUTE OF

MEHRSPRACHIGKEIT  
PLURILINGUISME  
PLURILINGUISMO  
PLURILINGUITAD  
MULTILINGUALISM

SWIKOweb

LOGOUT

 ⓘSearch into  Output from  display format  ☐ Case Sensitive

Hide other filters

file

student

task


language

rating

remove all filters

Files project  status student student  canton  year  gender  grade Task file type  mode  Task Number  topic Language language of schooling  language of document Rating overall  lexicon  grammar  spelling  text

# Fehlerannotation Potenzial I: SWIKOweb




The screenshot shows the SWIKOweb search interface. At the top left is a 'Menu' button. Next to it is the logo of the 'INSTITUT FÜR MEHRSPRACHIGKEIT' (Institute for Multilingualism), which also lists 'INSTITUT DE PLURILINGUISME', 'ISTITUTO DI PLURILINGUISMO', 'INSTITUT DA PLURILINGUITAD', and 'INSTITUTE OF MULTILINGUALISM'. The 'SWIKOweb' logo is to the right. On the far right is a 'LOGOUT' link. Below the header is a search bar containing the query 'kein.\* WITH #g=.\*cha'. To the right of the search bar is a 'Suchen' (Search) button with an information icon. Below the search bar are several dropdown menus: 'Search into' (set to 'Corrected Token'), 'Output from' (set to 'Token'), and 'display format' (set to 'file list'). There is also a checkbox for 'Case Sensitive'. Below these are several filter buttons: 'Hide other filters', 'Datei', 'Lemende(r)', 'task', 'Sprache', 'rating', and 'Alle Filter entfernen'.

Suche auf mehreren Ebenen möglich, z. B.

kein.\* WITH #g=.\*cha #lgPOS=NN

«kein» mit Flexionsfehler gefolgt von Nomen

# Fehlerannotation Potenzial I: SWIKOweb



Menu

INSTITUT FÜR  
INSTITUT DE  
INSTITUTO DI  
INSTITUT DA  
INSTITUTE OF

MEHRSPRACHIGKEIT  
PLURILINGUISME  
PLURILINGUISMO  
PLURILINGUITAD  
MULTILINGUALISM

SWIKOweb

LOGOUT

Suchen ⓘ

Search into

Corrected Token

Output from

Token

display format

file list

☐ Case Sensitive

Hide other filters

Datei

Lernende(r)

task



















Sprache

rating

Alle Filter entfernen

13 von insgesamt 2616 gefundenen Einträgen Ordnen nach undefined.

Filter: Token key=a.TokenSearchable REGEXP 'kein[^\~]\*<[^\~]\*>[^\~]\*cha<[^\~]\*~', Suchbegriff like kein.\* WITH #g=.\*cha, Projekt=SWIKO, Status=5: EXMARaLDA/CLAN (at least)

	#	Konkordanz	Projekt	Jahr	Kanton	Schulsprache	Zielsprache	Thema	Lernende(r)	Geschlecht	Klassenstufe	Text
<input type="checkbox"/>	191	nish in 17 h 30 weil ich habe kein ferien in meinen Freunden after Schu	SWIKO	2017		Französisch	Deutsch	Unterrichtszeiten	Mo122	F	8	
<input type="checkbox"/>	217	studieren. Für mich, das ist kein gut idea für Schuler.	SWIKO	2017		Französisch	Deutsch	Klassenaustausch	Mo131	M	8	
<input type="checkbox"/>	244	Wasser zu gleiten. Ich mag kein Kieg oder wenn Leute sich streiten o	SWIKO	2018		Deutsch	Deutsch	Kleines Portrait	Th141	F	9	
<input type="checkbox"/>	247	bei unserer Schule. Ich mag kein Spinat oder wem ich die Wäsche mache	SWIKO	2018		Deutsch	Deutsch	Kleines Portrait	Fr156	M	10	
<input type="checkbox"/>	686	ufgaben machen haben sie dann kein Freizeit mehr. Es währe aber gut spä	SWIKO	2018		Deutsch	Deutsch	Unterrichtszeiten	Fr157	M	10	
<input type="checkbox"/>	765	Familie oder Freund, man hat kein direkten Kontakt zu ihnen und ich fi	SWIKO	2018		Deutsch	Deutsch	Klassenaustausch	Ba226	M	9	
<input type="checkbox"/>	3582	it mein Freund are good weil kein parent. Ich weil vacation aus der sc	SWIKO	2022		Französisch	Deutsch	Coole Ferien	RI101	F	9	
<input type="checkbox"/>	3743	e es ist Kompliziert, weil du kein Zeit hast. Aber ich will lange schla	SWIKO	2022		Französisch	Deutsch	Unterrichtszeiten	RI212	M	9	
<input type="checkbox"/>	4092	lernen schnell und wir haben kein hilfe von dein Eltern.	SWIKO	2022		Französisch	Deutsch	Klassenaustausch	RI408	F	9	

Deutsch


Deutsch

Kleines Portrait


Fr161


M


10



5: EXMARaLDA/CLAN







# Fehlerannotation Potenzial I: SWIKOweb

Menu

INSTITUT FÜR  
INSTITUT DE  
INSTITUTO DI  
INSTITUT DA  
INSTITUTE OF

MEHRSPRACHIGKEIT  
PLURILINGUISME  
PLURILINGUISMO  
PLURILINGUITAD  
MULTILINGUALISM

SWIKOweb

LOGOUT

kein.\* WITH #g=.\*cha

Suchen

1

Search into

Corrected Token

Output from

Token

display format

file list

☐ Case Sensitive

Hide other filters

Datei

Lernende(r)

task

Sprache

rating

Alle Filter entfernen

ke kein.\* WITH #g=.\*cha #lgPOS=NN, Projekt=SWIKO

chulsprache	Zielsprache	Thema	Lernende(r)	Geschlecht	Klassenstufe	Text	Status	Transkription	POS- getaggt	EXMARaLDA	Insgesamt	Lexikon	Grammatik	Rechtschreibung	Text
anzösisch	Deutsch	Unterrichtszeiten	Mo122	F	8	 5: EXMARaLDA/CLAN					A1.2	A1.2	A1.2	A1.2	A1.2
deutsch	Deutsch	Kleines Portrait	Th141	F	9	 5: EXMARaLDA/CLAN									
deutsch	Deutsch	Kleines Portrait	Fr156	M	10	 5: EXMARaLDA/CLAN									
deutsch	Deutsch	Unterrichtszeiten	Fr157	M	10	 5: EXMARaLDA/CLAN									
anzösisch	Deutsch	Coole Ferien	Ri101	F	9	 5: EXMARaLDA/CLAN					A1.1	A1.1	A1.1	A1.1	A1.1
anzösisch	Deutsch	Unterrichtszeiten	Ri212	M	9	 5: EXMARaLDA/CLAN					A2.2	A2.2	A2.2	A2.2	A2.2
anzösisch	Deutsch	Klassenaustausch	Ri408	F	9	 5: EXMARaLDA/CLAN					A2.1	A2.1	A2.1	A2.2	A2.1
anzösisch	Deutsch	Coole Ferien	Ri501	F	9	 5: EXMARaLDA/CLAN					A2.2	A2.2	A2.2	B1.1	A2.2
anzösisch	Deutsch	Unterrichtszeiten	Ri521	F	9	 5: EXMARaLDA/CLAN					B1.2	B1.1	B1.1	B1.2	B1.2
deutsch	Deutsch	Kleines Portrait	Fr161	M	10	 5: EXMARaLDA/CLAN									

# Fehlerannotation Potenzial II: Vergleiche

(Contrastive Interlanguage Analysis: Granger 2015)

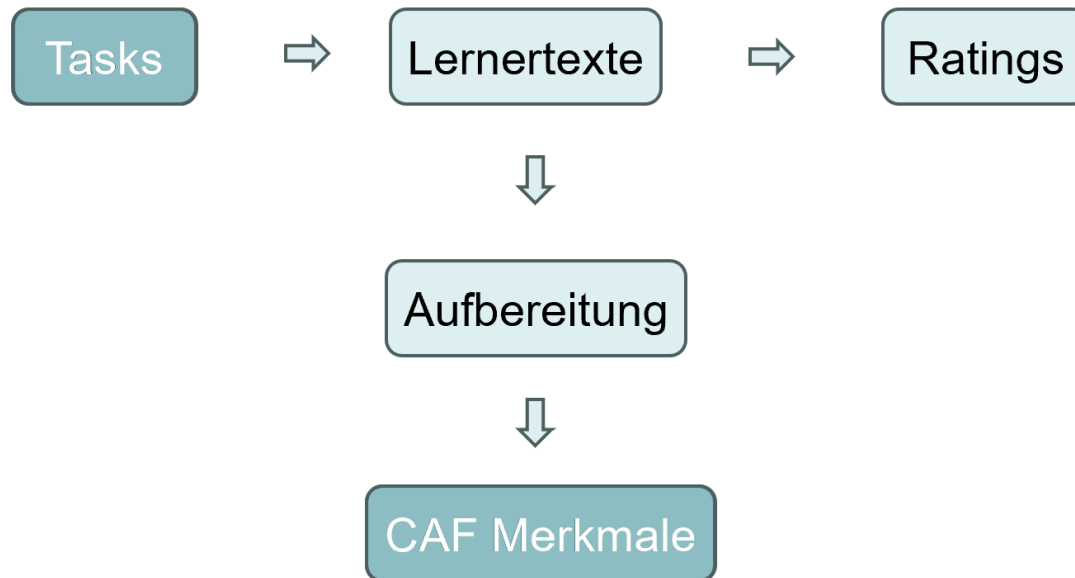
	Deutsch als Fremdsprache (n = 543)	Deutsch als Schulsprache (n = 341)
Anz. “nicht”	251	189
“nicht” + Satzbau	47 (18.7%)	2 (1.1%)
Anz. “kein”	37	55
“kein” + Flexionsfehler	8 (21.6%)	5 (9.1%)
“nicht” statt “kein”	40 (15.9%)	1 (0.5%)

# Inhalt

- SWIKO
- Fehlerannotation
- **Ausgewählte Ergebnisse**
- Fazit

# Ausgewählte Ergebnisse I

Inwiefern beeinflusst die Aufgabe die sprachlichen CAF-Merkmale der Produktion?



## Tasks

Texttyp

Thema

Struktur

## sprachliche Merkmale

Lexikalische Komplexität

Syntaktische Komplexität

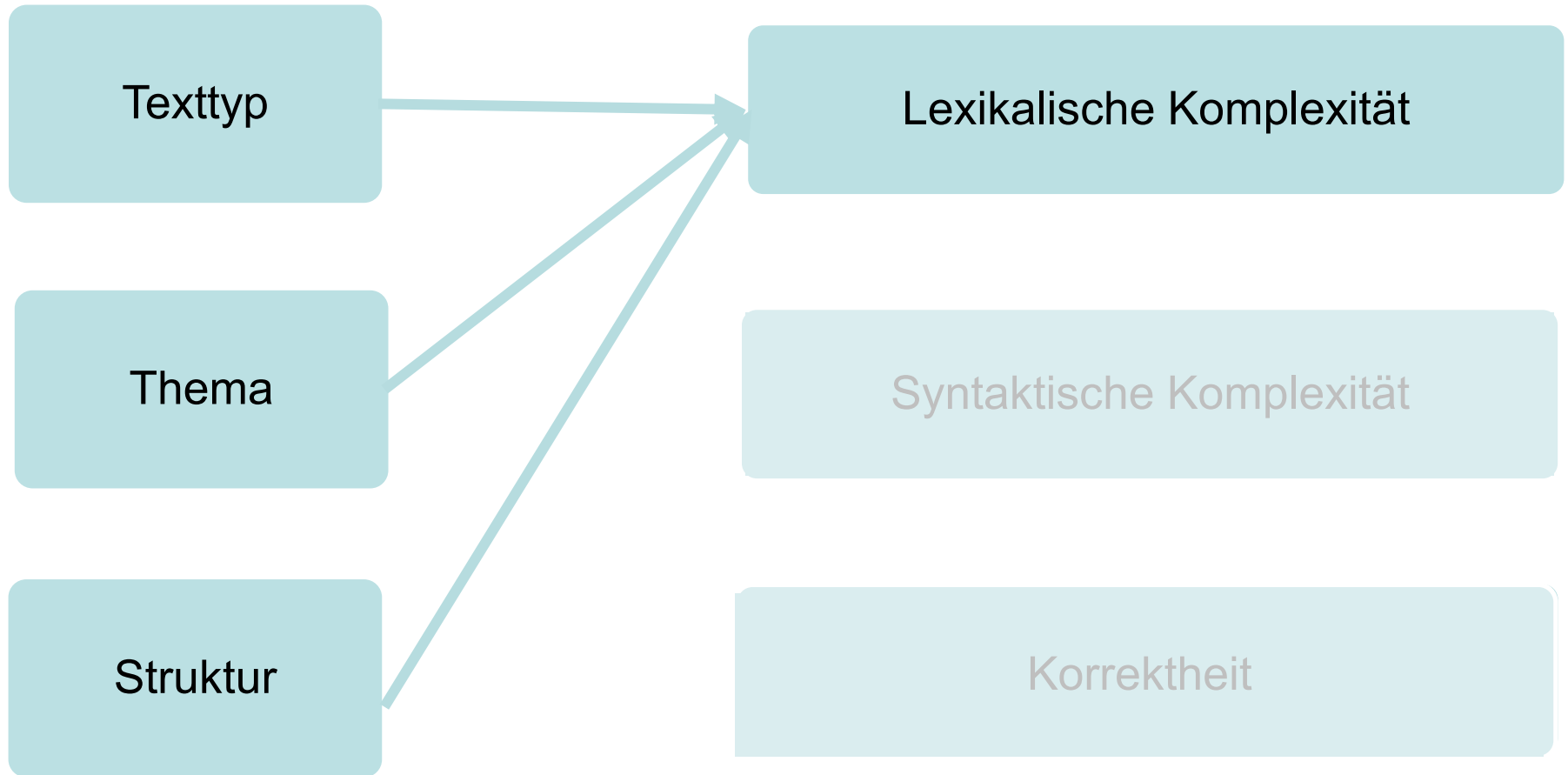
Korrektheit

(Hicks & Studer 2024, Karges et al. 2022, Studer & Hicks 2022, Weiss et al. 2022)



## Tasks

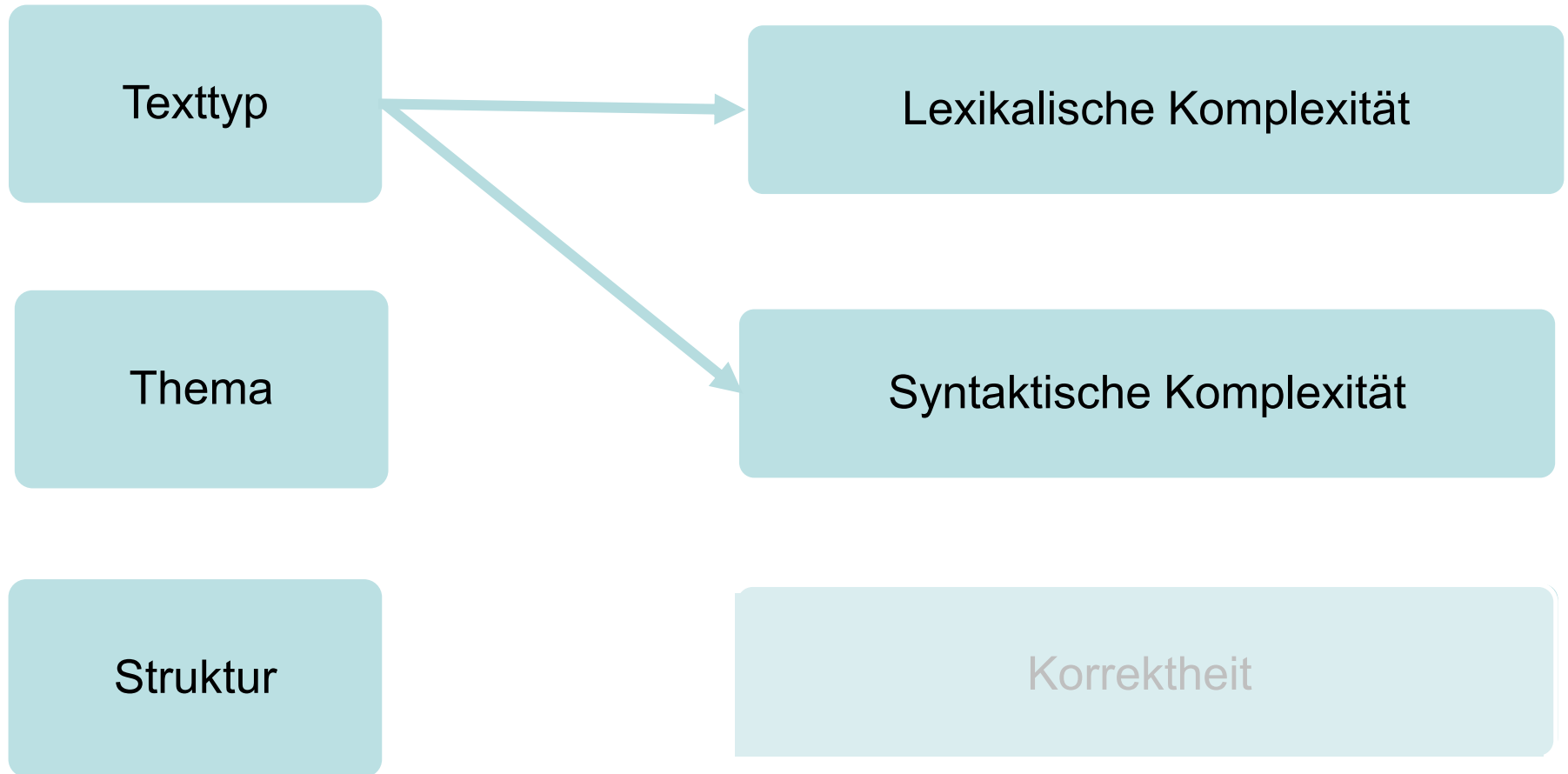
## sprachliche Merkmale



(Hicks & Studer 2024, Karges et al. 2022, Studer & Hicks 2022, Weiss et al. 2022)

## Tasks

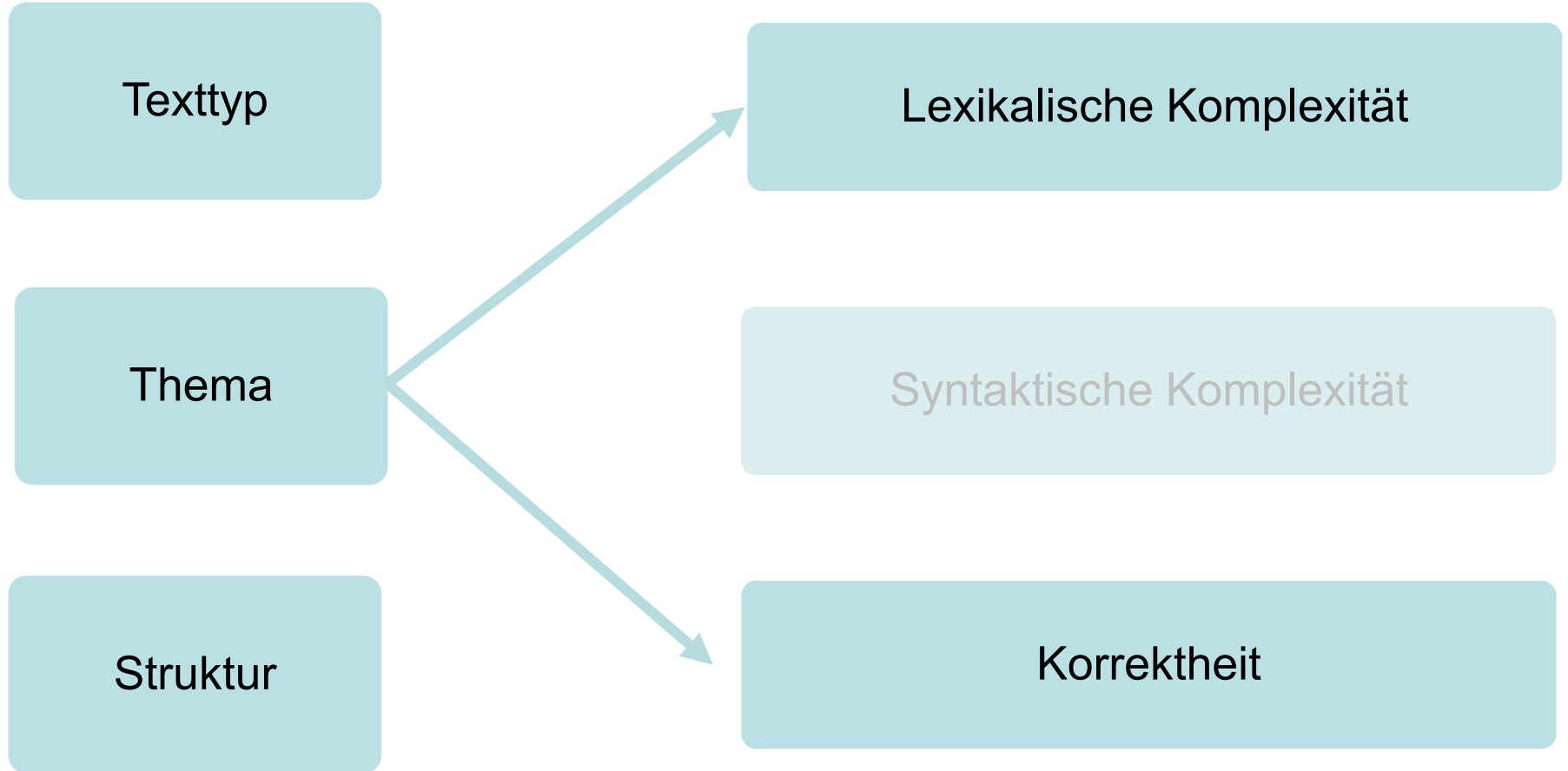
## sprachliche Merkmale



(Hicks & Studer 2024, Karges et al. 2022, Studer & Hicks 2022, Weiss et al. 2022)

## Tasks

## sprachliche Merkmale



(Hicks & Studer 2024, Karges et al. 2022, Studer & Hicks 2022, Weiss et al. 2022)

## SWI05 (Selbstportrait)

*Hallo! Ich heisse Sandra und ich bin 15 Jahre alt. (...) Ich habe zwei katzen, Sie heissen Simba und Luna. Mein Schwester heisst Laura und sie ist 18 aber ich habe keine Brüder. (...)*

### kaum Fehler

kurze, einfache Hauptsätze

viele Substantive

## SWI04 (Erfindungen)

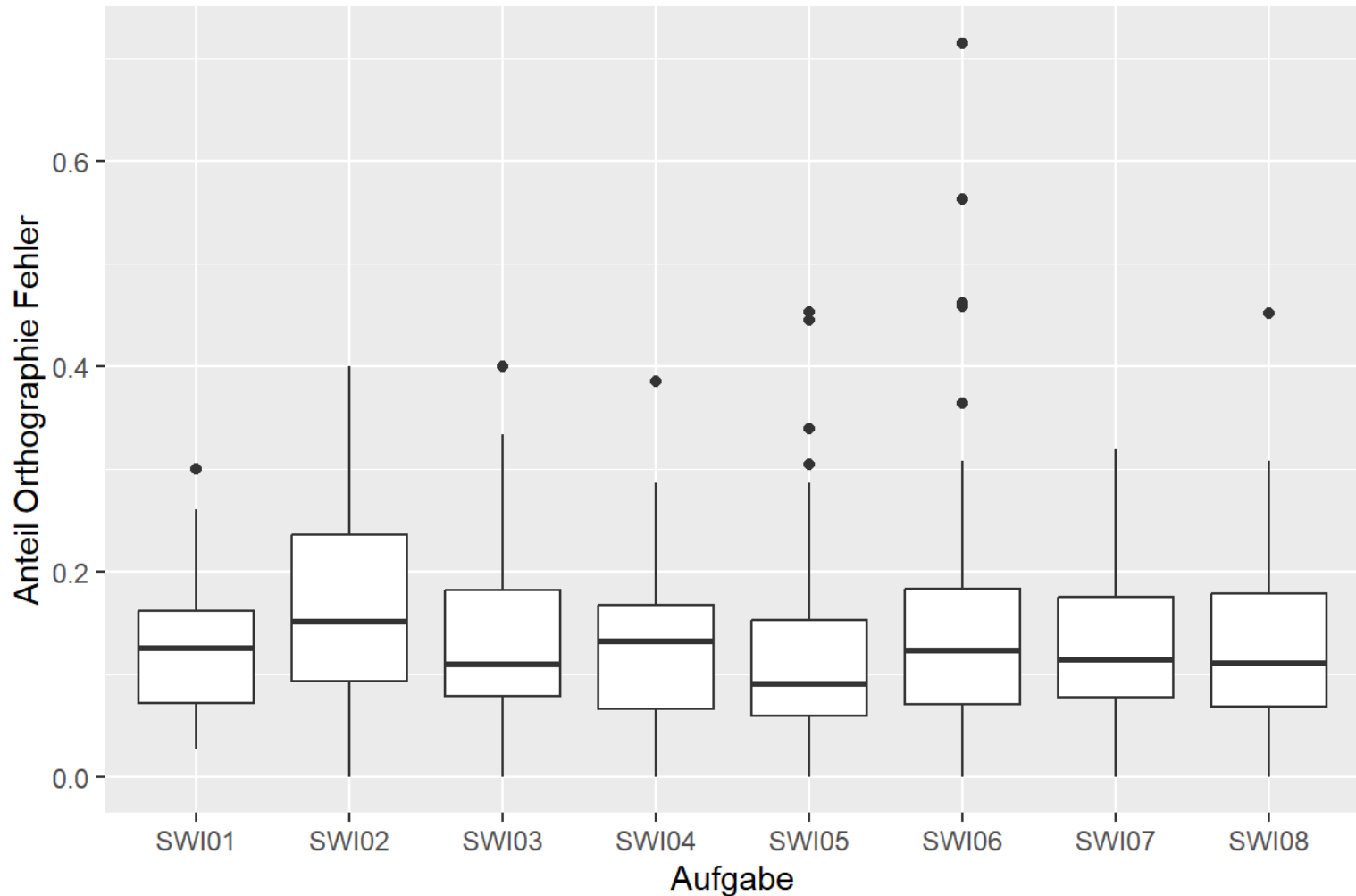
*Für mich die Electricite ist im ersten platz weil ohne electricite es keine Lampe, keine Computer, keine Smartphone mehr gibt. (...) Der Flieger ist für mich in die richtige platz und der bus auch aber der Stieft muss nicht in die liste bin weil es ist nicht ein sehr grossen invention.*

### Fehler auf Token- und Satzebene

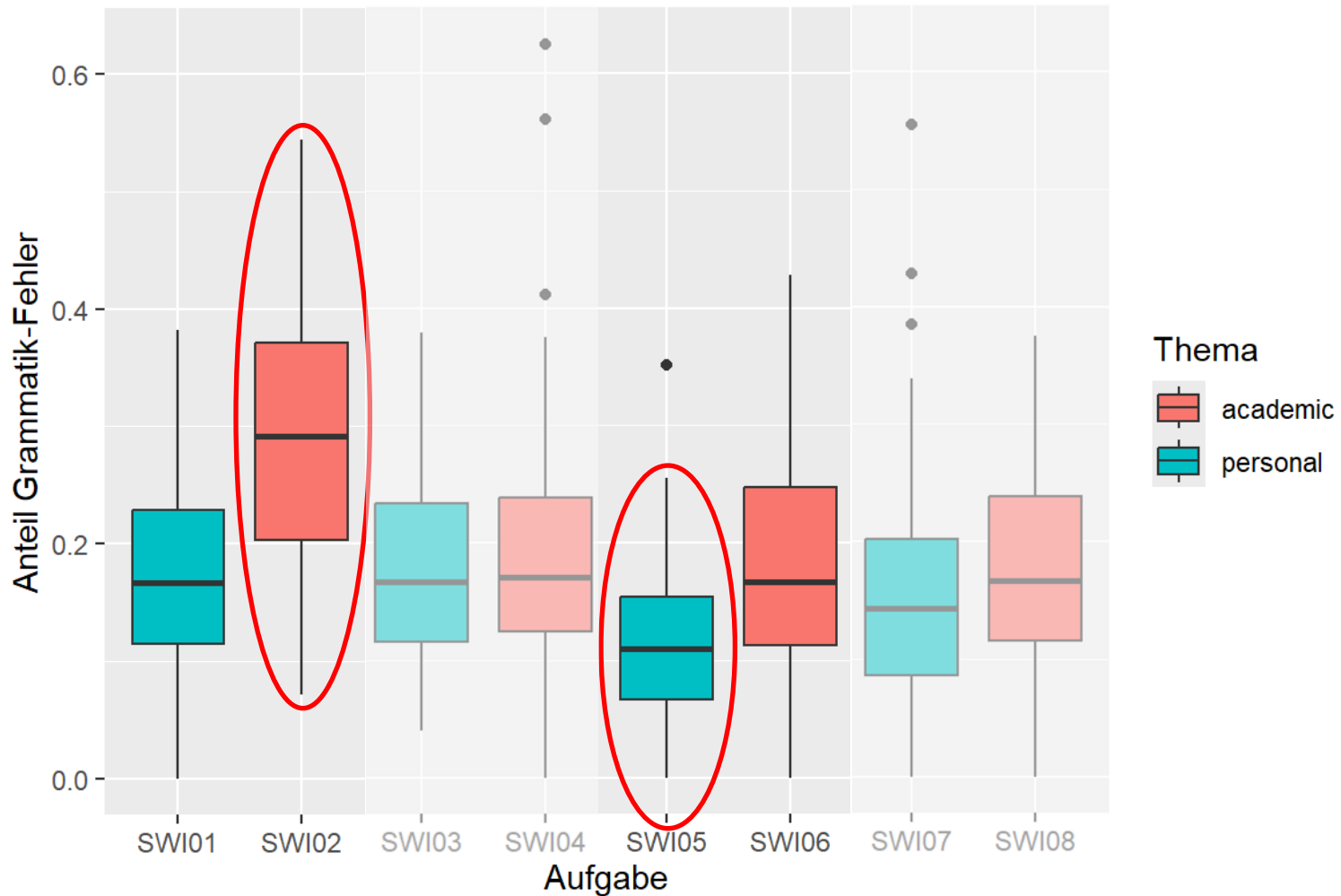
lange Sätze, viele Nebensätze

mehr Funktionswörter & Adjektive

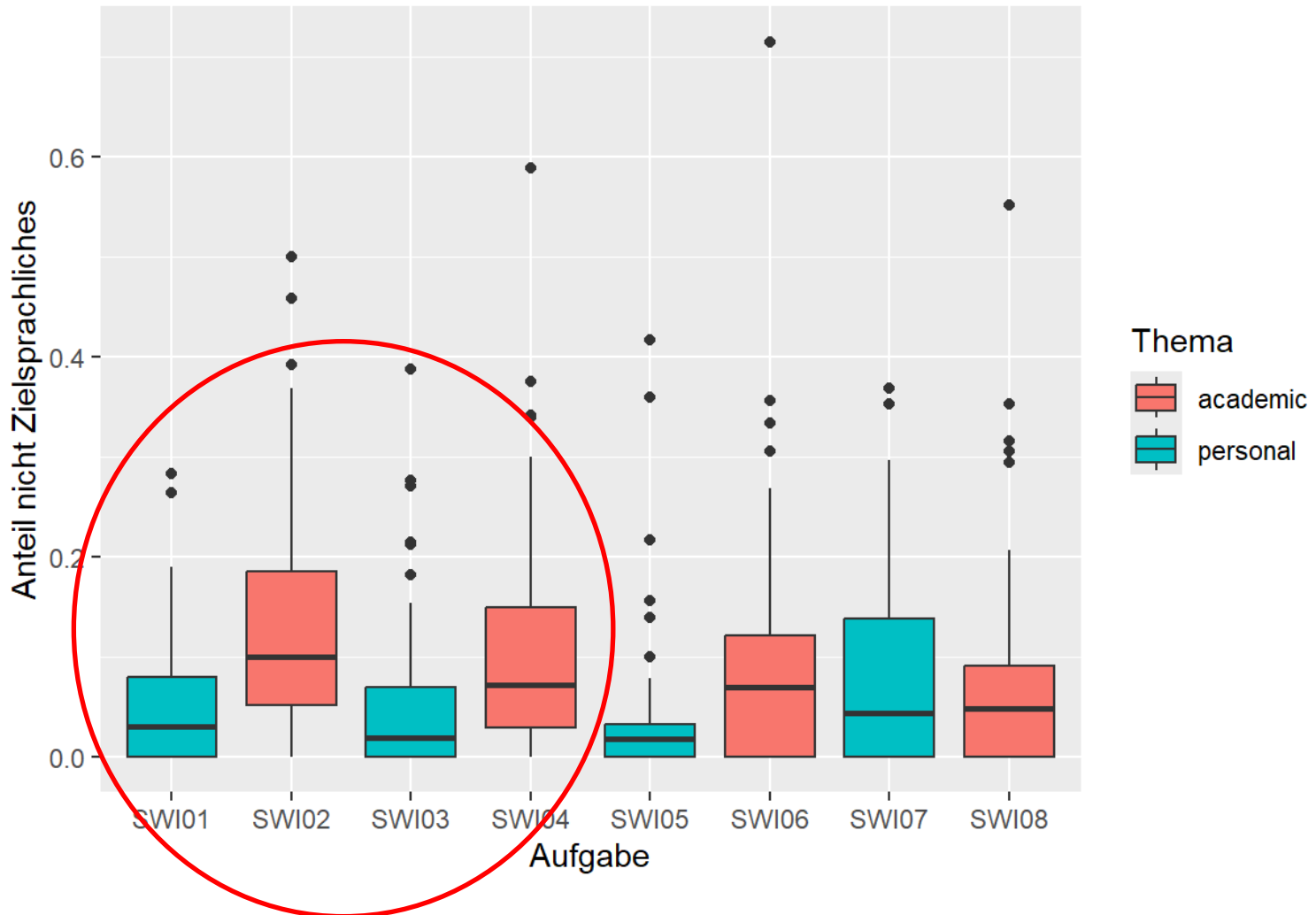
# Aufgabe & Anteil Orthographiefehler



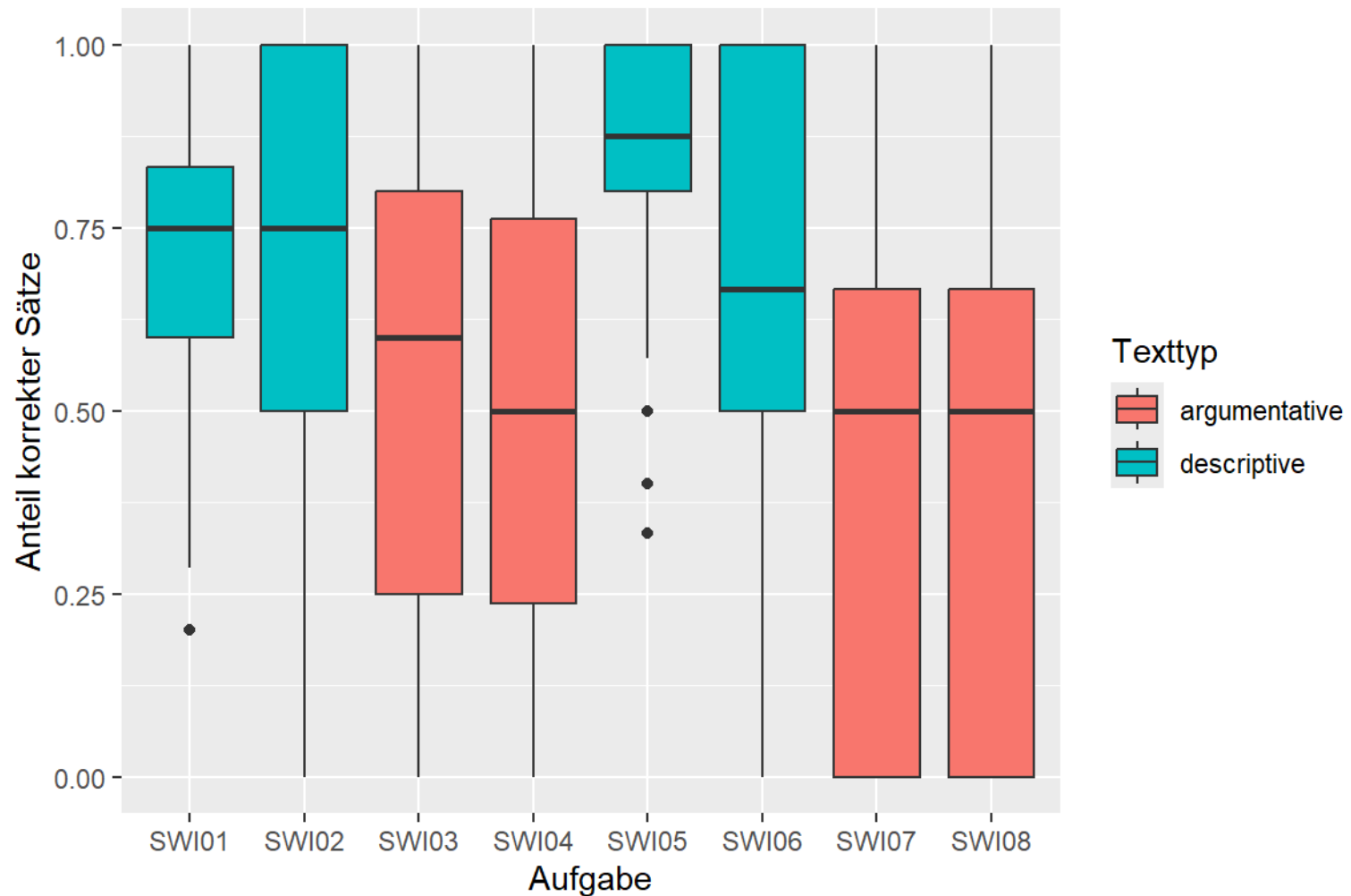
# Aufgabe & Anteil Grammatikfehler



# Aufgabe & Anteil nicht Zielsprachliches



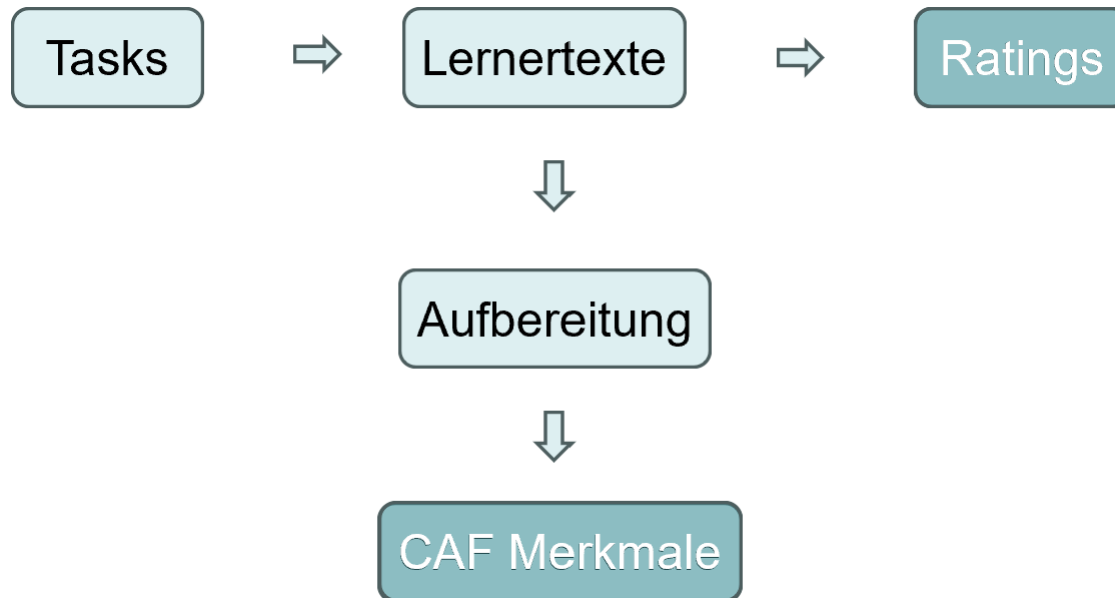
# Aufgabe & Anteil Sätze mit korrektem Satzbau



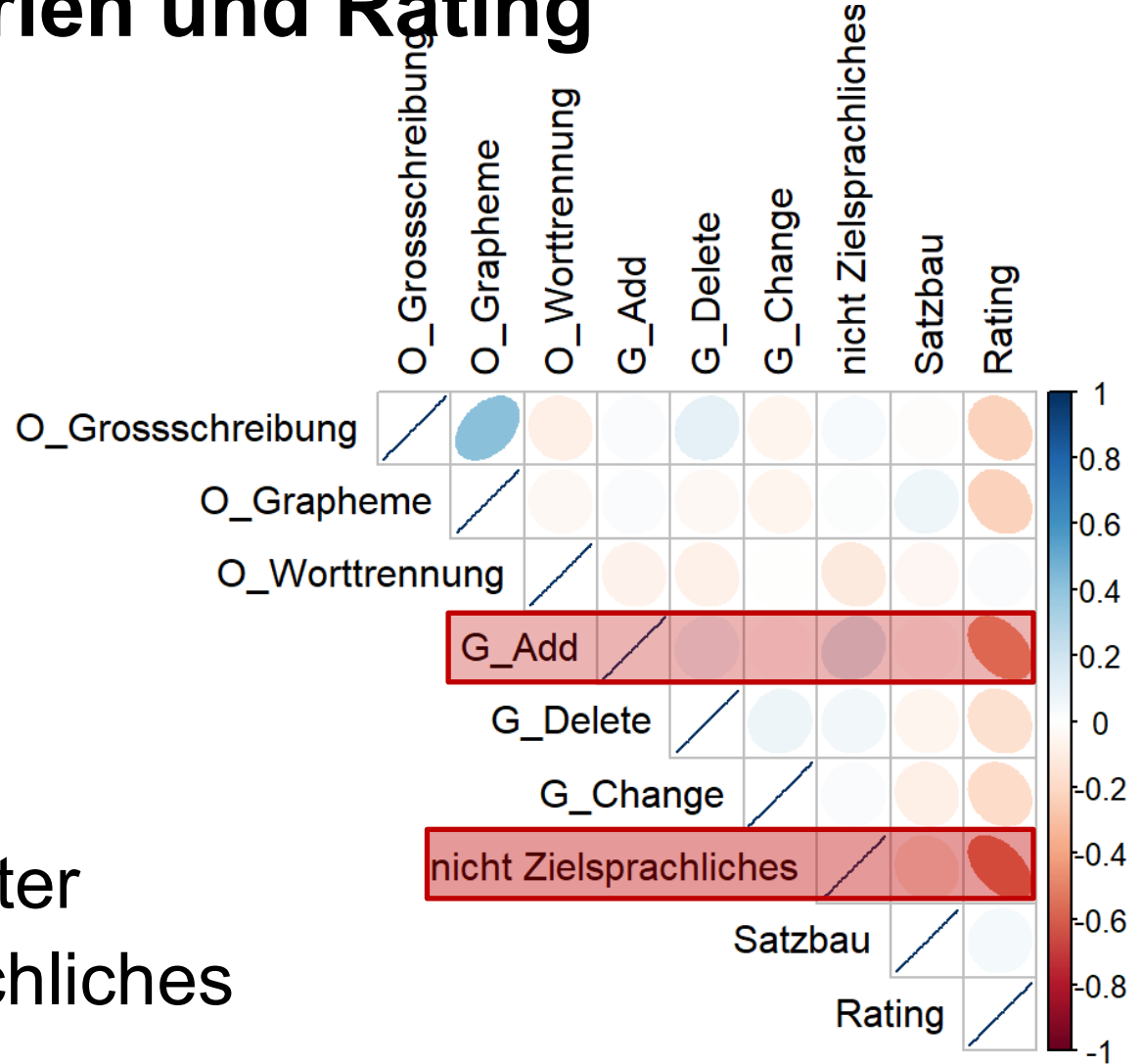


# Ausgewählte Ergebnisse II

Inwiefern korrelieren sprachliche CAF-Merkmale und das Rating der Produktion?

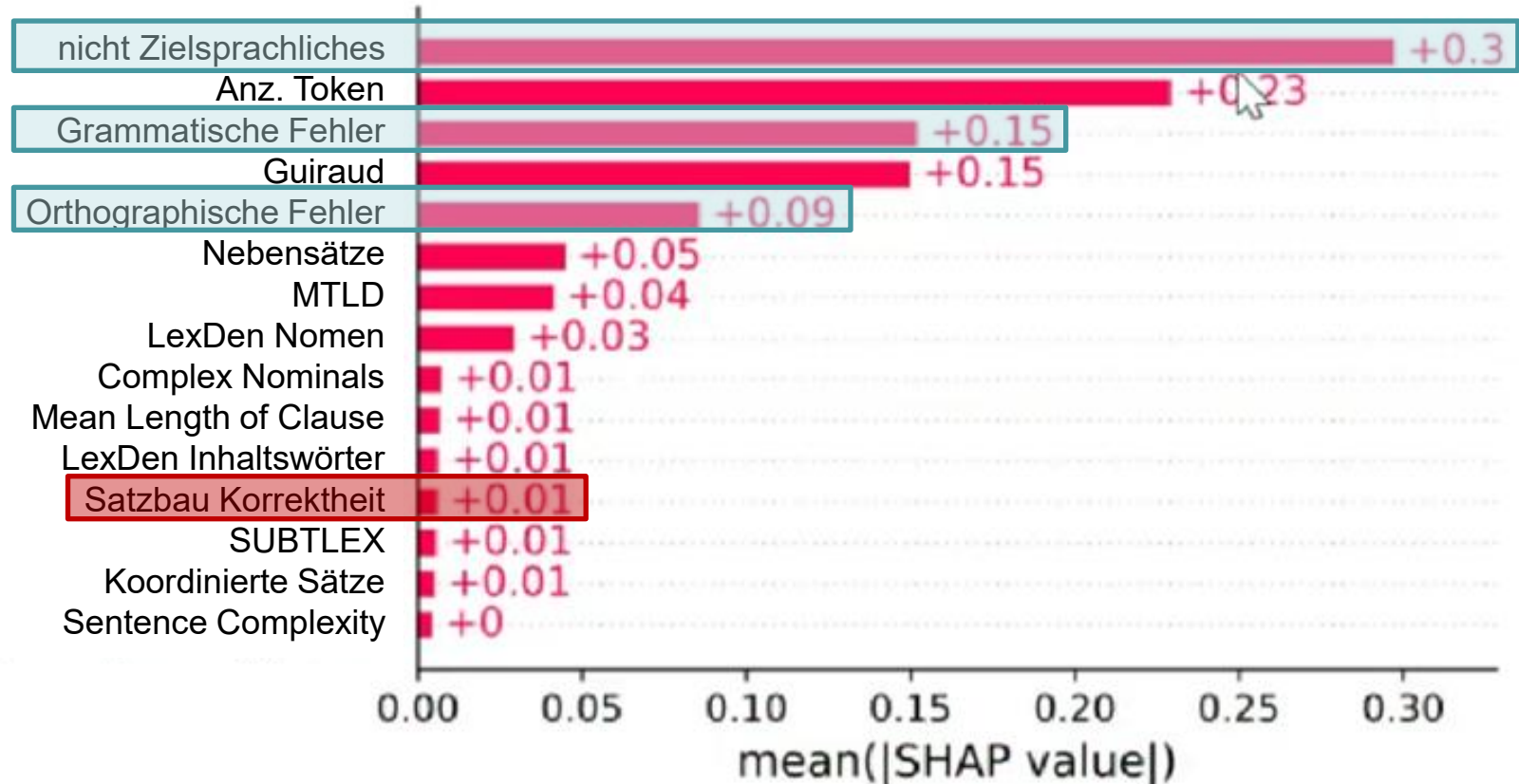


# Fehlerkategorien und Rating



- Fehlende Wörter
- nicht Zielsprachliches

# Shapley: CAF Merkmale & Rating



(Berthele & Hicks i.V.)

# Inhalt

- SWIKO
- Fehlerannotation
- Ausgewählte Ergebnisse
- **Ausblick**

# Recap

- Schweizer Lernerkorpus SWIKO
  - 8 systematisch variierte Aufgaben, 3 Sprachen
  - GER Rating
  - Fehlerannotation: manuelle minimale Zielhypothese, automatisch generierte orthographische & grammatische Feintags
- Ausgewählte Ergebnisse
  - Thema → nicht Zielsprachliches & Grammatikfehler
  - Texttyp → Syntax; Orthographie aufgabenunabhängig
  - Nicht Zielsprachliches und Fehler auf Wortebene beeinflussen Rating, Syntax nicht

# Ausblick

- Aktuelles Erweiterungsprojekt:  
Synergien zwischen den Sprachen in der Schule  
«SWIKO triple S»
- Kontinuierliche Erweiterung SWIKOweb  
[ifm-swiko.unifr.ch](http://ifm-swiko.unifr.ch)

# Bibliografie

- Bertschy, I., Cuenat, M. E., & Stotz, D. (2015). *Lehrplan Französisch und Englisch*. Passepartout - Fremdsprachen an der Volksschule. [https://be.lehrplan.ch/passepartout/Lehrplan\\_Passepartout.pdf](https://be.lehrplan.ch/passepartout/Lehrplan_Passepartout.pdf)
- Chen, X., & Meurers, D. (2016). CTAP: A web-based tool supporting automatic complexity analysis. *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CL4LC)*, 113–119. <https://www.aclweb.org/anthology/W16-4113>
- Corder, S. P. (1981). *Error Analysis and Interlanguage*. Oxford University Press.
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press. Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment: companion volume*. Council of Europe Publishing.
- Eckes, T. (2015). *Introduction to Many-Facet Rasch Measurement*. Peter Lang. <https://www.peterlang.com/document/1045610>
- EDK. (2011). *Grundkompetenzen für die Fremdsprachen: Nationale Bildungsstandards*.
- EDK. (2017). *Empfehlungen zum Fremdsprachenunterricht (Landessprachen und Englisch) in der obligatorischen Schule*.
- Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2020). *Task-based language teaching: Theory and practice*. Cambridge University Press.
- Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *Learner Corpus Research*, 1, 7-24.
- Hicks, N., & Studer, T. (2024). Learner corpora in foreign language education: Examples from the multilingual SWIKO corpus. *Babylonia Journal of Language Education*, 2, 26–35. <https://doi.org/10.55393/babylonia.v2i.388>
- Housen, A., De Clercq, B., Kuiken, F., & Vedder, I. (2019). Multiple approaches to complexity in second language research. *Second Language Research*, 35(1), 3–21. <https://doi.org/10.1177/0267658318809765>
- Housen, A., Kuiken, F., & Vedder, I. (Hrsg.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. John Benjamins.
- Jarvis, S. (2013). Defining and measuring lexical diversity. In M. Daller & S. Jarvis (Hrsg.), *Vocabulary Knowledge: Human ratings and automated measures*. John Benjamins.
- Karges, K., Studer, T., & Hicks, N. S. (2022). Lernersprache, Aufgabe und Modalität: Beobachtungen zu Texten aus dem Schweizer Lernerkorpus SWIKO. *Zeitschrift für germanistische Linguistik*, 50(1), 104–130. <https://doi.org/10.1515/zgl-2022-2050>
- Kempfert, I., & Köhn, C. (2018). An Automatic Error Tagger for German. In I. Pilán, E. Volodina, D. Alfter, & L. Borin (Hrsg.), *Proceedings of the 7th workshop on NLP for Computer Assisted Language Learning* (S. 32–40). LiU Electronic Press. <https://aclanthology.org/W18-7104/>
- Kyle, K., & Crossley, S. A. (2015). Automatically Assessing Lexical Sophistication: Indices, Tools, Findings, and Application. *TESOL Quarterly*, 49(4), 757–786. <https://doi.org/10.1002/tesq.194>
- Lenz, P., & Studer, T. (2008). *Lingualevel: Instrumente zur Evaluation von Fremdsprachenkompetenzen: 5.-9. Schuljahr*. Schulverlag.
- Lenz, P., & Wieden Keller, E. (2019). *Kurzbericht zum Projekt „Ergebnisbezogene Evaluation des Französischunterrichts in der 6. Klasse (HarmoS 8) in den sechs Passepartout-Kantonen“*. Institut für Mehrsprachigkeit.

- Linacre, J. M. (1994). *Many-facet Rasch measurement* (2nd ed). MESA Press.
- Linacre, J. M. (2022). *Facets computer program for many-facet Rasch measurement* (Version 3.84.0) [Software]. Winsteps.com.
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of College-level ESL writers' language development. *TESOL Quarterly*, 45/1, 36–62.
- Lüdeling, A. (2008). Mehrdeutigkeiten und Kategorisierung: Probleme bei der Annotation von Lernerkorpora. In M. Walter & P. Grommes (Hrsg.), *Fortgeschrittene Lernervarietäten* (S. 119–140). Max Niemeyer Verlag. <https://doi.org/10.1515/9783484970342.2.119>
- Lüdeling, A., Doolittle, S., Hirschmann, H., Schmidt, K., & Walter, M. (2008). Das Lernerkorpus Falko. *Deutsch als Fremdsprache*, 45(2), 67–73.
- Lüdeling, A., & Hirschmann, H. (2015). Error annotation systems. In F. Meunier, G. Gilquin, & S. Granger (Hrsg.), *The Cambridge Handbook of Learner Corpus Research* (S. 135–158). Cambridge University Press. <https://doi.org/10.1017/CBO9781139649414.007>
- Lüdeling, A., Hirschmann, H., Shadrova, A., & Wan, S. (2021). Tiefe Analyse von Lernerkorpora. In *Tiefe Analyse von Lernerkorpora* (S. 235–284). De Gruyter. <https://doi.org/10.1515/9783110731514-013>
- Lüdeling, A., Walter, M., Kroymann, E., & Adolphs, P. (2005). Multi-level error annotation in learner corpora. *Proceedings of Corpus Linguistics*, 14–17. <https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko/pdf/FALKO-CL2005.pdf>
- Michalke, M. (2017). *koRpus: An R Package for Text Analysis* (Version 0.10-2) [Software]. reaktanz.de. <http://reaktanz.de/?c=hacking&s=koRpus>
- Peyer, E., Andexlinger, M., Kofler, K., & Lenz, P. (2016). *Projekt Fremdsprachenevaluation BKZ: Schlussbericht zu den Sprachkompetenztests*. Institut für Mehrsprachigkeit.
- R Core Team. (2022). *R: A Language and Environment for Statistical Computing* (Version 4.0.2) [Software]. R Foundation for Statistical Computing. <http://www.R-project.org>
- Schmid, H. (2013). Probabilistic Part-of-Speech Tagging Using Decision Trees. In D. B. Jones & H. Somers (Hrsg.), *New Methods in Language Processing* (S. 154–164). Routledge. <https://www.cis.lmu.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf>
- Schmidt, T., & Wörner, K. (2009). EXMARaLDA – creating, analysing and sharing spoken language corpora for pragmatic research. *Pragmatics*, 19(4), 565–582. <https://doi.org/10.1075/rag.19.4.06sch>
- Selinker, L. (1972). *Interlanguage*. 10(1–4), 209–232. <https://doi.org/10.1515/iral.1972.10.1-4.209>
- Studer, T., & Hicks, N. S. (2022). *The interplay of task variables, linguistic measures, and human ratings: Insights from the multilingual learner corpus SWIKO*. European Second Language Acquisition Conference, Fribourg.
- Ure, J. (1971). Lexical density and register differentiation. In G. E. Perren & J. L. M. Trim (Hrsg.), *Applications of linguistics* (S. 443–452). Cambridge University Press.
- Volodina, E., Bryant, C., Caines, A., De Clercq, O., Frey, J.-C., Ershova, E., Rosen, A., & Vinogradova, O. (2023). MultiGED-2023 shared task at NLP4CALL: Multilingual Grammatical Error Detection. In D. Alfter, E. Volodina, T. François, A. Jönsson, & E. Rennes (Hrsg.), *Proceedings of the 12th Workshop on NLP for Computer Assisted Language Learning* (S. 1–16). LiU Electronic Press. <https://aclanthology.org/2023.nlp4call-1.1/>
- Weiss, Z. (2020). *EXMARaLDA Converter for Time-Aligned Mult-Layer Speech Transcriptions* [Python]. LingMod Research Group. <https://github.com/lingmod-tue/exmaralda-converter> (Original work published 2020)
- Weiss, Z., Hicks, N. S., Meurers, D., & Studer, T. (2022). *Using linguistic complexity to probe into genre differences? Insights from the multilingual SWIKO learner corpus*. Learner Corpus Research Conference, Padua.
- Wisniewski, K., Schöne, K., Nicolas, L., Vettori, C., Boyd, A., Meurers, D., Abel, A., & Hana, J. (2013). MERLIN: An online trilingual learner corpus empirically grounding the European Reference Levels in authentic learner data. *ICT for Language Learning. Conference Proceedings 2013*.



# Vielen Dank für Ihre Aufmerksamkeit!



Nina Hicks  
[nina.hicks@unifr.ch](mailto:nina.hicks@unifr.ch)

Thomas Studer  
[thomas.studer@unifr.ch](mailto:thomas.studer@unifr.ch)