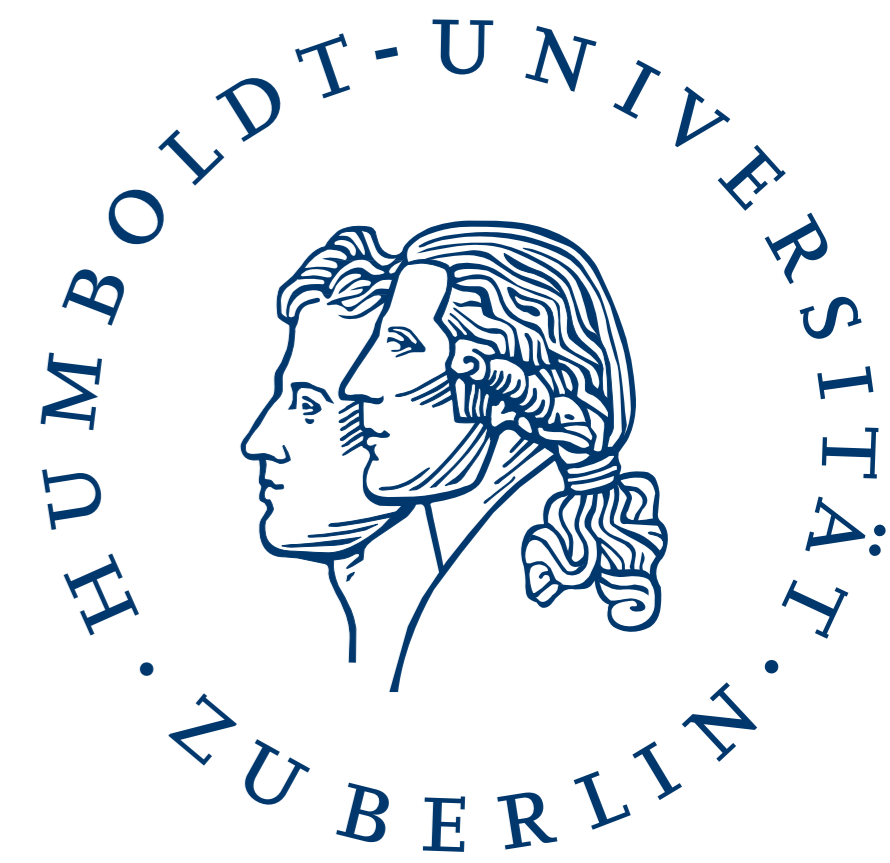


Repair disfluencies in German native and non-native speech



Institut für deutsche
Sprache und Linguistik,
Humboldt-Universität zu Berlin,
<https://u.hu-berlin.de/belz>

Malte Belz

1.1 Background

- Repair disfluencies are disfluencies within self-repairs.
 - Disfluencies: unfilled pauses and filled pauses.
- Advanced non-native speakers (L2) still show different disfluency behaviours than native speakers (L1). (cf. Declerck & Kormos 2012, Belz & Klapi 2013)
- No systematic studies comparing German L1 and L2 data so far.

2.1 Terminology

- Repairs consist of (cf. Shriberg 1994)
 - a **reparandum (RD)** – the utterance to be repaired.
 - an optional **interregnum (IR)** – the temporal region between RD and RS.
 - a **reparans (RS)** – the repairing utterance.
- Tokens in the RS are classified into subrepair categories
 - **repetitions (r)**.
 - **substitutions (s)**.
 - **insertions (i)**.
- Subrepair categories **r** and **s** will later be merged into one category **rs**.

Examples

taken from
BeMaTaC.L1.2013-02 and
BeMaTaC.L2.2013-02 at
<https://u.hu-berlin.de/annis3>

L1	nach 0.8s links to the 0.8s left	0.5s 0.5s	waagrecht horizontally	nach links to the left				
	RD	IR		RS				
					i	r	r	

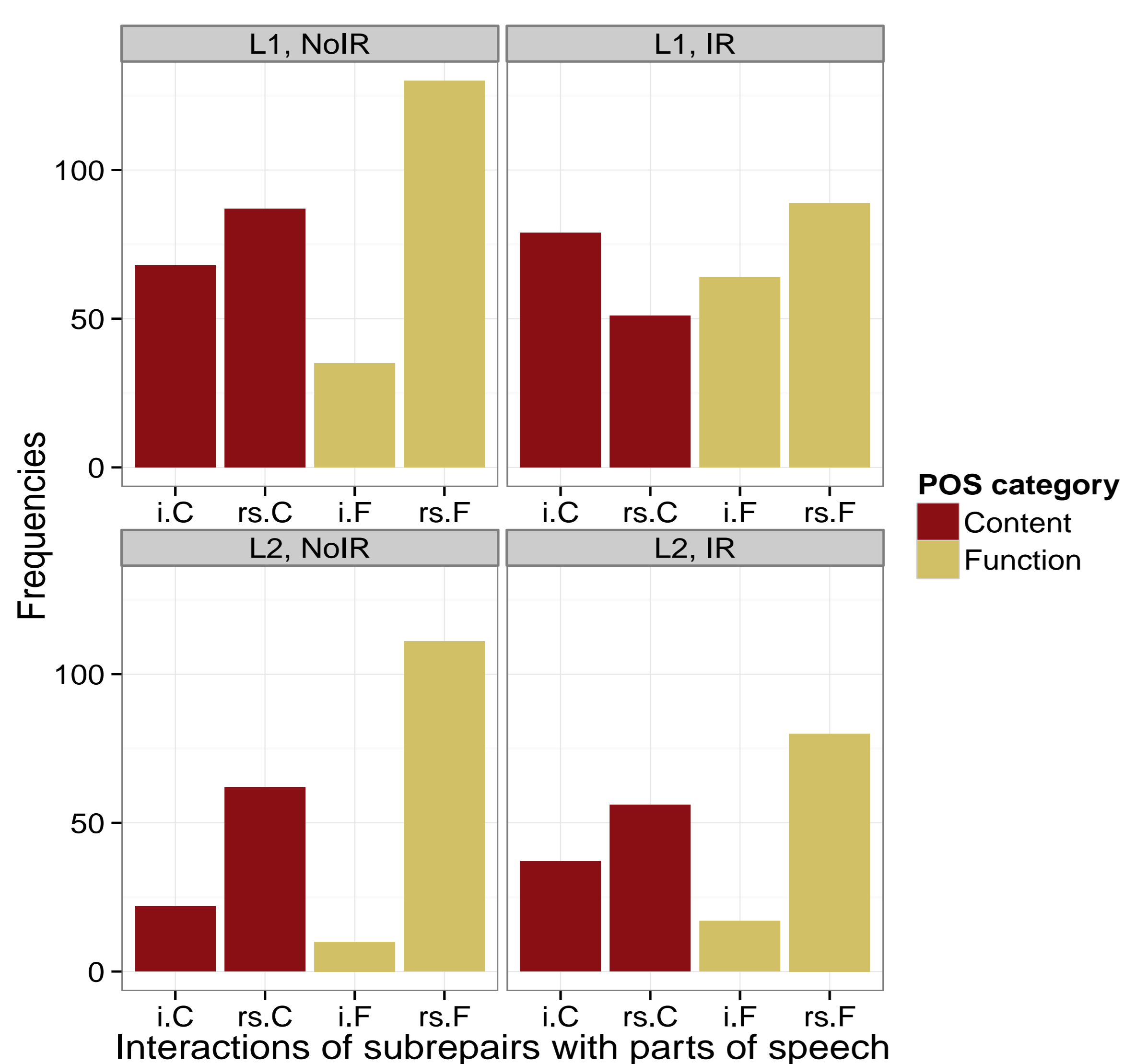
L2	linken Ecke left corner	0.4s 0.4s	unteren bottom	linken left	Ecke corner			
	RD	IR		RS				
						i	r	r

L1	geh/ go	gehst go						
	RD	RS						
							s	

L2	ich hab I have	du hast you have						
	RD	RS						
							s	s

3.2 Analysis by distributions

- Conditions L1/L1 and IR/No IR.
- Interactions of **rs** and **i** with **content words (C)** and **function words (F)**.



- ⇒ L1 speakers tend to repair content words that are repeated or substituted without using IR and content words that are inserted together with IR ($\chi^2 = 7.4$; $df = 1$; $p < 0.01$).
- ⇒ L2 with IR = L1 without IR (n.s.).

References

Belz, M. & M. Klapi (2013) Pauses following Fillers in L1 and L2 German Map Task Dialogues. In R. Eklund, ed., *Proceedings of DISS 2013: The 6th Workshop on Disfluency in Spontaneous Speech*, 9–12. Declerck, M. & J. Kormos (2012) The effect of dual task demands and proficiency on second language speech production. *Bilingualism: Language and Cognition* 15 (4): 782–796. Sauer, S. & A. Lüdeling (2013) BeMaTaC: A Flexible Multilayer Spoken Dialogue Corpus for Contrastive SLA Analyses. In *ICAME 34*, 46–47. Shriberg, E. E. (1994) *Preliminaries to a Theory of Speech Disfluencies*. Unpublished Ph.D. thesis, University of California, Berkeley. Schmidt, T. & K. Wörner (2009) EXMARaLDA – Creating, analysing and sharing spoken language corpora for pragmatic research. *Pragmatics* 19 (4): 565–582. Zeldes, A., J. Ritz, A. Lüdeling & C. Chiarcos (2009) ANNIS: A Search Tool for Multi-Layer Annotated Corpora. In *Proceedings of Corpus Linguistics*.



1.2 Questions

- How are repair disfluencies influenced by their surroundings, namely reparandum and reparans?
- Are there differences between L1 and advanced L2 speakers?
- Do subrepair categories, like insertions, repetitions or substitutions influence repair disfluencies?
- Are parts of speech influencing repair disfluencies?

2.2 Method

- L1 & L2 spontaneous speech: Berlin Map Task Corpus (BeMaTaC) (Sauer & Lüdeling 2013)
 - instructors describe a map with landmarks to instructees.
- German L2 speakers beyond C1 level (Common European Framework of Reference for Languages).
- Annotation of repair instances with EXMARaLDA (Schmidt & Wörner 2009)
 - Repair tier with repair frame: **RD – IR – RS**.
 - Subrepair tier with subrepairs within RS: **r, s, i**.
- Query and export via ANNIS (Zeldes et al. 2009).
- Distributional and multivariate analysis.

3.1 Results

	No IR (%)	IR (%)	Repairs (%)	Tokens	Duration	Dialogues	Subjects
L1	141 (0.59)	98 (0.41)	239 (0.02)	11.192	66min	12	16
L2	148 (0.58)	109 (0.42)	257 (0.01)	21.330	77min	5	6

- ⇒ The frequency distribution of L1 and L2 repairs with IR does not deviate significantly from the expected one ($\chi^2 = 0.1$; $df = 1$; $p = 0.75$).
- ⇒ The distribution of the subrepair relations **i**, **r** and **s** differs significantly for L1 and L2 ($\chi^2 = 45.8$; $df = 2$; $p < 0.001$).

3.3 Analysis by linear mixed-effects model

This model takes speaker specific variation into account.

- No effect is found for differences between L1 and L2.
- Subrepair variant **rs** only significant predictor for IR occurrence (Estimate -0.74, Std.Error 0.15, z value -4.78, $p < 0.001$).
 - ⇒ **rs** tends not to be preceded by an IR.
- ⇒ For L1 and L2 speakers, **insertions** tend to be preceded by an **interregnum**.

Conclusion

- Advanced L2 speakers produce more disfluencies when paralleling L1 repair patterns.
- Speaker specific variation shows **no difference between L1 and L2**.
- **Subrepair phenomena** may influence the utterance of interregna.
 - ⇒ It seems that the ease of planning repetitions and substitutions on the one hand and the difficulty of planning insertions on the other hand are mirrored by the nonexistence or existence, as the case may be, of an IR for both L1 and L2.