

# Richtlinien zur Annotation von Reparaturen in BeMaTaC

Malte Belz

Humboldt-Universität zu Berlin

18. März 2014

# 1 Vorbemerkungen

Diese Richtlinien sind im Rahmen meiner Masterarbeit über Disfluencies und Reparaturen bei Muttersprachlern und Lernern entstanden (Belz 2013). Sie dienen als Entscheidungshilfe zur Annotation von Reparaturen, die bei Schegloff et al. (1977) „selbstinitiierte Selbstreparaturen“ heißen. Die Reparaturebenen im Berlin Map Task Corpus (BeMaTaC)<sup>1</sup> können durch diese Richtlinien nachverfolgt und konsistent erweitert werden. Das verwendete Schema orientiert sich vor allem an Shriberg (1994). Dabei gibt zwei mögliche Arten zu reparieren: direkt oder indirekt. Direkte Reparaturen enthalten einen linken Teil, das Reparandum, und einen rechten Teil, das Reparans. Bei indirekten Reparaturen wird ein dritter Teil eingeschoben, das Interregnum, welches Disfluencies enthält.

- Zwei Reparaturschemata:
  - Direkte Reparatur: Reparandum – Reparans
  - Indirekte Reparatur: Reparandum – Interregnum – Reparans

Eine weitere Spezifizierung erfolgt für das Reparans, in welchem die Art der Reparatur für jedes Token annotiert wird. Dies nenne ich Subreparaturen. Hier folge ich leicht modifizierend Fox & Jasperson (1995) in der Klassifikation.

- Mögliche Subreparaturen im Reparans:
  - Repetition (Wiederholung eines Tokens)
  - Substitution (Ersetzung eines Tokens)
  - Insertion (Einfügung eines Tokens)

Abweichend zu den Annotationsrichtlinien in Belz (2013: 73) sind Löschungen im Reparandum nicht mehr beschrieben. Im Zuge der Datenanalyse hatte sich herausgestellt, dass solche Phänomene in den Korpusdaten sehr niedrigfrequent sind.

---

<sup>1</sup><http://u.hu-berlin.de/bematac>

## 2 Annotationsrichtlinien

Reparaturen	Beschreibung	Tags
instructor_repair instructee_repair	Ebene für die Annotation von Reparandum, Interregnum, Reparans. Bezugsebene für die Segmentierung ist ⟨instructor_dipl⟩ bzw. ⟨instructee_dipl⟩.	rd, ir, rs
Reparandum	Der Teil der Aussage, der vom Sprecher später repariert wird. Meist folgt danach ein Abbruch, eine ungefüllte Pause oder eine gefüllte Pause. Wenn <i>rd</i> annotiert wird, muss auch <i>rs</i> annotiert werden. Spannenannotation.	rd
Interregnum (fakultativ!)	Interregnum ist das, was zwischen Reparandum und Reparans steht, meist Stille (vgl. Ebene ⟨break⟩), gefüllte Pausen (vgl. Ebene ⟨df⟩), explizite Editierungen (bspw. <i>oder</i> , <i>beziehungsweise</i> , <i>Quatsch</i> , <i>ich meine</i> , Bsp. 4). Mehrere vorhandenen Phänomene bilden auch nur ein Interregnum (Bsp. 1, 4). Wird <i>ir</i> annotiert, muss unmittelbar davor ein <i>rd</i> und unmittelbar danach ein <i>rs</i> stehen. Spannenannotation.	ir
Reparans	Die Reparatur der Aussage. Ihr geht entweder ein Reparandum voraus oder ein Interregnum. Die Reparatur hört dort auf, wo die reparierte Sinneinheit (bspw. Silbe, Wort, Phrase) zu Ende ist oder eine neue Äußerung beginnt. Dies ist v. a. bei Insertionen nicht immer eindeutig (s. Insertion). Ist kein Interregnum vorhanden, beginnt <i>rs</i> unmittelbar nach <i>rd</i> (Bsp. 4). Spannenannotation.	rs
Subreparaturen		
instructor_subrepair instructee_subrepair	Ebene für die Annotation von konkreten Reparaturphänomenen. Bezugsebene für die Segmentierung ist ⟨instructor_dipl⟩ bzw. ⟨instructee_dipl⟩.	d⟨n⟩, i⟨n⟩, r⟨n⟩, s⟨n⟩
Insertion (Einfügung)	Nur im Reparans zu annotieren. Damit wird getaggt, was im Reparandum noch nicht vorhanden war, sondern neu hinzukam. Im Zweifelsfall muss man untersuchen, ob die fragliche Insertion ohne die Reparatur auch geäußert worden wäre. Falls ja, ist sie nicht mehr Teil des Reparans und wird nicht annotiert. Tokenannotation. n steht für eine Zahl von 1 bis n, also i1, i2 etc.	i1, i2, ...

Substitution (Ersetzung)	Nur im Reparans zu annotieren. Damit wird getaggt, was im Reparans in Bezug auf das Reparandum ersetzt worden ist. Sobald das neue Token nicht mehr exakt das alte Token darstellt, ist es eine Substitution (Bsp. 1). Tokenannotation. n steht für eine Zahl von 1 bis n, also s1, s2 etc.	s1, s2, ...
Repetition (Wiederholung)	Nur im Reparans zu annotieren. Damit wird getaggt, was im Reparans in Bezug auf das Reparandum gleich blieb. Wiederholung von identischen Tokens oder Wiederholung mit Verkürzung des Tokens (Bsp. 1). Tokenannotation. n steht für eine Zahl von 1 bis n, also r1, r2 etc. (Bsp. 6).	r1, r2, ...
Residuen „Weiß-nicht-Tag“	Wird vergeben, wenn man sich zwischen den Subreparaturtags nicht entscheiden kann. Tokenannotation.	x

#### Anmerkungen:

- Jede Ebene `_repair` bzw. `_subrep` kann zu `_repair2` bzw. `_subrep2` verdoppelt werden, um verschachtelten Reparaturen besser Rechnung zu tragen.
- Die Annotationen der Laufziffer n stellt eine zusätzliche Information dar. Sie auszuwerten bedeutet jedoch einen nicht unerheblichen Aufwand. Je nach Forschungsfrage bleibt zu überlegen, ob die Annotation dieser Laufziffer nicht als fakultativ betrachtet werden sollte.

## Beispiele

### 1. Beispiel

<b>instructor_dipl</b>	dann	schlägst	ähm		schä/	schlägst	du	sozusagen
<b>break</b>				1,5				
<b>instructor_repair</b>		rd		ir		rs		
<b>instructor_subrep</b>					x	r1	i1	

Das Reparans besteht hier aus der Auflösung der Enklise *schlägst* und der Fehlaussprache *schä/* während dieses Prozesses. Der Diskursmarkierer *sozusagen* zählt nicht mit ins Reparans, er hätte auch zusammen mit der Enklise realisiert werden können.

### 2. Beispiel

<b>instructor_dipl</b>	des	Kai/	des	Kaiserbrötchens
<b>instructor_repair</b>		rd		rs
<b>instructor_subrep</b>			r1	s1

### 3. Beispiel

<b>instructor_dipl</b>	du	gehst	sozusagen	einen	ähm	in	einem	Fünfundvierziggradwinkel
<b>instructor_repair</b>				rd	ir		rs	
<b>instructor_subrep</b>						i1	s1	

Hier endet das Reparans bei *einem*, da der Kasus repariert wurde (*einen/Akk* -> *in einem/PP*).

### 4. Beispiel

<b>instructor_dipl</b>	und	machst	rechts	oder	beziehungsweise	gehst	rechts	herum
<b>instructor_repair</b>			rd		ir		rs	
<b>instructor_subrep</b>						s1	r1	i1

Die VP *machst rechts* kann auch alleine stehen. *herum* ist nicht zwingend schon mental geplant worden, daher zählt es als Insertion mit ins Reparans.

5. Beispiel

<b>instructor_dipl</b>		und		gehst		äh		grade/		grad/		in		gerader		Linie		nach		rechts
<b>instructor_repair</b>								rd						rs						
<b>instructor_subrep</b>												i1		s1		i2				

*grade/* und *grad/* werden schnell hintereinander abgebrochen und bilden ein gemeinsames Reparaturandum. Sie werden durch eine Adverbialphrase ersetzt, daher reicht das Reparans von *in* bis *Linie*.

6. Beispiel

<b>instructor_dipl</b>		einen		runden		Kreis		um		den		Zah/		um		den		Arzt		herum
<b>instructor_repair</b>								rd						rs						
<b>instructor_subrep</b>														r1		r2		s1		

Hier ist die Reparatur eigentlich auf das Nomen begrenzt, doch der Sprecher ersetzt die ganze PP *um den Zah/* mit *um den Arzt*. *herum* zählt nicht mit in das Reparans, da *um* und *herum* eine feste Zirkumfigierung bilden und die Reparatur sich innerhalb dieser abspielt.

## Literatur

- Belz, Malte (2013): Disfluencies und Reparaturen bei Muttersprachlern und Lernern – eine kontrastive Analyse. Masterarbeit. Humboldt-Universität zu Berlin. URL: <https://edoc.hu-berlin.de/docviews/abstract.php?lang=ger%5C&id=40482>.
- Fox, Barbara A. & Jasperson, Robert (1995): A Syntactic Exploration of Repair in English Conversation. In: Davis, Philip W. (Hrsg.): *Alternative linguistics*. (= Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory, Bd. 102). Amsterdam, Philadelphia: J. Benjamins. 77–134.
- Schegloff, Emanuel A.; Jefferson, Gail & Sacks, Harvey (1977): The preference for self-correction in the organization of repair in conversation. In: *Language* 53(2). 361–382.
- Shriberg, Elizabeth E. (1994): Preliminaries to a Theory of Speech Disfluencies. Unveröffentlichte Dissertation. Berkeley: University of California.