

Multi-Word Units in Multilingual Learners

# The MuLeCo Project

**LMU**

LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

Patricia de Crignis  
Annika Moratzky  
Cecilia Ugartemendía  
Florian Zacherl

# Outline

---



- ***Munich Learner Corpus*** (MuLeCo)
- Methodology
- Multi-Word example: DOM in Spanish (preliminary results)
- Perspectives

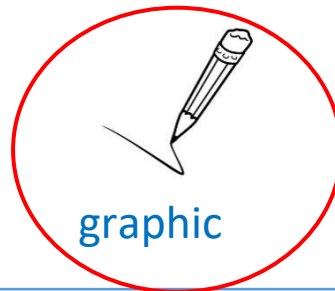
# Munich Learner Corpus (MuLeCo)

---

Error-annotated learner corpus  
& exercise platform for learners



Medium




(cf. Franke 2017, Gilquin 2022, 2023, Granger et al. 2009, Granger/Gilquin/Meunier 2015, Reznicek et al. 2012...)

# Munich Learner Corpus (MuLeCo)


## Virtual research and learning environment



digital research and learning environment



digitally available research data and learning material



computer-assisted procedures and methods for research



Romance Philology: up to now no sufficient support for L1 German

# Munich Learner Corpus (MuLeCo)

Interfaces for needs/use in classroom, learners

Linguistics

errors/level of competence?  
difficult categories?

feedback-exercises based on linguistic data

We need errors for learning processes

classroom/learners

Errors as a basis for evaluation

**Language Awareness**

awareness of difficult categories

awareness of individual weaknesses

individual, needs-oriented

**Autonomous learning:**

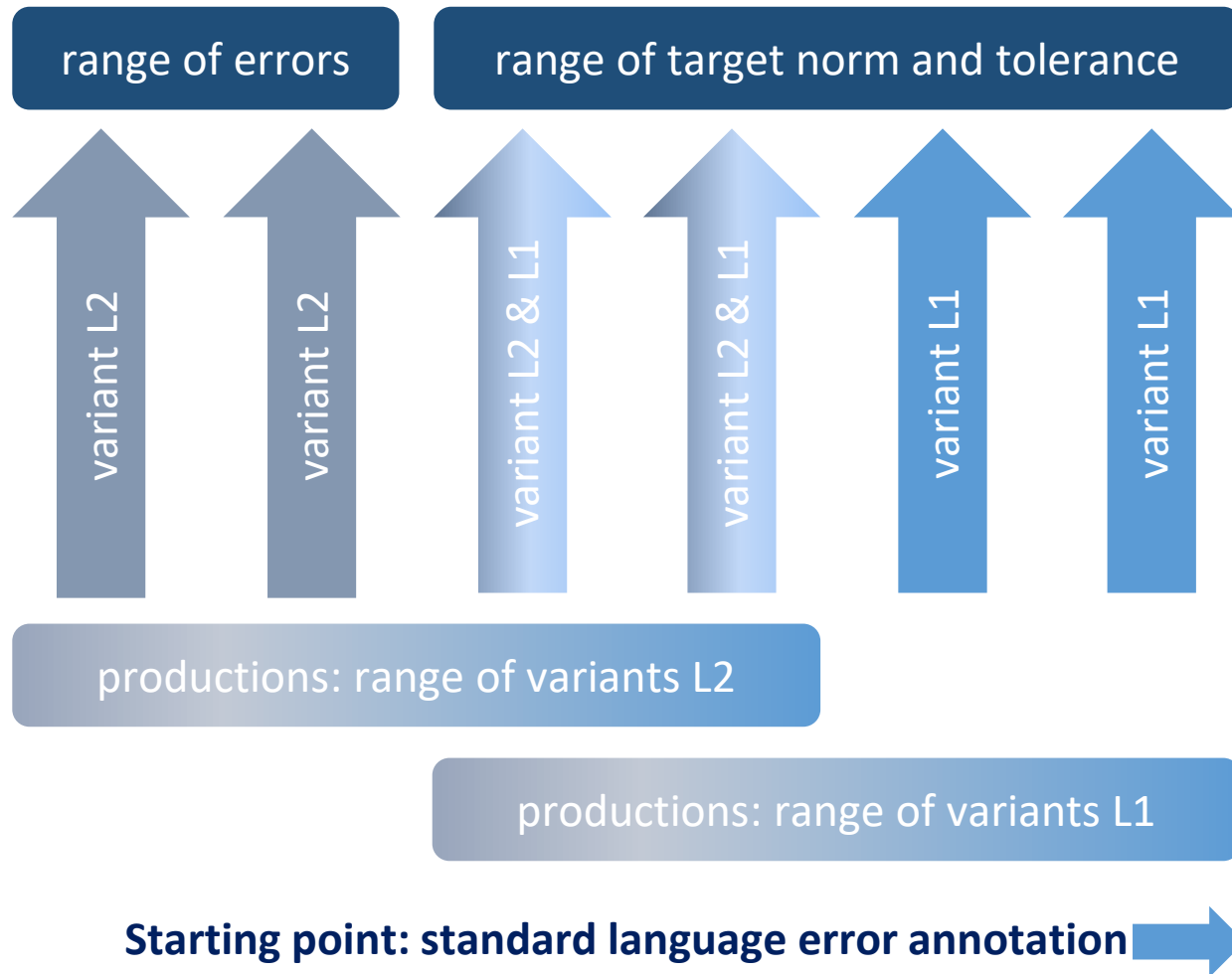
reenforcing/enhancing strategies for self-correction

something learners want/need

(cf. Schlaak 2020)

# Munich Learner Corpus (MuLeCo)

**Goal: rethinking errors**



But, what is an error?

# Munich Learner Corpus (MuLeCo)

---

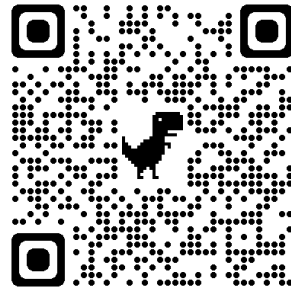
## Our team

### Informatics

#### IT-Group

#### Humanities

- Stephan Lücke
- Christian Riepl
- Florian Zacherl



Operation beyond the  
development phase  
(cf. Franke 2017)

### Linguistics & Didactics



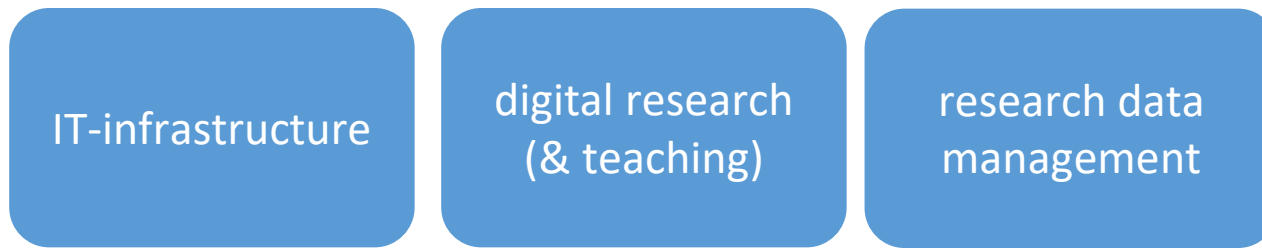
#### Romance Philology

- Patricia de Crignis
- Annika Moratzky
- Cecilia Ugartemendía
- Johanna Wolf
- (Marion Le Brun)

# Munich Learner Corpus (MuLeCo)

## IT-Gruppe Geisteswissenschaften (ITG) – LMU Center for Digital Humanities

(cf. Krefeld/Lücke  
2007, 2009, 2015)



FAKULTÄT FÜR SPRACH- UND LITERATURWISSENSCHAFTEN  
DEPARTMENT II  
ROMANISCHE PHILOLOGIE

Atlante AsiA  
sintattico della Calabria

ASD

metropol  
italia  
social language tagging

VERBA ALPINA  
Der alpine Kulturraum im Spiegel seiner Mehrsprachigkeit

GEISTES- UND KULTURWISSENSCHAFTLICHE FAKULTÄTEN  
IT-GRUPPE GEISTESWISSENSCHAFTEN

## sustainable research data management (RDM): FAIR-principles

Auffindbar  
(Findable)



Zugänglich  
(Accessible)



Interoperabel  
(Interoperable)



Wiederverwendbar  
(Reusable)





# Methodology

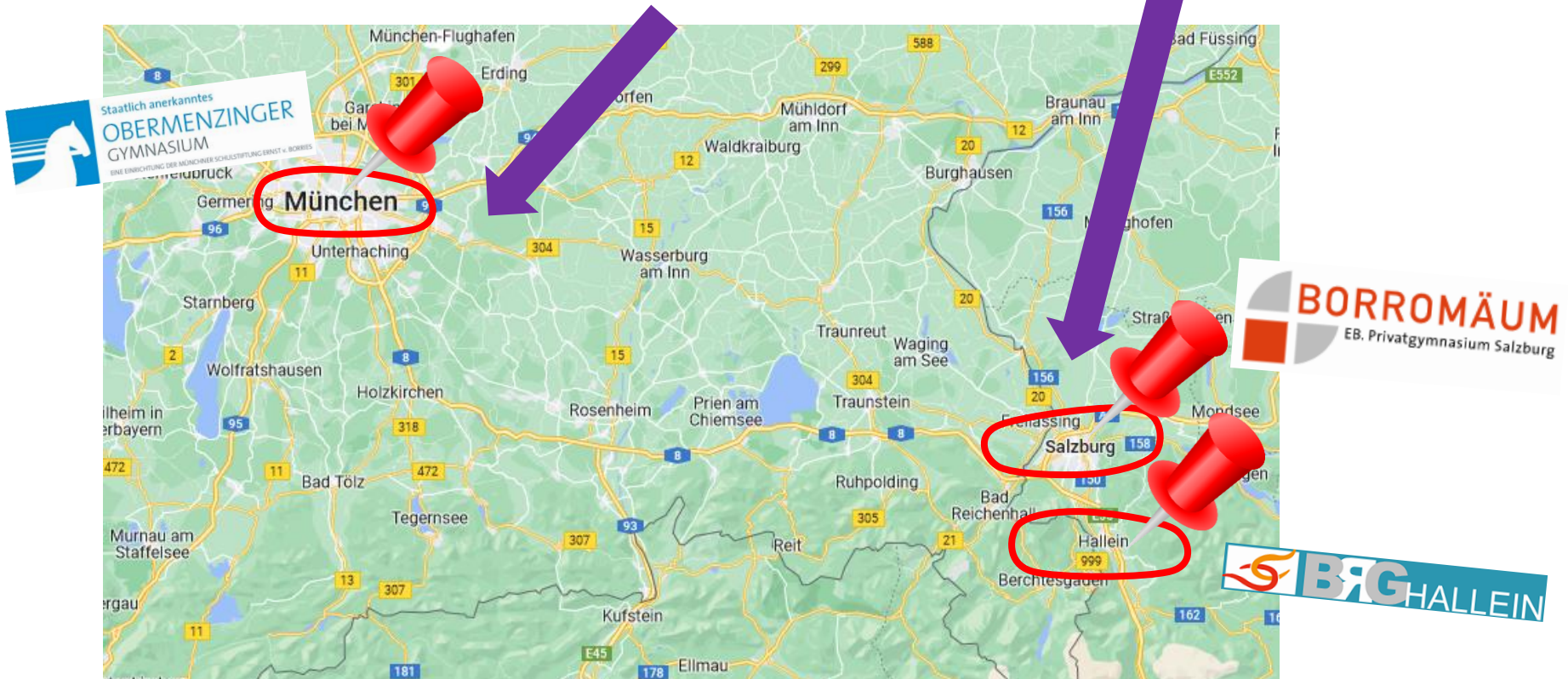
„the variety that the learners represented in the corpus are most likely to have been exposed to“ (Gilquin 2022: 9)

+ L1 SP control group Spain (Salamanca)

Where we started: **Private** schools

2 Germany: **FR**  
(Useni 2022)

1 Austria: **SP**  
(Nicolussi/Pilz 2021)



# Munich Learner Corpus (MuLeCo)

## In the schools

- Letter of consent
- Introduction to the task
- Questionnaire on the learning situation

**Fragebogen zur Lernsituation**  
(adaptiert nach Ender 2007: 229)

*Dieser Fragebogen dient zur Erhebung persönlicher Daten der Sprachlernerfahrungen des Teilnehmers bzw. der Teilnehmerin. Alle Daten werden anonym behandelt.*

Anfangsbuchstaben des Namens (Vorname und Nachname): FB

Alter: 16 Jahre

Geschlecht:  weiblich  männlich

Geburtsort (Land, Stadt): Salzburg, Salzburg

Wohnort (Land, Stadt): Salzburg, Salzburg

L1 (Erstsprache bzw. Muttersprache): Deutsch

L2 (Zweitsprache bzw. erste gelernte Fremdsprache): Englisch

L3 (Drittsprache bzw. zweite gelernte Fremdsprache): Spanisch

Weitere: \_\_\_\_\_

Welches Sprachniveau hast du in Spanisch?

A1  A2  B1  B2  C1  C2

Seit wie vielen Jahren lernst du Spanisch? (Lernjahre)

1  2  3  4  5  6 mehr: \_\_\_\_\_

Wie viele Stunden Spanischunterricht hast du in der Woche?

1  2  3  4  5  6 mehr: \_\_\_\_\_

Beschäftigst du dich mit dem Spanischen auch außerhalb deiner schulischen Verpflichtungen?

nein

ja; was genau machst du? (z.B. Filme schauen) \_\_\_\_\_; wie oft? (z. B. einmal die Woche) \_\_\_\_\_

Besondere Lernerfahrungen in Spanisch (z.B. Austauschprogramme, Sprachkurse, ...)

nein

ja (welche? \_\_\_\_\_; wie lange? \_\_\_\_\_)

Wurde bei dir jemals eine Lese- und Rechtschreibschwäche festgestellt?  ja  nein

2

# Methodology

## Data elicitation

picture story: *Frog, where are you?*

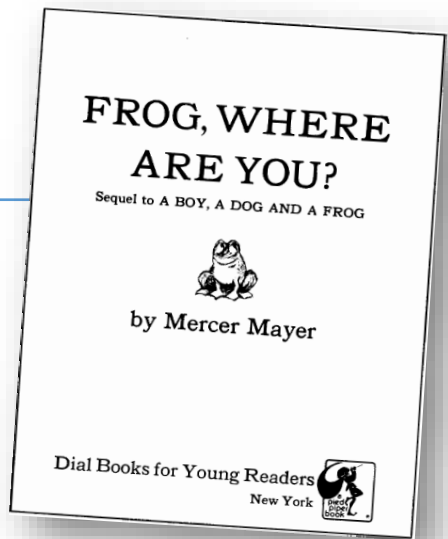


Bild 1



la rana	der Frosch
el perro	der Hund
pequeño/a	klein
fascinado/a	fasziniert
observar	beobachten

Bild 1



la grenouille	der Frosch
le chien	der Hund
petit/e	klein
observer avec fascination	fasziniert beobachten

- written production
- description of 29 pictures
- 1 sentence per picture
- use of given vocabulary (cognitive relief; cf. Walter 2004, 2007)
- working time: 45-60 minutes
- acquisition of grammar and writing skills

# Methodology

---

## Participants

### learners

	Salzburg & Hallein	München
<b>language (L2/L3)</b>	SPA	FRA
<b>students</b>	39 (age: 15-18)	67 (age: 11-17)
<b>level</b>	A1, A2, B1	A2, B1
<b>L1</b>	mostly German; Turkish, Serbo-Croatian, English, Arabic etc.	

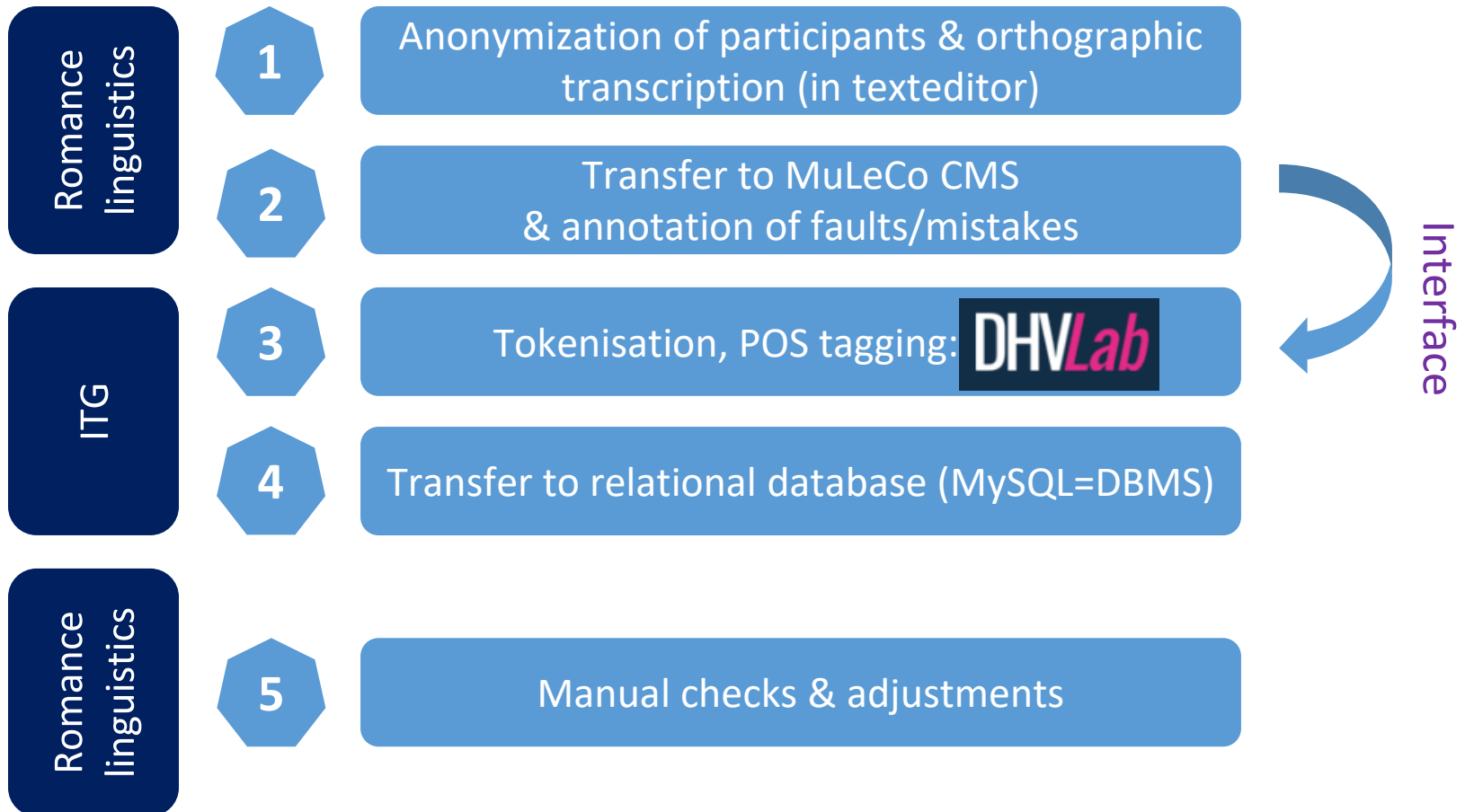
### control group

	Salamanca	<i>in progress ...</i>
<b>language (L1)</b>	SPA	FRA
<b>students</b>	29 (age: 16)	
<b>L2; L3</b>	English; French	

# Methodology

---

## Data handling



# Methodology

---

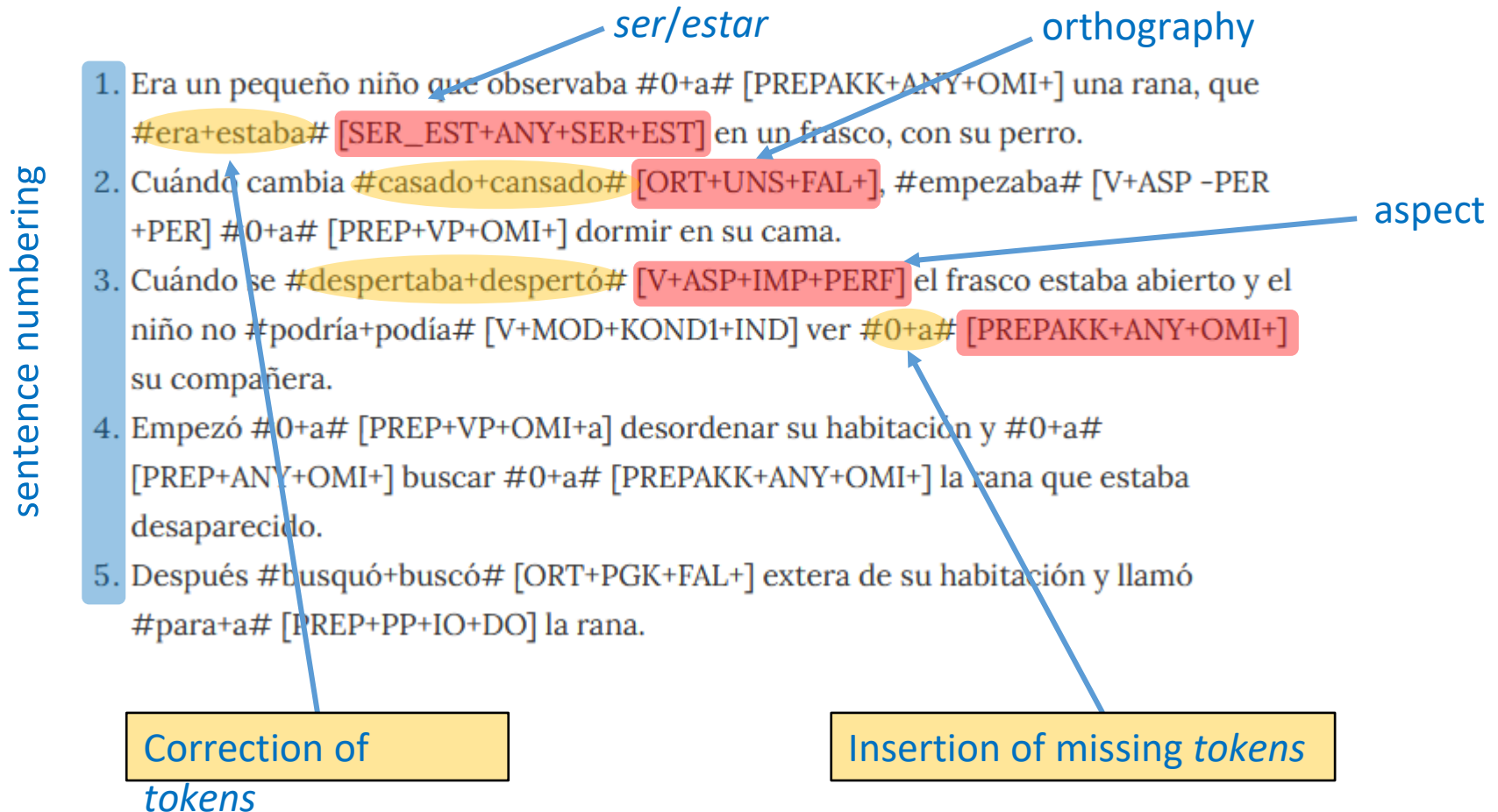
## Participants: anonymous code

	1	2	3	4	5	6	7
<i>Parameter</i>	L1 (ISO 639-1)	Mehrsprachigkeit	Sprache Produktion	Sprach- niveau	Textart	Medium	ID text
<i>Ausprägung</i>	<b>DE/TR/PL</b> ...	<b>N</b> (ein) <b>J</b> (a)	<b>ES/FR/IT</b>	<b>A/B/C</b> <b>1/2</b>	<b>B</b> (ilderge- schichte) <b>F</b> (reie) Produktion	<b>S</b> (graph.) <b>M</b> (phon.)	<b>001</b> , <b>002</b> , <b>002</b> , <b>004</b> , ...
<i>Beispiel</i>	<b>DE</b>	<b>N</b>	<b>ES</b>	<b>A2</b>	<b>B</b>	<b>S</b>	<b>001</b>

**DENESA2BS001**

# Methodology

## Error annotation (SP), an example (DENESA2BS001)



# Methodology

AND

correlation with:

Competence level, years of learning, L1/2/3, age, gender, etc.

## Relational database

### An example (DENESA2BS001)

Anonymized code

Language <small>Ländercode nach ISO839-1 (müssen manuell in sql-Sc...</small>	sigle	SatzNr	TokenNr	Token	Token_Ort <small>orthographisch korrekte Form des Tokens (sofern fe...</small>	Flag	Analyse	pos <small>Part of Speech (Worart) des Tokens; ermittelt mit Treetagger</small>	lemma <small>Lemma; ermittelt mit Treetagger</small>
es	DENESA2BS001	2	7	dormir	NULL	NULL		VLinf	dormir
es	DENESA2BS001	2	8	en	NULL	NULL		PREP	en
es	DENESA2BS001	2	9	su	NULL	NULL		PPO	suyo
es	DENESA2BS001	2	10	cama	NULL	NULL		NC	cama
es	DENESA2BS001	2	11	.	NULL	NULL		FS	.
es	DENESA2BS001	3	1	Cuándo	NULL	NULL		ADV	cuándo
es	DENESA2BS001	3	2	se	NULL	NULL		SE	se
es	DENESA2BS001	3	3	despertaba	NULL	NULL	V ASP -PER +PER	VLfin	despertar
es	DENESA2BS001	3	4	el	NULL	NULL		ART	el
es	DENESA2BS001	3	5	frasco	NULL	NULL		NC	frasco
es	DENESA2BS001	3	6	estaba	NULL	NULL		VEfin	estar
es	DENESA2BS001	3	7	abierto	NULL	NULL		ADJ	abierto

Search options

Error annotation



# Methodology – Example statistical analysis

```
SELECT
  level,
  fraction,
  total,
  ROUND(fraction / total * 100, 1) AS percent
FROM (
  SELECT
    SUBSTR(sigle,6,2) AS level,
    SUM(IF(Analyse LIKE 'V%', 1, 0)) AS
fraction,
    SUM(IF(Analyse != '', 1, 0)) AS total
  FROM texte
  GROUP BY level) sq;
```



level	fraction	total	percent
A2	355	1296	27.4
B1	220	958	23.0
B2	29	229	12.7

Percentage of verb  
related mistakes  
according to  
competence levels

# Example: DOM in Spanish

*El niño abre la ventana y llama a la rana.* (L1; VAA6)  
'The boy opens the window and calls the frog.'

## What is it about?

subject	verb	object marker	dir. object
<i>el niño</i>	<i>abre</i>		<i>la ventana</i>
'the boy'	opens		'the window'
<i>el niño</i>	<i>llama</i>	<i>a</i>	<i>la rana</i>
'the boy'	calls	TO	'the frog'

- **linguistics**  
complex interaction of two continua (animacy & specificity)  
[?] subject-like dir. objects [?] a-marking & variation
- **didactics (A1)**  
discrete categories (animacy)  
[?] animate reference: a-marking  
[?] inanimate reference: unmarking (∅)



# Example: DOM in Spanish



Does school grammar fit the 'speakers reality'?

DOs (learners): 357	
animate reference	inanimate reference
61 % (216)	39 % (141)

<b>50 % (109)</b> a-marking	50 % (107) ∅	0 % (0) a-marking	<b>100 % (141)</b> ∅
error annotation			

DOs (L1 control group): 275			
animate reference		inanimate reference	
63 % (174)		37 % (101)	
a-marking	∅	a-marking	∅
<b>80 % (139)</b>	<b>20 % (35)</b>	4 % (4)	<b>96 % (97)</b>



When unmark animate DOs?

# Example: DOM in Spanish

---

What did the students do?

animate reference		inanimate reference	
61 % (216 )		39 % (141)	
a-marking	∅	a-marking	∅
10 % (21 )	<b>90 % (195)</b>	1 % (1)	<b>99 % (140)</b>



unmarking is predominant  
[?] awareness?

# Example: DOM in Spanish

How many students use DOM?

A2	B1
17 % (5/29)	30 % (3/10)



unknown grammar rule for many

Do all L1 speakers use DOM in the same way?



tolerance range  
& progression



SDH9

animate (human)	animate (animal)		inanimate
(3/3)	7/13	6/13	(11/11)
100 % a-marking	54 % a-marking	46 % ∅	100 % ∅

VAA6

animate (human)	animate (animal)		inanimate
(5/5)	7/8	1/8	5/5)
100 % a-marking	87,5 % a-marking	12,5 % ∅	100 % ∅

# Perspectives

---

## **DOM in Spanish (MuLeCo)**

- definition of ‘error’ (L1 error annotation vs. L1 control group)
- lack of awareness (A2 & B1)
- L1: variation within animate DOs ☐ tolerance range
- didactics: unmarking of animate DOs, e.g. special verbs (*tener, haber*)

## **MuLeCo**

### Things to do

- more data
- Italian
- spoken language etc.

### Problems

- syntactic annotation
- word order and/or position errors?
- annotation of transfer phenomena? (= interpretation of data vs. invisibility; CLI as a specific problem area of learners)

---

Thank you!

# Bibliography

---

Bossong, Georg (1982): „Historische Sprachwissenschaft und empirische Universalienforschung“, *Romanistisches Jahrbuch* 33, 17-51.

Bossong, Georg (1998): „Le marquage différentiel de l’objet dans les langues d’Europe“, Feuillet, Jack (ed.), *Actance et valence dans les langues d’Europe*, Berlin/New York: de Gruyter, 193-258.

Delbecque, N. (2002). A construction grammar approach to transitivity in Spanish. In: *The nominative and accusative and their counterparts*, (81-130). Amsterdam: Benjamins.

Fábregas, Antonio (2013): „Differential Object Marking in Spanish: State of the Art“, *Borealis: An International Journal of Hispanic Linguistics* 2 (2), 1-80.

Franke, Michael (2017): „Erfolgsfaktoren für virtuelle Forschungsumgebungen“, in: Seng, Eva-Maria/Keil, Reinhard/Oevel, Gudrun (Eds.): *Kooperative Forschungsumgebungen in den eHumanities*, Berlin, Boston: De Gruyter, 57-65.

Gilquin, G. 2022. ‘One norm to rule them all? Corpus-derived norms in learner corpus research and foreign language teaching’. *Language Teaching* 55.1: 87-99.

Gilquin, G. 2023. ‘Written learner corpora to inform teaching’. In R.R. Jablonkai & E. Csomay (eds) *The Routledge Handbook of Corpora and English Language Teaching and Learning* (pp. 281-295). London: Routledge.

Gómez Torrego, Leonardo (192011): *Gramática didáctica del español*. Madrid: Ediciones SM.

Krefeld, Thomas/Lücke, Stephan: (seit 2007): Atlante sintattico dell'italiano meridionale (ASICA), <http://www.asica.gwi.uni-muenchen.de>

Krefeld, Thomas/Lücke, Stephan: (seit 2009): Audioatlas siebenbürgisch-sächsischer Dialekte (ASD), <http://www.asd.gwi.uni-muenchen.de/?projektinfo=true>

Krefeld, Thomas/Lücke, Stephan: (seit 2015): VerbaAlpina (VA), <http://www.verba-alpina.gwi.uni-muenchen.de>

Granger, S., Dagneaux, E., Meunier, F. and Paquot, M. (2009). *The International Corpus of Learner English. Version 2*. Handbook and CD-ROM. Louvain-la-neuve: Presses universitaires de Louvain.

Granger, S., Gilquin, G., & Meunier, F. (Eds.). (2015). *The Cambridge Handbook of Learner Corpus Research* (Cambridge Handbooks in Language and Linguistics). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139649414

Pomino, Natascha (2012). Der präpositionale Akkusativ. *Handbuch Spanisch. Sprache, Literatur, Kultur, Geschichte in Spanien und Hispanoamerika*. Berlin: Erich Schmidt Verlag, 307-313.

Prévost, P. (1997): *Truncation in second language acquisition*. Dissertation McGill University, Montreal, Quebec.

Radatz, Hans-Ingo (2021): *Spanische Grammatik im Fokus. Klassische Beschreibungsprobleme aus neuer Sicht*. Berlin, Boston: de Gruyter.

Reznicek, Mark/Lüdeling, Anke/Krummes Cedric/Schwantuschke, Franziska/Walter, Maik/Schmidt, Karin/Hirschmann, Hagen/Andreas, Torsten (2012): *Das Falko-Handbuch. Korpusaufbau und Annotationen (Version 2.01)*, Berlin: HU.

Schlaak, Claudia (2020): „Fehlertoleranz bei und von Fremdsprachenlernern im ersten Lernjahr“, in: Veldre-Gerner, Georgia/Ossenkop, Christina (Eds.): *Fehler-Abweichung-Variation*, Frankfurt am Main: Lang, 209-224.

Pustka, Elissa (2022): „Feldforschung in der Schule – ein Bericht aus der Forschungspraxis“, in: *Zeitschrift für Romanische Sprachen und ihre Didaktik* 16.1, 89-122.