

Bachelorarbeit

Analyse und Korrektur von Disfluenzen in gesprochener Sprache

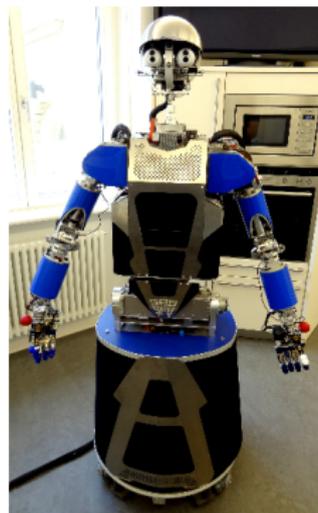
Robert Hochweiß

Betreut von Sebastian Weigelt und Tobias Hey

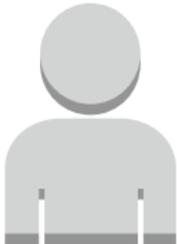
IPD Tichy, Fakultät für Informatik



Motivation



Motivation

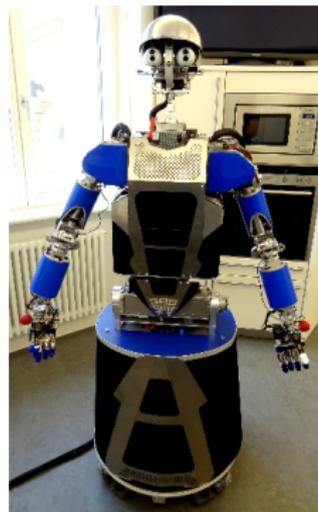


Go to go to the table





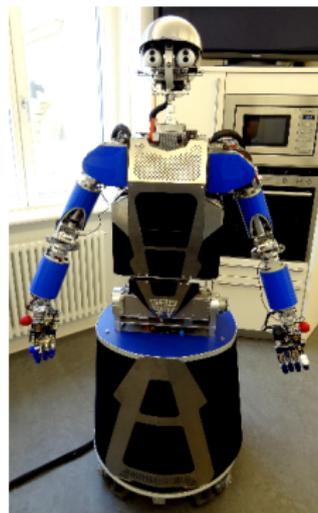
Armar take the red cup
uh the green cup



■ Disfluenzen erschweren Verarbeitung der Äußerungen



Armar take the red cup
uh the green cup



- **Disfluenzen** erschweren Verarbeitung der Äußerungen

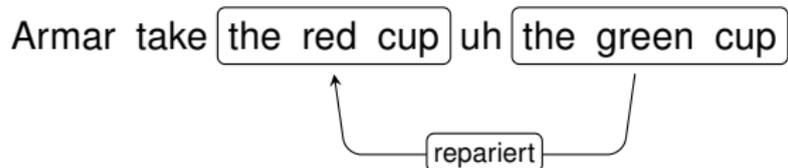
- Ziel: Disfluenzen erkennen und auflösen
 - Detektion
 - Analyse
 - Korrektur
- Kontext: Projekt PARSE

- Ziel: Disfluenzen erkennen und auflösen
 - Detektion
 - Analyse
 - Korrektur
- Kontext: Projekt PARSE

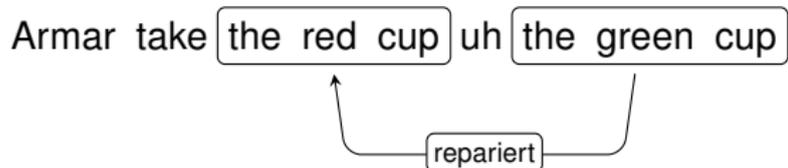
Armar take the red cup uh the green cup



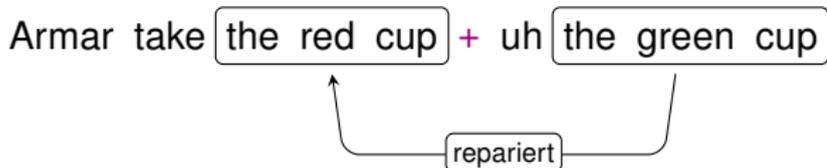
■ Struktur einer Disfluenz



■ Struktur einer Disfluenz

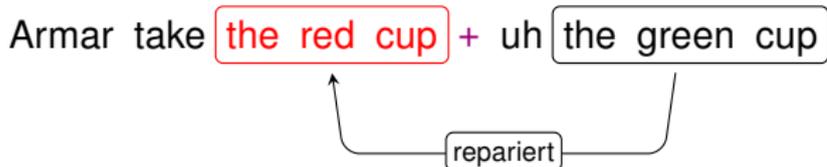


■ Struktur einer Disfluenz



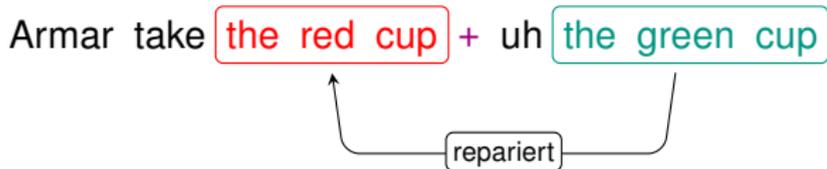
■ Struktur einer Disfluenz

- + : Unterbrechungspunkt



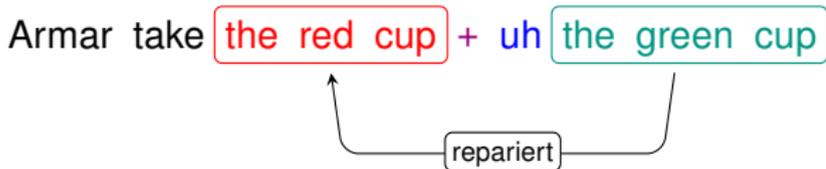
■ Struktur einer Disfluenz

- + : Unterbrechungspunkt
- **Reparandum** : Fehlerhafter Bestandteil



■ Struktur einer Disfluenz

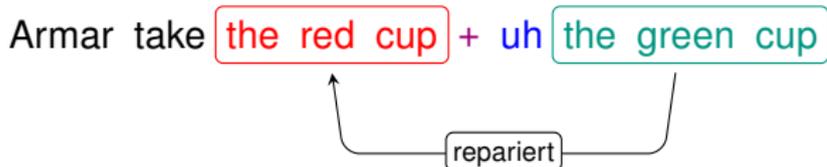
- + : Unterbrechungspunkt
- **Reparandum** : Fehlerhafter Bestandteil
- **Reparans** : Reparierender Bestandteil



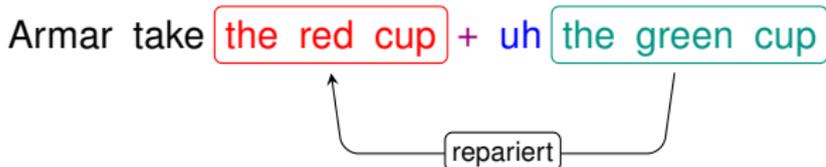
■ Struktur einer Disfluenz

- + : Unterbrechungspunkt
- **Reparandum** : Fehlerhafter Bestandteil
- **Reparans** : Reparierender Bestandteil
- **Interregnum** : Zwischen Reparandum und Reparans

Verfeinerung der Zielsetzung

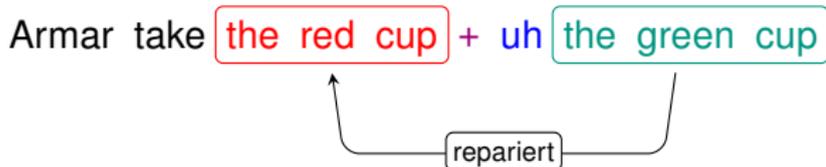


- **Detektion:**
Unterbrechungspunkt ?
- **Analyse:**
Verzögerung oder Reparatur ?
- **Korrektur**
 - Zuweisung Wörter zu Bestandteile
 - Zuweisung Reparans zu Reparandum



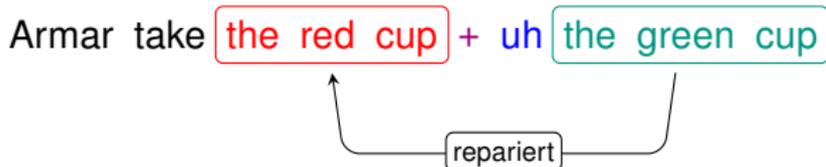
- **Detektion:**
Unterbrechungspunkt ?
- **Analyse:**
Verzögerung oder Reparatur ?
- **Korrektur**
 - Zuweisung Wörter zu Bestandteile
 - Zuweisung Reparans zu Reparandum

Verfeinerung der Zielsetzung



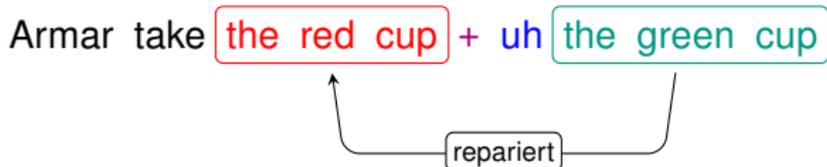
- **Detektion:**
Unterbrechungspunkt ?
- **Analyse:**
Verzögerung oder Reparatur ?
- **Korrektur**
 - Zuweisung Wörter zu Bestandteile
 - Zuweisung Reparans zu Reparandum

Verfeinerung der Zielsetzung

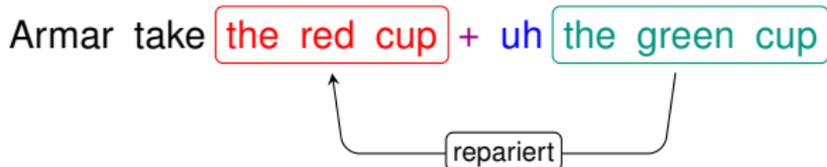


- **Detektion:**
Unterbrechungspunkt ?
- **Analyse:**
Verzögerung oder Reparatur ?
- **Korrektur**
 - Zuweisung Wörter zu Bestandteile
 - Zuweisung Reparans zu Reparandum

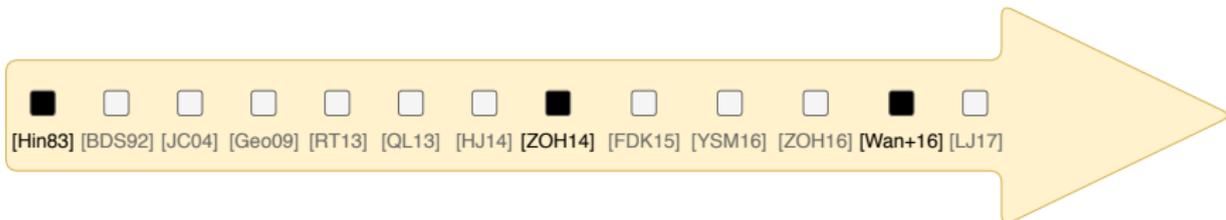
Verfeinerung der Zielsetzung

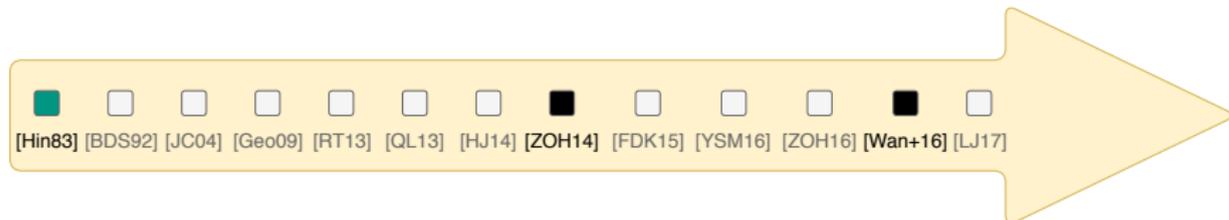


- **Detektion:**
Unterbrechungspunkt ?
- **Analyse:**
Verzögerung oder Reparatur ?
- **Korrektur**
 - Zuweisung Wörter zu Bestandteile
 - Zuweisung Reparans zu Reparandum

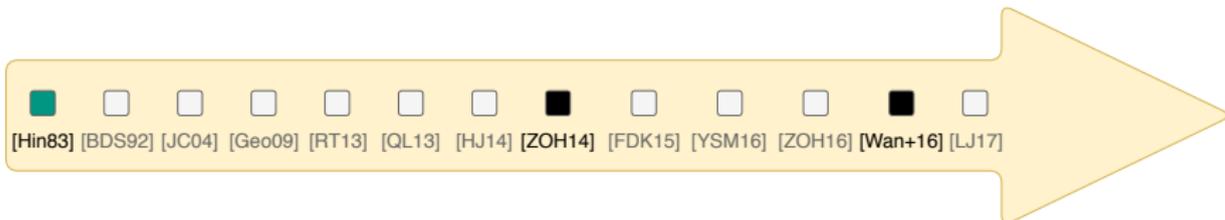


- **Detektion:**
Unterbrechungspunkt ?
- **Analyse:**
Verzögerung oder Reparatur ?
- **Korrektur**
 - Zuweisung Wörter zu Bestandteile
 - Zuweisung Reparans zu Reparandum



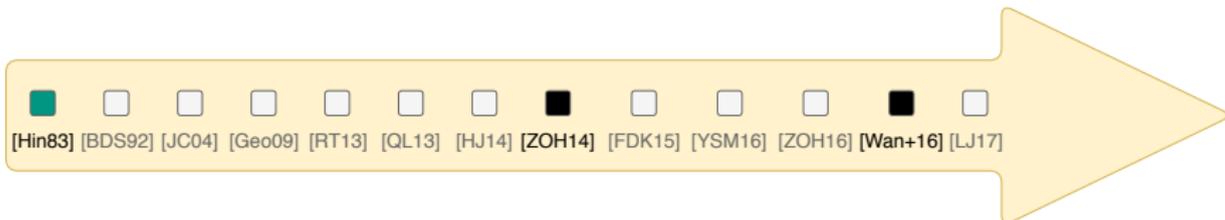


- Regelbasiert + Bearbeitungssignal [Hin83]
 - Überprüft:
Wiederholungen/Ähnlichkeiten von Wörter und Wortgruppen
 - Aber: Unterbrechungspunkt vorgegeben

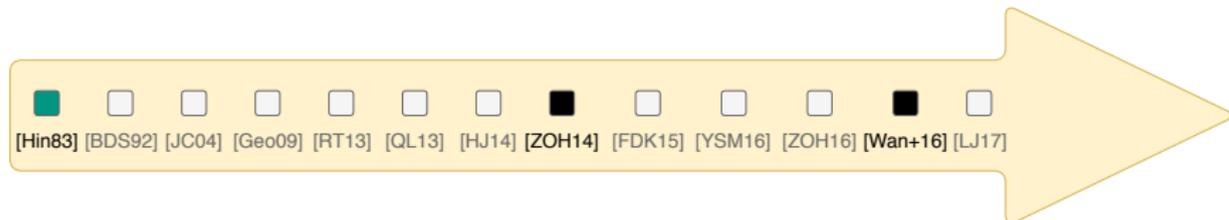


■ Regelbasiert + Bearbeitungssignal [Hin83]

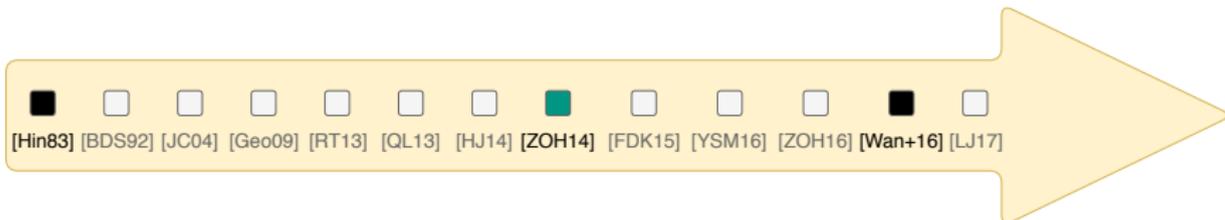
- Überprüft:
Wiederholungen/Ähnlichkeiten von Wörter und Wortgruppen
- Aber: Unterbrechungspunkt vorgegeben



- Regelbasiert + Bearbeitungssignal [Hin83]
 - Überprüft:
Wiederholungen/Ähnlichkeiten von Wörter und Wortgruppen
 - Aber: Unterbrechungspunkt vorgegeben

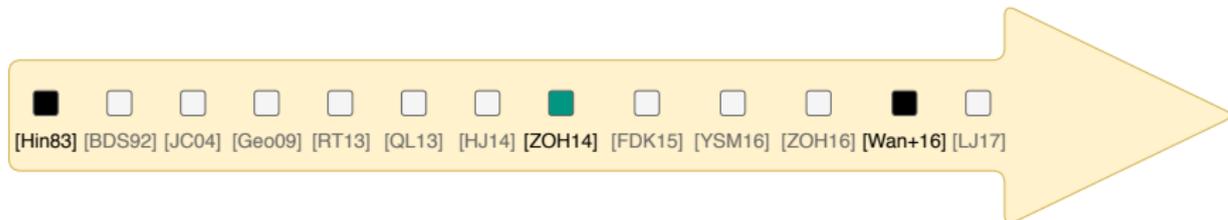


- Regelbasiert + Bearbeitungssignal [Hin83]
 - Überprüft:
Wiederholungen/Ähnlichkeiten von Wörter und Wortgruppen
 - Aber: Unterbrechungspunkt vorgegeben



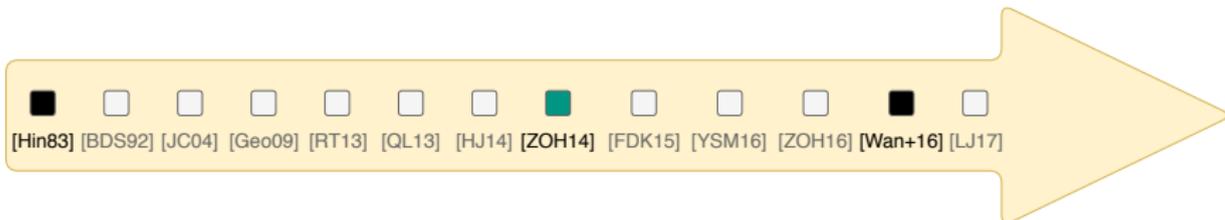
■ CRF [ZOH14]

- Conditional Random Field (CRF)
- Reduktion: Anzahl Wortarten



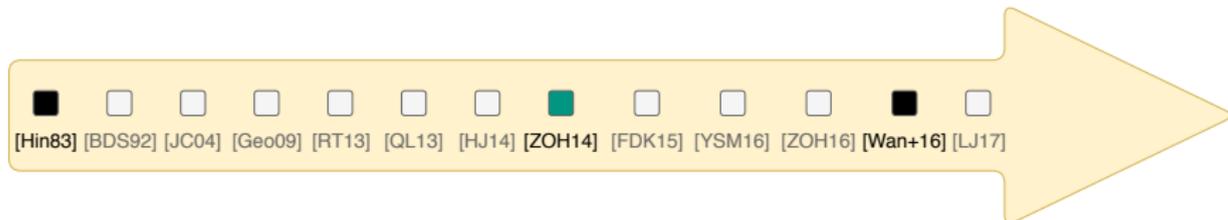
■ CRF [ZOH14]

- Conditional Random Field (CRF)
- Reduktion: Anzahl Wortarten



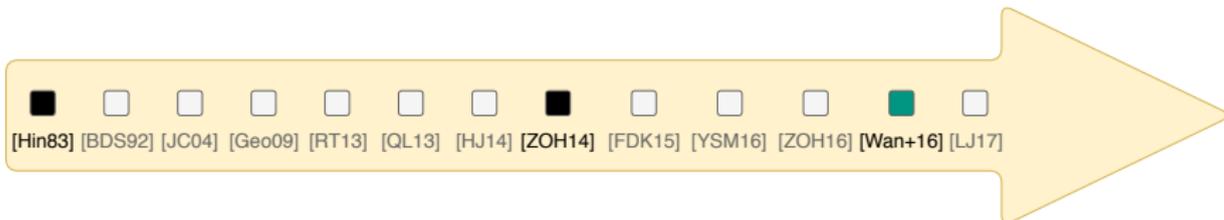
■ CRF [ZOH14]

- Conditional Random Field (CRF)
- Reduktion: Anzahl Wortarten



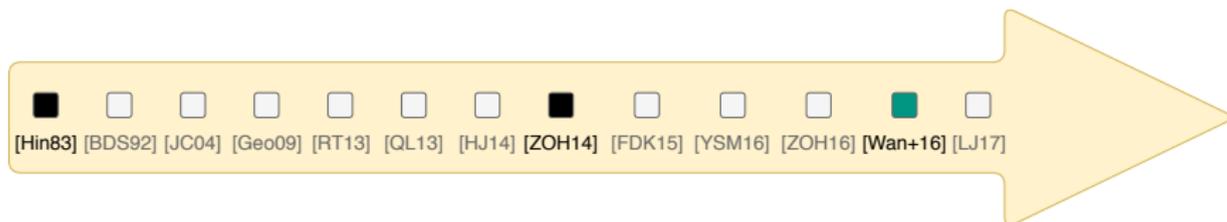
■ CRF [ZOH14]

- Conditional Random Field (CRF)
- Reduktion: Anzahl Wortarten



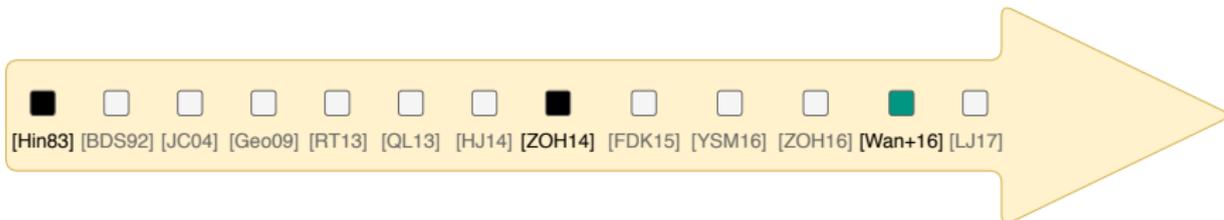
■ BI-LSTM + CRF-Schicht [Wan+16]

- Kombiniert bidirektionales Long Short-Term Memory (LSTM) mit CRF-Schicht
- Umgebungsvektor als Eingabemerkmale



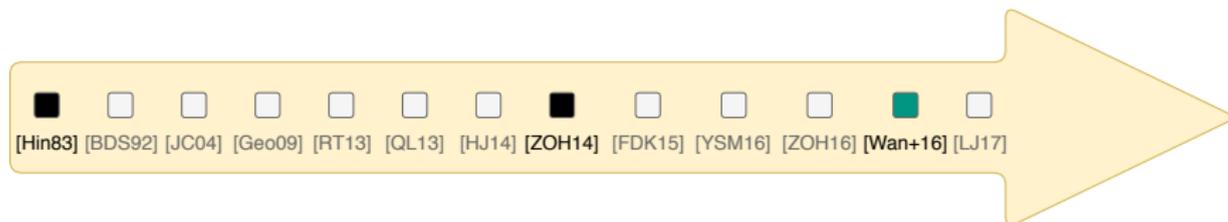
■ BI-LSTM + CRF-Schicht [Wan+16]

- Kombiniert bidirektionales Long Short-Term Memory (LSTM) mit CRF-Schicht
- Umgebungsvektor als Eingabemerkmale



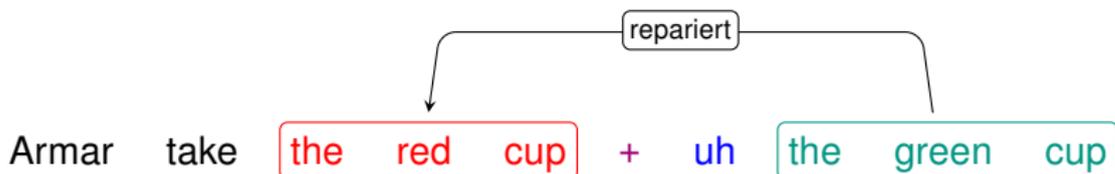
■ BI-LSTM + CRF-Schicht [Wan+16]

- Kombiniert bidirektionales Long Short-Term Memory (LSTM) mit CRF-Schicht
- Umgebungsvektor als Eingabemerkmale

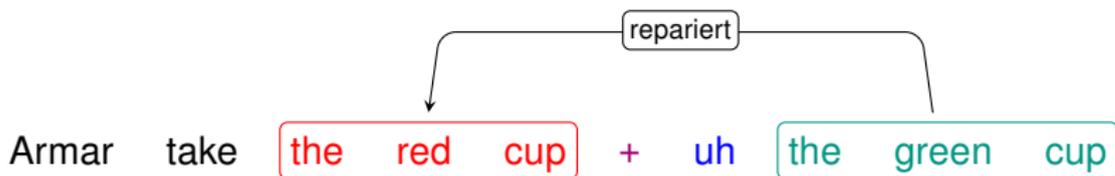


■ BI-LSTM + CRF-Schicht [Wan+16]

- Kombiniert bidirektionales Long Short-Term Memory (LSTM) mit CRF-Schicht
- Umgebungsvektor als Eingabemerkmale

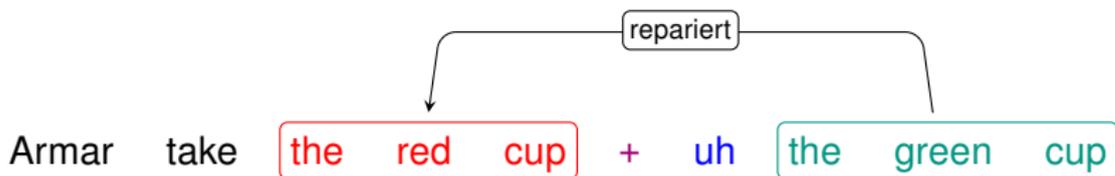


- Untersuchung: Gleichheit oder Ähnlichkeit
 - Wörter
 - Wortarten
 - Wortgruppen (mit IOB-Informationen)
- Füllwörter hilfreich ?

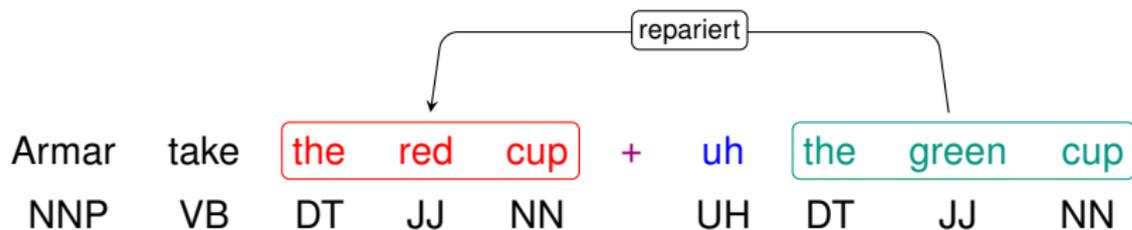


■ Untersuchung: Gleichheit oder Ähnlichkeit

- Wörter
- Wortarten
- Wortgruppen (mit IOB-Informationen)
- Füllwörter hilfreich ?



- Untersuchung: Gleichheit oder Ähnlichkeit
 - Wörter
 - Wortarten
 - Wortgruppen (mit IOB-Informationen)
- Füllwörter hilfreich ?



- Untersuchung: Gleichheit oder Ähnlichkeit
 - Wörter
 - Wortarten
 - Wortgruppen (mit IOB-Informationen)
- Füllwörter hilfreich ?

repariert

Armar	take	the	red	cup	+	uh	the	green	cup
NNP	VB	DT	JJ	NN		UH	DT	JJ	NN
NP	VP	NP	NP	NP		INTJ	NP	NP	NP

- Untersuchung: Gleichheit oder Ähnlichkeit
 - Wörter
 - Wortarten
 - Wortgruppen (mit IOB-Informationen)
- Füllwörter hilfreich ?

repariert

Armar	take	the	red	cup	+	uh	the	green	cup
NNP	VB	DT	JJ	NN		UH	DT	JJ	NN
NP	VP	NP	NP	NP		INTJ	NP	NP	NP

- Untersuchung: Gleichheit oder Ähnlichkeit
 - Wörter
 - Wortarten
 - Wortgruppen (mit IOB-Informationen)
- Füllwörter hilfreich ?

- Realistischere Bedingungen mit maschinellen Lernverfahren
- Beste Erkennungsraten mit LSTMs
- Bidirektionales LSTM

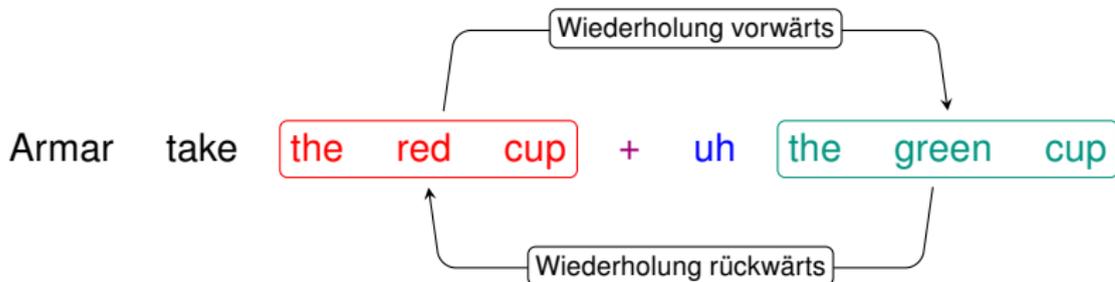


Analyse – Erkenntnisse verwandte Arbeiten

- Realistischere Bedingungen mit maschinellen Lernverfahren
- Beste Erkennungsraten mit LSTMs
- Bidirektionales LSTM



- Realistischere Bedingungen mit maschinellen Lernverfahren
- Beste Erkennungsraten mit LSTMs
- Bidirektionales LSTM



■ Bidirektionales LSTM

■ Eingabemerkmale

- Wortvektoren (Word Embeddings): Vortrainiertes Global-Vectors (GloVe)-Modell
- Wortarten
- Wortgruppen (mit IOB-Information)
- Umgebungsvektor (15 Wörter auf beiden Seiten)

■ Datensatz: Switchboard-Korpus

- ca. 2400 Telefongespräche
- 543 Sprecher, Teilnehmer aus ganz USA
- 70 verschiedene Themen

- Bidirektionales LSTM
- Eingabemerkmale
 - Wortvektoren (Word Embeddings): Vortrainiertes Global-Vectors (GloVe)-Modell
 - Wortarten
 - Wortgruppen (mit IOB-Information)
 - Umgebungsvektor (15 Wörter auf beiden Seiten)
- Datensatz: Switchboard-Korpus
 - ca. 2400 Telefongespräche
 - 543 Sprecher, Teilnehmer aus ganz USA
 - 70 verschiedene Themen

- Bidirektionales LSTM
- Eingabemerkmale
 - Wortvektoren (Word Embeddings): Vortrainiertes Global-Vectors (GloVe)-Modell
 - Wortarten
 - Wortgruppen (mit IOB-Information)
 - Umgebungsvektor (15 Wörter auf beiden Seiten)
 - Datensatz: Switchboard-Korpus
 - ca. 2400 Telefongespräche
 - 543 Sprecher, Teilnehmer aus ganz USA
 - 70 verschiedene Themen

- Bidirektionales LSTM
- Eingabemerkmale
 - Wortvektoren (Word Embeddings): Vortrainiertes Global-Vectors (GloVe)-Modell
 - Wortarten
 - Wortgruppen (mit IOB-Information)
 - Umgebungsvektor (15 Wörter auf beiden Seiten)
- Datensatz: Switchboard-Korpus
 - ca. 2400 Telefongespräche
 - 543 Sprecher, Teilnehmer aus ganz USA
 - 70 verschiedene Themen

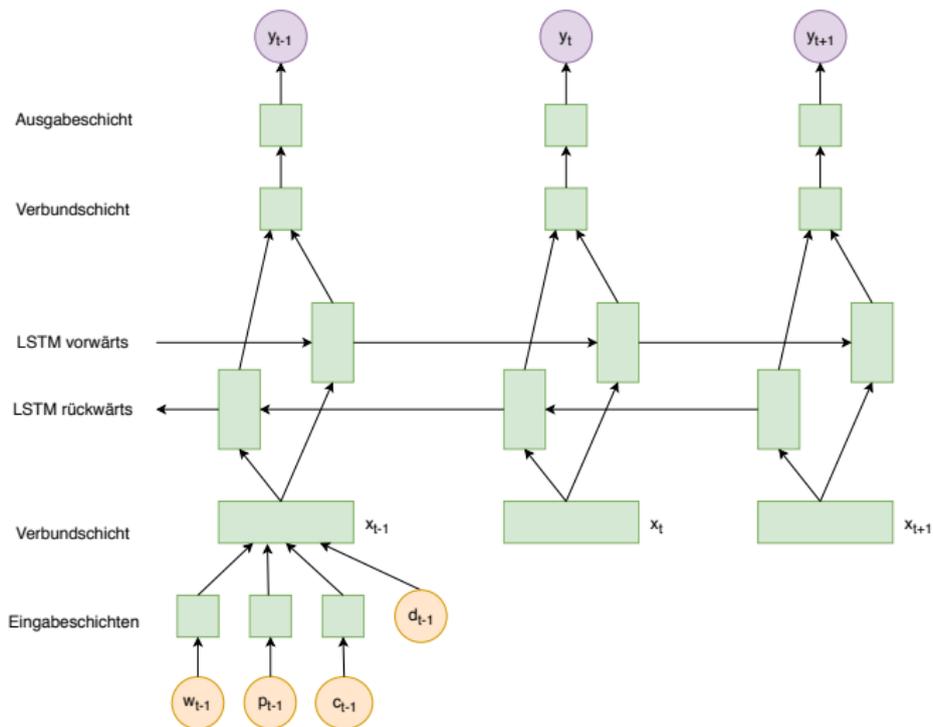
- Bidirektionales LSTM
- Eingabemerkmale
 - Wortvektoren (Word Embeddings): Vortrainiertes Global-Vectors (GloVe)-Modell
 - Wortarten
 - Wortgruppen (mit IOB-Information)
 - Umgebungsvektor (15 Wörter auf beiden Seiten)
- Datensatz: Switchboard-Korpus
 - ca. 2400 Telefongespräche
 - 543 Sprecher, Teilnehmer aus ganz USA
 - 70 verschiedene Themen

- Bidirektionales LSTM
- Eingabemerkmale
 - Wortvektoren (Word Embeddings): Vortrainiertes Global-Vectors (GloVe)-Modell
 - Wortarten
 - Wortgruppen (mit IOB-Information)
 - Umgebungsvektor (15 Wörter auf beiden Seiten)
- Datensatz: Switchboard-Korpus
 - ca. 2400 Telefongespräche
 - 543 Sprecher, Teilnehmer aus ganz USA
 - 70 verschiedene Themen

- Bidirektionales LSTM
- Eingabemerkmale
 - Wortvektoren (Word Embeddings): Vortrainiertes Global-Vectors (GloVe)-Modell
 - Wortarten
 - Wortgruppen (mit IOB-Information)
 - Umgebungsvektor (15 Wörter auf beiden Seiten)
- Datensatz: Switchboard-Korpus
 - ca. 2400 Telefongespräche
 - 543 Sprecher, Teilnehmer aus ganz USA
 - 70 verschiedene Themen

- Bidirektionales LSTM
- Eingabemerkmale
 - Wortvektoren (Word Embeddings): Vortrainiertes Global-Vectors (GloVe)-Modell
 - Wortarten
 - Wortgruppen (mit IOB-Information)
 - Umgebungsvektor (15 Wörter auf beiden Seiten)
- Datensatz: Switchboard-Korpus
 - ca. 2400 Telefongespräche
 - 543 Sprecher, Teilnehmer aus ganz USA
 - 70 verschiedene Themen

Struktur des neuronalen Netzwerks



Ausgabeklasse	Markierung
Disfluenz (außerhalb)	O-DF
Reparandum (Beginn)	B-RM
Reparandum (innen)	I-RM
Reparans (Beginn)	B-RS
Reparans (innen)	I-RS
Gefüllte Pause	FP
Expliziter Bearbeitungsbegriff	EE
Diskursmarker	DM

Armar take the red cup + uh the green cup

Ausgabeklasse	Markierung
Disfluenz (außerhalb)	O-DF
Reparandum (Beginn)	B-RM
Reparandum (innen)	I-RM
Reparans (Beginn)	B-RS
Reparans (innen)	I-RS
Gefüllte Pause	FP
Expliziter Bearbeitungsbegriff	EE
Diskursmarker	DM

Armar take the red cup + uh the green cup

Ausgabeklasse	Markierung
Disfluenz (außerhalb)	O-DF
Reparandum (Beginn)	B-RM
Reparandum (innen)	I-RM
Reparans (Beginn)	B-RS
Reparans (innen)	I-RS
Gefüllte Pause	FP
Expliziter Bearbeitungsbegriff	EE
Diskursmarker	DM

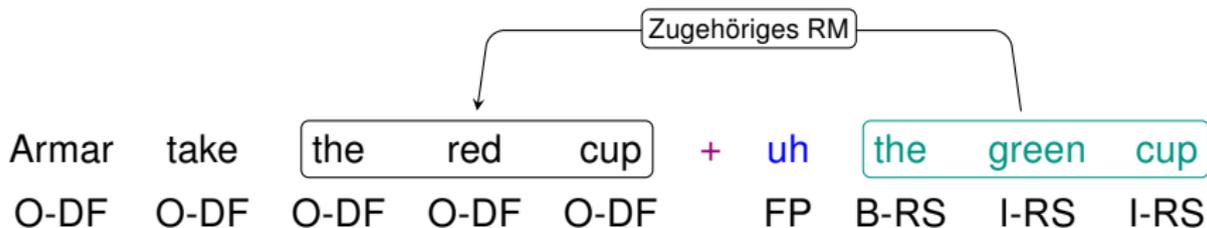
Armar take the red cup + uh the green cup

O-DF O-DF B-RM I-RM I-RM FP B-RS I-RS I-RS

- Nachverarbeitung der Schätzungen:



- Nachverarbeitung der Schätzungen:



■ Nachverarbeitung der Schätzungen:

Armar take the red cup + uh the green cup
O-DF O-DF B-RM I-RM I-RM FP B-RS I-RS I-RS

Klasse	Präzision	Ausbeute	F1
O-DF	0,951	0,990	0,970
B-RM	0,895	0,750	0,816
I-RM	0,848	0,591	0,697
B-RS	0,896	0,754	0,819
I-RS	0,824	0,434	0,559
FP	0,989	0,993	0,991
EE	0,970	0,938	0,953
DM	0,960	0,903	0,931
Alle	0,947	0,947	0,947
Alle außer O-DF	0,922	0,766	0,837

Klasse	Präzision	Ausbeute	F1
O-DF	0,951	0,990	0,970
B-RM	0,895	0,750	0,816
I-RM	0,848	0,591	0,697
B-RS	0,896	0,754	0,819
I-RS	0,824	0,434	0,559
FP	0,989	0,993	0,991
EE	0,970	0,938	0,953
DM	0,960	0,903	0,931
Alle	0,947	0,947	0,947
Alle außer O-DF	0,922	0,766	0,837

Klasse	Präzision	Ausbeute	F1
O-DF	0,951	0,990	0,970
B-RM	0,895	0,750	0,816
I-RM	0,848	0,591	0,697
B-RS	0,896	0,754	0,819
I-RS	0,824	0,434	0,559
FP	0,989	0,993	0,991
EE	0,970	0,938	0,953
DM	0,960	0,903	0,931
Alle	0,947	0,947	0,947
Alle außer O-DF	0,922	0,766	0,837

Klasse	Präzision	Ausbeute	F1
O-DF	0,951	0,990	0,970
B-RM	0,895	0,750	0,816
I-RM	0,848	0,591	0,697
B-RS	0,896	0,754	0,819
I-RS	0,824	0,434	0,559
FP	0,989	0,993	0,991
EE	0,970	0,938	0,953
DM	0,960	0,903	0,931
Alle	0,947	0,947	0,947
Alle außer O-DF	0,922	0,766	0,837

Klasse	Präzision	Ausbeute	F1
O-DF	0,951	0,990	0,970
B-RM	0,895	0,750	0,816
I-RM	0,848	0,591	0,697
B-RS	0,896	0,754	0,819
I-RS	0,824	0,434	0,559
FP	0,989	0,993	0,991
EE	0,970	0,938	0,953
DM	0,960	0,903	0,931
Alle	0,947	0,947	0,947
Alle außer O-DF	0,922	0,766	0,837

Klasse	Präzision	Ausbeute	F1
O-DF	0,951	0,990	0,970
B-RM	0,895	0,750	0,816
I-RM	0,848	0,591	0,697
B-RS	0,896	0,754	0,819
I-RS	0,824	0,434	0,559
FP	0,989	0,993	0,991
EE	0,970	0,938	0,953
DM	0,960	0,903	0,931
Alle	0,947	0,947	0,947
Alle außer O-DF	0,922	0,766	0,837

■ Metriken für zusammengefasste RM-Klasse

Model	Präzision	Ausbeute	F1
Basisversion	0,906	0,702	0,792
Ohne Wortgruppe	0,908	0,701	0,791
Ohne gefüllte Pausen	0,896	0,702	0,787
BI-LSTM-CRF [Wan+16]	0,910	0,836	0,871

■ Metriken für zusammengefasste RM-Klasse

Model	Präzision	Ausbeute	F1
Basisversion	0,906	0,702	0,792
Ohne Wortgruppe	0,908	0,701	0,791
Ohne gefüllte Pausen	0,896	0,702	0,787
BI-LSTM-CRF [Wan+16]	0,910	0,836	0,871

■ Metriken für zusammengefasste RM-Klasse

Model	Präzision	Ausbeute	F1
Basisversion	0,906	0,702	0,792
Ohne Wortgruppe	0,908	0,701	0,791
Ohne gefüllte Pausen	0,896	0,702	0,787
BI-LSTM-CRF [Wan+16]	0,910	0,836	0,871

■ Metriken für zusammengefasste RM-Klasse

Model	Präzision	Ausbeute	F1
Basisversion	0,906	0,702	0,792
Ohne Wortgruppe	0,908	0,701	0,791
Ohne gefüllte Pausen	0,896	0,702	0,787
BI-LSTM-CRF [Wan+16]	0,910	0,836	0,871

Evaluation – PARSE-Korpus

- Anweisungen an Roboter ARMAR-III
- Küchenumgebung
- Eingaben in Befehlsform, keine Sätze
- Bisheriger Korpus enthält nur 4 Reparaturen
⇒ Erweiterung notwendig



Evaluation – PARSE-Korpus

- Anweisungen an Roboter ARMAR-III
- Küchenumgebung
- Eingaben in Befehlsform, keine Sätze
- Bisheriger Korpus enthält nur 4 Reparaturen
⇒ Erweiterung notwendig



Evaluation – PARSE-Korpus

- Anweisungen an Roboter ARMAR-III
- Küchenumgebung
- Eingaben in Befehlsform, keine Sätze

- Bisheriger Korpus enthält nur 4 Reparaturen
⇒ Erweiterung notwendig



Evaluation – PARSE-Korpus

- Anweisungen an Roboter ARMAR-III
- Küchenumgebung
- Eingaben in Befehlsform, keine Sätze

- Bisheriger Korpus enthält nur 4 Reparaturen
⇒ Erweiterung notwendig



- 2 Szenarien
- 10 Probanden

	Sz. 1	Sz. 2	Gesamt
Aufnahmen	15	14	29
Wörter Transkriptionen	1269	1392	2661
Gefüllte Pausen	54	17	71
Reparaturen	12	15	27
Eingaben ohne Reparaturen	8	9	17

- Version ohne gefüllte Pausen mit Nachverarbeitung und (ohne Nachverarbeitung)

Ausgabeklasse	Präzision	Ausbeute	F1
B-RM	0,690 (0,667)	0,667 (0,467)	0,678 (0,549)
I-RM	0,688 (0,688)	0,611 (0,611)	0,647 (0,647)
B-RS	0,870 (0,833)	0,667 (0,500)	0,755 (0,625)
I-RS	0,846 (0,818)	0,524 (0,429)	0,647 (0,563)
RM+RS	0,765 (0,742)	0,626 (0,495)	0,689 (0,594)

- Version ohne gefüllte Pausen mit Nachverarbeitung und (ohne Nachverarbeitung)

Ausgabeklasse	Präzision	Ausbeute	F1
B-RM	0,690 (0,667)	0,667 (0,467)	0,678 (0,549)
I-RM	0,688 (0,688)	0,611 (0,611)	0,647 (0,647)
B-RS	0,870 (0,833)	0,667 (0,500)	0,755 (0,625)
I-RS	0,846 (0,818)	0,524 (0,429)	0,647 (0,563)
RM+RS	0,765 (0,742)	0,626 (0,495)	0,689 (0,594)

- Version ohne gefüllte Pausen mit Nachverarbeitung und (ohne Nachverarbeitung)

Ausgabeklasse	Präzision	Ausbeute	F1
B-RM	0,690 (0,667)	0,667 (0,467)	0,678 (0,549)
I-RM	0,688 (0,688)	0,611 (0,611)	0,647 (0,647)
B-RS	0,870 (0,833)	0,667 (0,500)	0,755 (0,625)
I-RS	0,846 (0,818)	0,524 (0,429)	0,647 (0,563)
RM+RS	0,765 (0,742)	0,626 (0,495)	0,689 (0,594)

- Version ohne gefüllte Pausen mit Nachverarbeitung und (ohne Nachverarbeitung)

Ausgabeklasse	Präzision	Ausbeute	F1
B-RM	0,690 (0,667)	0,667 (0,467)	0,678 (0,549)
I-RM	0,688 (0,688)	0,611 (0,611)	0,647 (0,647)
B-RS	0,870 (0,833)	0,667 (0,500)	0,755 (0,625)
I-RS	0,846 (0,818)	0,524 (0,429)	0,647 (0,563)
RM+RS	0,765 (0,742)	0,626 (0,495)	0,689 (0,594)

- Ziel: Detektion, Analyse und Korrektur von Disfluenzen
- Ansatz: Bidirektionales LSTM und Nachverarbeitung der Schätzungen
- Ergebnis
 - F1-Wert 0,837 für alle Klassen außer O-DF auf Switchboard-Korpus
 - F1-Wert 0,689 für RM- + RS-Klassen auf PARSE-Korpus
- Ausblick
 - Andere Eingabemerkmale, z.B. semantische Rolleninformationen
 - CRF-Schicht
 - Einbindung von Korreferenz-Informationen

- Ziel: Detektion, Analyse und Korrektur von Disfluenzen
- Ansatz: Bidirektionales LSTM und Nachverarbeitung der Schätzungen
- Ergebnis
 - F1-Wert 0,837 für alle Klassen außer O-DF auf Switchboard-Korpus
 - F1-Wert 0,689 für RM- + RS-Klassen auf PARSE-Korpus
- Ausblick
 - Andere Eingabemerkmale, z.B. semantische Rolleninformationen
 - CRF-Schicht
 - Einbindung von Korreferenz-Informationen

- Ziel: Detektion, Analyse und Korrektur von Disfluenzen
- Ansatz: Bidirektionales LSTM und Nachverarbeitung der Schätzungen
- Ergebnis
 - F1-Wert 0,837 für alle Klassen außer O-DF auf Switchboard-Korpus
 - F1-Wert 0,689 für RM- + RS-Klassen auf PARSE-Korpus
- Ausblick
 - Andere Eingabemerkmale, z.B. semantische Rolleninformationen
 - CRF-Schicht
 - Einbindung von Korreferenz-Informationen

- Ziel: Detektion, Analyse und Korrektur von Disfluenzen
- Ansatz: Bidirektionales LSTM und Nachverarbeitung der Schätzungen
- Ergebnis
 - F1-Wert 0,837 für alle Klassen außer O-DF auf Switchboard-Korpus
 - F1-Wert 0,689 für RM- + RS-Klassen auf PARSE-Korpus
- Ausblick
 - Andere Eingabemerkmale, z.B. semantische Rolleninformationen
 - CRF-Schicht
 - Einbindung von Korreferenz-Informationen

- Ziel: Detektion, Analyse und Korrektur von Disfluenzen
- Ansatz: Bidirektionales LSTM und Nachverarbeitung der Schätzungen
- Ergebnis
 - F1-Wert 0,837 für alle Klassen außer O-DF auf Switchboard-Korpus
 - F1-Wert 0,689 für RM- + RS-Klassen auf PARSE-Korpus
- Ausblick
 - Andere Eingabemerkmale, z.B. semantische Rolleninformationen
 - CRF-Schicht
 - Einbindung von Korreferenz-Informationen

- Ziel: Detektion, Analyse und Korrektur von Disfluenzen
- Ansatz: Bidirektionales LSTM und Nachverarbeitung der Schätzungen
- Ergebnis
 - F1-Wert 0,837 für alle Klassen außer O-DF auf Switchboard-Korpus
 - F1-Wert 0,689 für RM- + RS-Klassen auf PARSE-Korpus
- Ausblick
 - Andere Eingabemerkmale, z.B. semantische Rolleninformationen
 - CRF-Schicht
 - Einbindung von Korreferenz-Informationen

- Ziel: Detektion, Analyse und Korrektur von Disfluenzen
- Ansatz: Bidirektionales LSTM und Nachverarbeitung der Schätzungen
- Ergebnis
 - F1-Wert 0,837 für alle Klassen außer O-DF auf Switchboard-Korpus
 - F1-Wert 0,689 für RM- + RS-Klassen auf PARSE-Korpus
- Ausblick
 - Andere Eingabemerkmale, z.B. semantische Rolleninformationen
 - CRF-Schicht
 - Einbindung von Korreferenz-Informationen

- Ziel: Detektion, Analyse und Korrektur von Disfluenzen
- Ansatz: Bidirektionales LSTM und Nachverarbeitung der Schätzungen
- Ergebnis
 - F1-Wert 0,837 für alle Klassen außer O-DF auf Switchboard-Korpus
 - F1-Wert 0,689 für RM- + RS-Klassen auf PARSE-Korpus
- Ausblick
 - Andere Eingabemerkmale, z.B. semantische Rolleninformationen
 - CRF-Schicht
 - Einbindung von Korreferenz-Informationen

- Ziel: Detektion, Analyse und Korrektur von Disfluenzen
- Ansatz: Bidirektionales LSTM und Nachverarbeitung der Schätzungen
- Ergebnis
 - F1-Wert 0,837 für alle Klassen außer O-DF auf Switchboard-Korpus
 - F1-Wert 0,689 für RM- + RS-Klassen auf PARSE-Korpus
- Ausblick
 - Andere Eingabemerkmale, z.B. semantische Rolleninformationen
 - CRF-Schicht
 - Einbindung von Korreferenz-Informationen

- Asfour, Tamim u. a. (2006). „ARMAR-III: An integrated humanoid platform for sensory-motor control“. In: *Humanoid Robots, 2006 6th IEEE-RAS International Conference on*. IEEE, S. 169–175.
- Bear, John, John Dowding und Elizabeth Shriberg (1992). „Integrating multiple knowledge sources for detection and correction of repairs in human-computer dialog“. In: *Proceedings of the 30th annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, S. 56–63.
- Besser, Jana (Nov. 2006). „A Corpus-Based Approach to the Classification and Correction of Disfluencies in Spontaneous Speech“. Bachelor's Thesis. Saarland University, Germany.
- Bußmann, Hadumod [Hrsg.], Hrsg. (2002). *Lexikon der Sprachwissenschaft*. 3., aktualisierte u. erw. Aufl. Stuttgart: Kröner.

- Ferguson, James, Greg Durrett und Dan Klein (2015). „Disfluency Detection with a Semi-Markov Model and Prosodic Features.“. In: *HLT-NAACL*, S. 257–262.
- Georgila, Kallirroi (2009). „Using integer linear programming for detecting speech disfluencies“. In: *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*. Association for Computational Linguistics, S. 109–112.
- Godfrey, John J, Edward C Holliman und Jane McDaniel (1992). „SWITCHBOARD: Telephone speech corpus for research and development“. In: *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*. Bd. 1. IEEE, S. 517–520.

- Hindle, Donald (1983). „Deterministic parsing of syntactic non-fluencies“. In: *Proceedings of the 21st annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, S. 123–128.
- Honnibal, Matthew und Mark Johnson (2014). „Joint incremental disfluency detection and dependency parsing“. In: *Transactions of the Association for Computational Linguistics 2*, S. 131–142.
- Hough, Julian und David Schlangen (2015). „Recurrent neural networks for incremental disfluency detection“. In: *Interspeech 2015*.
- John, Godfrey und Edward Holliman (1993). „Switchboard-1 Release 2 LDC97S62“. In: *Web Download. Philadelphia: Linguistic Data Consortium*.

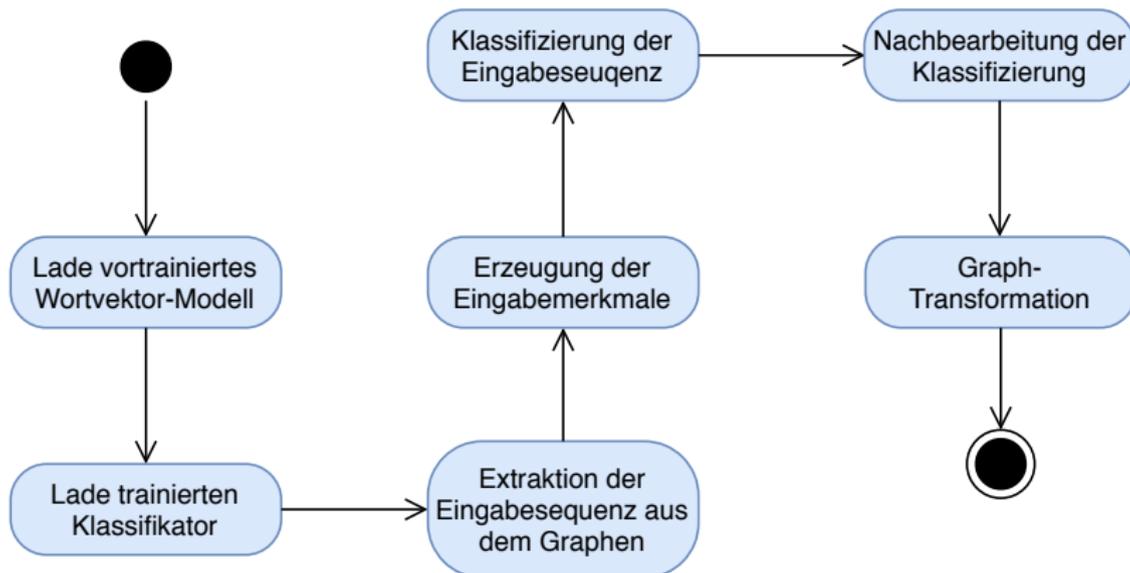
- Johnson, Mark und Eugene Charniak (2004). „A TAG-based noisy channel model of speech repairs“. In: *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, S. 33.
- Jurafsky, Dan und James H. Martin (2009). *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition*. 2. ed., Pearson International Edition. Prentice Hall, Pearson Education International.
- Lou, Paria Jamshid und Mark Johnson (2017). „Disfluency Detection using a Noisy Channel Model and a Deep Neural Language Model“. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Bd. 2, S. 547–553.

- Qian, Xian und Yang Liu (2013). „Disfluency detection using multi-step stacked learning“. In: *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, S. 820–825.
- Rasooli, Mohammad Sadegh und Joel R Tetreault (2013). „Joint Parsing and Disfluency Detection in Linear Time.“. In: *EMNLP*, S. 124–129.
- Shriberg, Elizabeth Ellen (1994). „Preliminaries to a theory of speech disfluencies“. Diss. University of California, Berkeley.
- Wang, Shaolei u. a. (2016). „Enhancing Neural Disfluency Detection with Hand-Crafted Features“. In: *China National Conference on Chinese Computational Linguistics*. Springer, S. 336–347.

- Weigelt, Sebastian und Walter F Tichy (2015). „Poster: ProNat: An Agent-Based System Design for Programming in Spoken Natural Language“. In: *Software Engineering (ICSE), 2015 IEEE/ACM 37th IEEE International Conference on*. Bd. 2. IEEE, S. 819–820.
- Yoshikawa, Masashi, Hiroyuki Shindo und Yuji Matsumoto (2016). „Joint Transition-based Dependency Parsing and Disfluency Detection for Automatic Speech Recognition Texts.“. In: *EMNLP*, S. 1036–1041.
- Zayats, Vicky, Mari Ostendorf und Hannaneh Hajishirzi (2016). „Disfluency detection using a bidirectional LSTM“. In: *arXiv preprint arXiv:1604.03209*.
- Zayats, Victoria, Mari Ostendorf und Hannaneh Hajishirzi (2014). „Multi-domain disfluency and repair detection“. In: *Fifteenth Annual Conference of the International Speech Communication Association*.

Nicht betrachtete Disfluenzen

- Prosodische Disfluenzen, z.B. Stottern: „fr-fridge“
- Wortfragmente, z.B. „frid“ statt „fridge“
- Nicht vom Sprecher reparierte Disfluenzen, z.B. „Go the to window and close it“



Evaluationsergebnisse – RS-Metriken

Switchboard-Korpus

Modell	Präzision	Ausbeute	F1
Basisversion	0,881	0,607	0,719
Ohne Wortgruppe	0,869	0,614	0,720
Ohne FP	0,865	0,624	0,725

- Version ohne gefüllte Pausen mit Nachverarbeitung und (ohne Nachverarbeitung)

Ausgabeklasse	Präzision	Ausbeute	F1
B-RM	0,690 (0,667)	0,667 (0,467)	0,678 (0,549)
I-RM	0,688 (0,688)	0,611 (0,611)	0,647 (0,647)
B-RS	0,870 (0,833)	0,667 (0,500)	0,755 (0,625)
I-RS	0,846 (0,818)	0,524 (0,429)	0,647 (0,563)
RM-Bin	0,733 (0,730)	0,688 (0,563)	0,710 (0,635)
RS-Bin	0,861 (0,828)	0,608 (0,471)	0,713 (0,600)
RM+RS	0,765 (0,742)	0,626 (0,495)	0,689 (0,594)