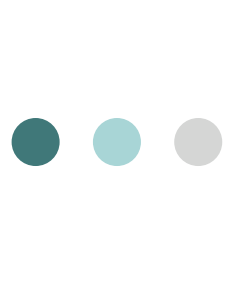




Koexpressivität von Sprache und Gestik in Wegbeschreibungen

Kirsten Bergmann
Arbeitsgruppe Künstliche Intelligenz
Universität Bielefeld



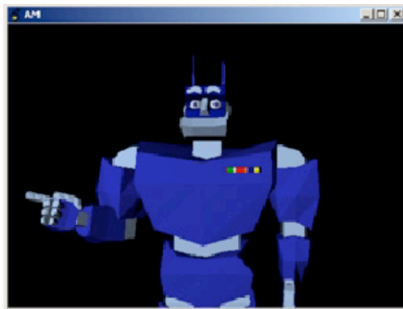
Multimodalität in der Mensch-Mensch Kommunikation

- Informationsaustausch über verbale und nonverbale Modalitäten:
 - Sprache
 - Gestik
 - Gesichtsmimik
 - Körperorientierung
- Bis zu 65% der Inhalte werden nonverbal kommuniziert
- Multimodalität wird auch für Mensch-Maschine-Schnittstellen angestrebt



Embodied Conversational Agents

- REA (Cassell 2000)

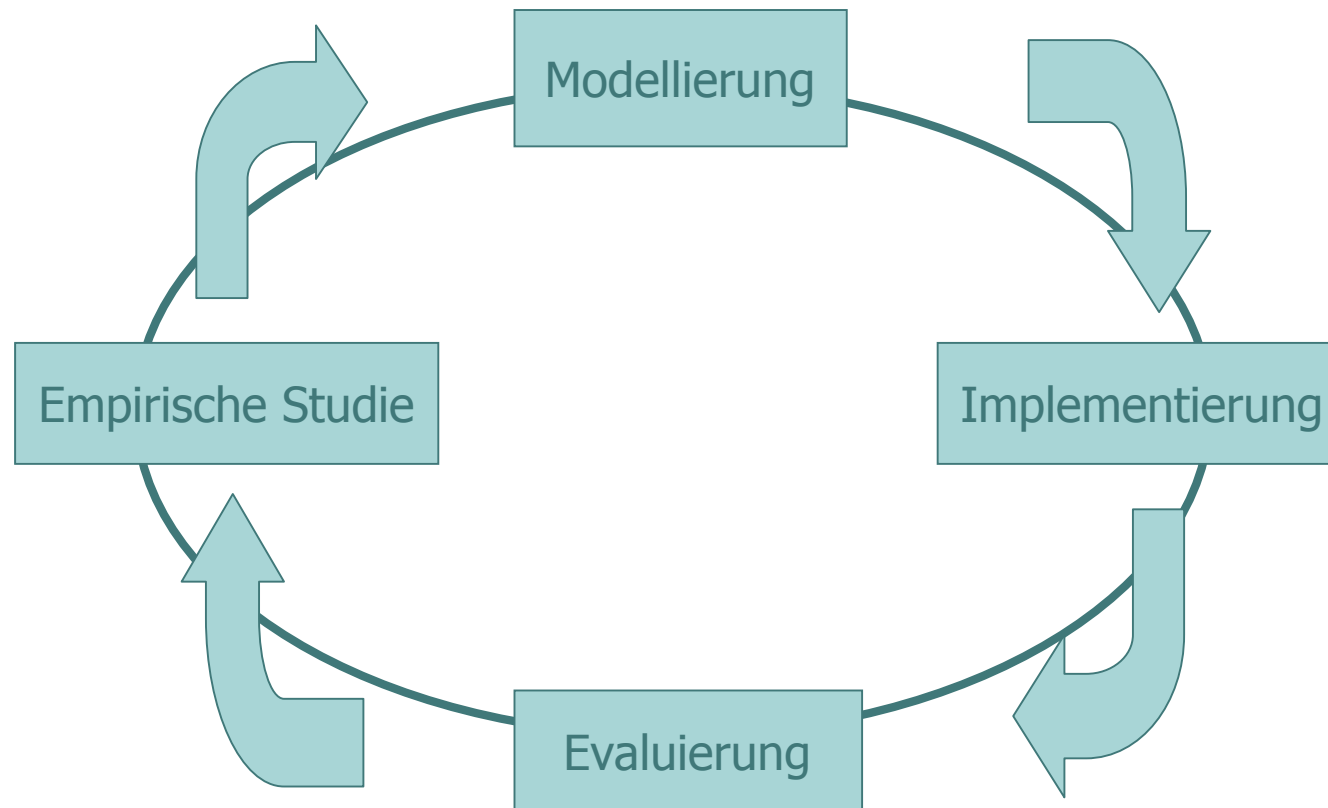


- NUMACK (Kopp et al. 2004)

- MAX
 - Kognitive Architektur
 - Emotionssystem
 - Sprach- und Gestenproduktion



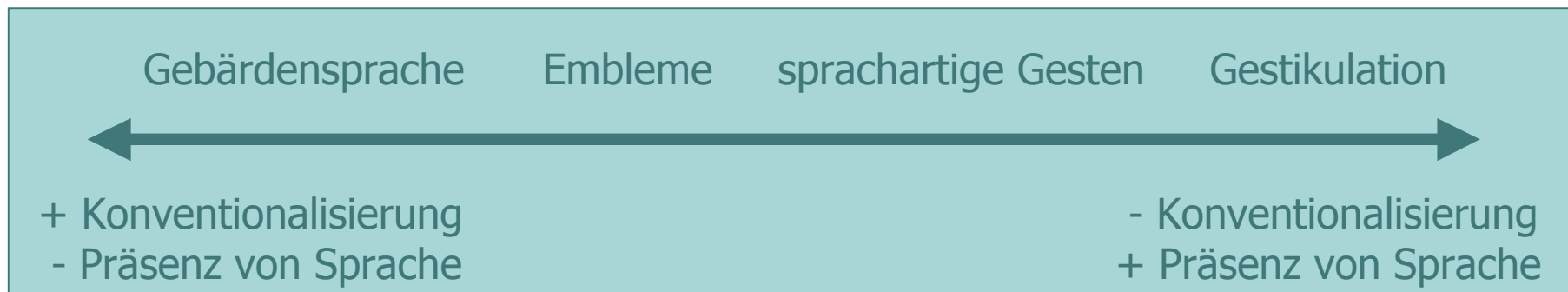
Modellierung der Kommunikationsfähigkeit in ECAs





Gestik

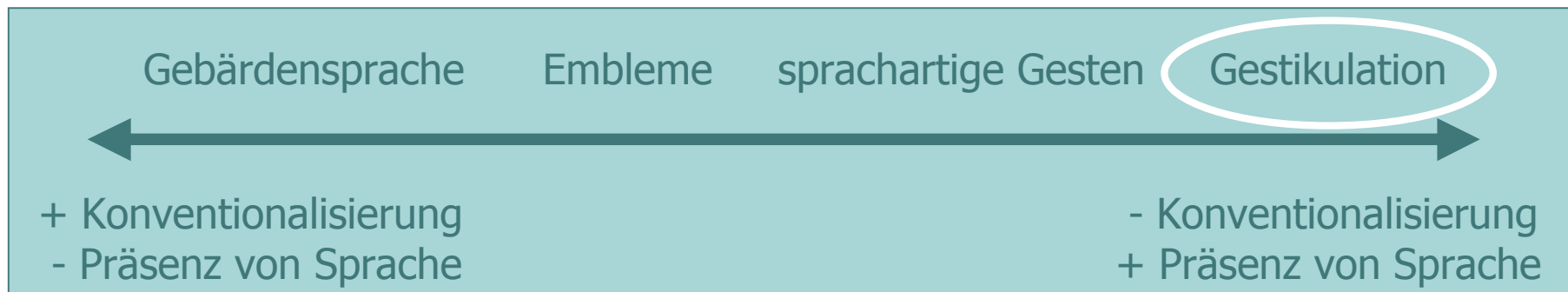
- Ursprung:
Art und Weise, wie Schauspieler und Redner ihren Körper einsetzen, um das Gesagte zu untermalen
- Arbeitsdefinition:
Bewegungen der Arme und Hände, die aufgrund einer Kommunikationsabsicht entstehen
- Kendon's Kontinuum (Kendon 1988), (McNeill 1992):





Gestik

- Ursprung:
Art und Weise, wie Schauspieler und Redner ihren Körper einsetzen, um das Gesagte zu untermalen
- Arbeitsdefinition:
Bewegungen der Arme und Hände, die aufgrund einer Kommunikationsabsicht entstehen
- Kendon's Kontinuum (Kendon 1988), (McNeill 1992):





Gestik

- Enge Verzahnung von Sprache und Gestik
 - Auch am Telefon wird gestikuliert (Williams 1977)
 - Rezipienten nutzen Informationen aus Gesten (Rogers 1978), (Cassell et al. 1999)
 - Zeitliche Synchronie (McNeill 1992)
- Klassifikation nach McNeill (1992, 2005):
 - Ikonische Gesten
 - Metaphorische Gesten
 - Deiktische Gesten
 - Beats

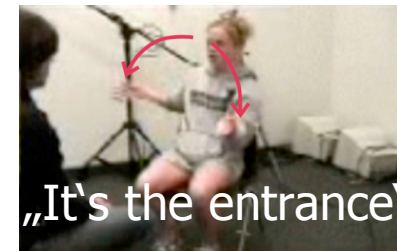
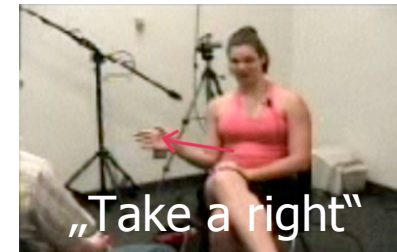


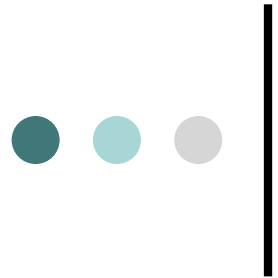
Gestik

- Enge Verzahnung von Sprache und Gestik
 - Auch am Telefon wird gestikuliert (Williams 1977)
 - Rezipienten nutzen Informationen aus Gesten (Rogers 1978), (Cassell et al. 1999)
 - Zeitliche Synchronie (McNeill 1992)
- Klassifikation nach McNeill (1992, 2005):
 - Ikonische Gesten
 - Metaphorische Gesten
 - Deiktische Gesten
 - Beats

● ● ● | Fragestellung

- Sprache und Gestik übermitteln nicht immer dieselbe Information
- Kontinuum der Koexpressivität
 - Redundanz
Sprache und Gestik vermitteln im Wesentlichen dieselbe Information
 - Komplementarität
Gesten enthalten Information, die nicht in der Sprache vorkommt
- Gibt es Faktoren die die Informationsverteilung zwischen Sprache und Gestik beeinflussen?





Korpus

- 28 Versuchspersonen geben einem „Ortsunkundigen“ eine Wegbeschreibung, die 5 markante Punkte auf dem Campus der Northwestern University (Chicago) verbindet
- Korpus: 10 Dialoge, 1500 Gesten
- Existierende Annotation (Kopp et al. 2004), (Kopp 2005):
 - Transkript der Sprache
 - Segmentierung der ikonischen/deiktischen Gesten
 - Gestenmorphologie nach McNeill (1992):
hand shape, hand orientation, hand location

Annotation in *Anvil*

The screenshot displays the Anvil 4.5.2 software interface, which is used for video annotation. The interface is divided into several windows:

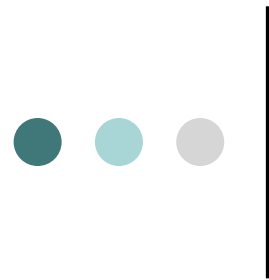
- Anvil 4.5.2 (Top Left):** Contains the menu bar (File, Edit, View, Tools, Bookmarks, ?) and a toolbar. It shows technical details: screen size: 352x234, frame rate: 29.97007, Audio format: ima4, 22050.0 Hz, 16-bit, Stereo, Video duration: 09:19:63 (16771 frames). Below this is the 'Current specification' section with a file path and a time display showing 00:31:66 and frame 949.
- Video: DG-1.mov (Top Center):** A video player showing a scene with four people in a room. A timestamp 00:00:38:14 is visible at the top.
- Track: discourse.dialogue acts DG (Top Right):** A panel for track management. It shows 'Track: discourse.dialogue acts DG', 'Referenced track: speech.DG', and 'Time:'. It includes an 'Attributes' text area, a 'Comment' text area, and buttons for 'start', 'edit', 'end', 'cut', 'extend', and 'del'.
- Annotation: DG-01_example.anvil (Bottom):** A large table for annotations. The top row shows a timeline from 00:31 to 00:34. The table has columns for different annotation types and their corresponding text or actions.

		00:31	00:32	00:33	00:34
speech	DF	mhm			
	DG	there..	a parking lot	directly behind this building.	So tha
	semantics	1, RELATIVE POSITION, new			
gesture	type	iconic + deictic; location of parking lot dir..			
	semantics	1, RELATIVE POSITION, new			
discourse	dialogue acts DG	Task, Statement			Task, Infl
	dialogue acts DF				Communication-m
	communicative goal	landmark with spatial orientation			comment



Reliabilität der Kodierung

- 2 Kodierer
- Kappa-Koeffizient (Cohen, 1960)
 - Gibt die Übereinstimmung von 2 Kodierern an
 - Berücksichtigt, dass auch bei zufälliger Zuordnung ein gewisses Maß an Übereinstimmung zu erwarten ist
 - Liefert Werte zwischen 0 und 1
 - Werte > 0.7 gelten als gute Übereinstimmung
- Durchschnittswert der Kodierung: $\kappa=0.774$

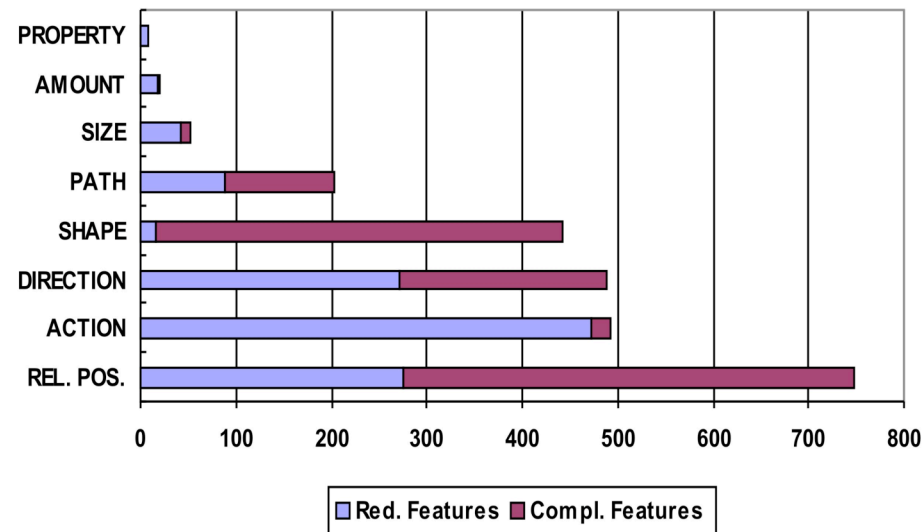


Semantischer Vergleich von Sprache und Gestik

- Annotation von *Semantic Features* (SFs) nach Jackendoff (1983) für Sprache und Gestik:
 - ENTITY
 - ACTION
 - DIRECTION
 - SHAPE
 - RELATIVE POSITION
 - SIZE
 - PATH
 - AMOUNT
 - PROPERTY
- Vergleich der SFs einer Geste mit denen des zugehörigen *lexical affiliate*:
 - Anzahl redundanter SFs
 - Anzahl komplementärer SFs



Typverteilung der SFs



- RELATIVE POSITION häufigstes SF
- ACTION meistens redundant
- SHAPE meistens komplementär



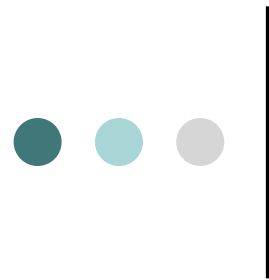
Koexpressivität von Sprache und Gesten

- 50:50 Verteilung redundanter/komplementärer SFs:
 - 49% aller SFs sind redundant
 - 51% aller SFs sind komplementär
- 38,9% aller Gesten sind ausschließlich komplementär
- 31,7% aller Gesten sind ausschließlich redundant
- 29,3% aller Gesten haben redundante und komplementäre Anteile



Hypothese 1: Probleme beim verbalen Enkodieren

- Zusammenhang zur Quantität der Gesten (Cohen 1977), (Bavelas et al. 2002)
- Annotation im Korpus:
 - Diskursmarker, sog. *Hedges*:
 - “kind of”, “sort of”, “like”, “somehow”
 - *Disfluencies* (Shriberg 1999):
 - Filled Pause: “um”, “uh”
 - Repetition: “the the”
 - Repair: “any health cov- any health insurance”
 - False Start: “It's fir- I could get it where I work”



Hypothese 1: Probleme beim verbalen Enkodieren

- Diskursmarker:
 - Geringerer Redundanzanteil in den Gesten ($p < 0.001$)
 - Höherer Komplementäranteil ($p < 0.001$)
- Disfluencies:
 - Geringerer Redundanzanteil in den Geste ($p < 0.001$)
 - Komplementäranteil unbeeinflusst



Hypothese 2: Feedback

- Feedback des *direction followers*:
 - Positives Feedback
 - Zwischenfragen
- Annotation nach DAMSL
(Dialog Act Markup in Several Layers – Allen & Core 1997):
 - Vorwärts gerichtete Funktion
 - *Statement*
 - *Influencing-addresse-future-action*
 - *Info-request*
 - Rückwärts gerichtete Funktion
 - *Understanding signals*
 - *Answers*



Hypothese 2: Feedback

- Zwischenfragen:
 - Kein signifikanter Einfluss
 - Tendenz zu leicht verringerter Redundanz in sprachbegleitenden Gesten zu Antworten ($p=0.195$)
- Positives Feedback:
 - Kein Einfluss von Verständnissignalen
- ABER:
 - *Direction follower* waren nur scheinbar ortsunkundig
 - Nur verbale Feedback-Signale berücksichtigt



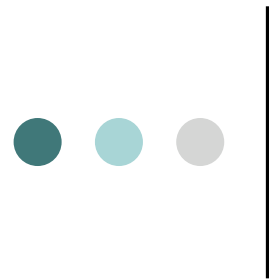
Hypothese 3: Kommunikative Ziele

- Kommunikative Ziele in Wegbeschreibungen (Denis 1997):
 - Instruktionen
 1. *Reorientation* („turn right“)
 2. *Locomotion* („keep on walking“)
 - Landmarken
 3. *Landmark* („Frances Searle“, Schilder, Straßen, Plätze)
 4. *Landmark description* („it's a small stone building“)
 5. *Landmark with spatial orientation* („Allen Center will be on your left“)
- 6. *Action+Landmark* („follow that path“, „you'll pass the Allen Center“)



Hypothese 3: Kommunikative Ziele

- Instruktionen (*reorientation, locomotion*):
 - Höhere Redundanz ($p < 0.001$)
 - Geringere Komplementarität ($p < 0.001$)
- Landmarken (*landmark, landmark description*):
 - Geringere Redundanz ($p < 0.001$)
 - Höhere Komplementarität ($p < 0.001$)
- Lokalisierung von Landmarken (*landmark with spatial orientation*)
 - Meistens RELATIVE POSITION als einziges redundantes SF ($p < 0.001$)
 - Geringere Komplementarität ($p < 0.001$)
- Kombination aus Instruktion und Landmarken (*action+landmark*)
 - Höhere Redundanz ($p < 0.001$)
 - Höhere Komplementarität ($p < 0.001$)



Hypothese 4: Informationsstatus

- Informationsstatus für jedes SF:
 - Neue Information
 - Gegebene Information: schon bekannt/erwähnt
- Insgesamt kein signifikanter Zusammenhang
- Für SFs des Typs ENTITY:
 - neu:
 - Geringere Redundanz ($p=0.001$)
 - gegeben:
 - Höhere Redundanz ($p=0.001$)

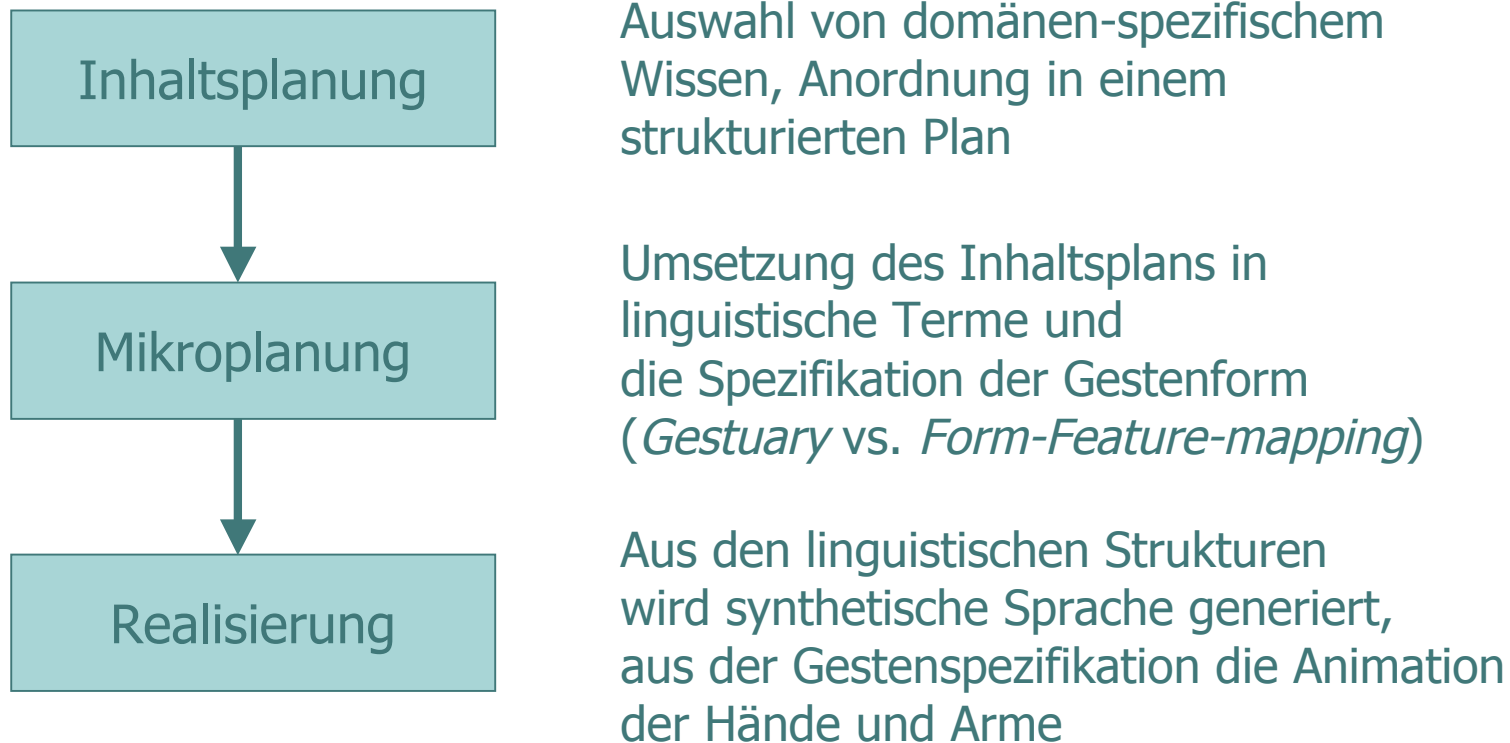


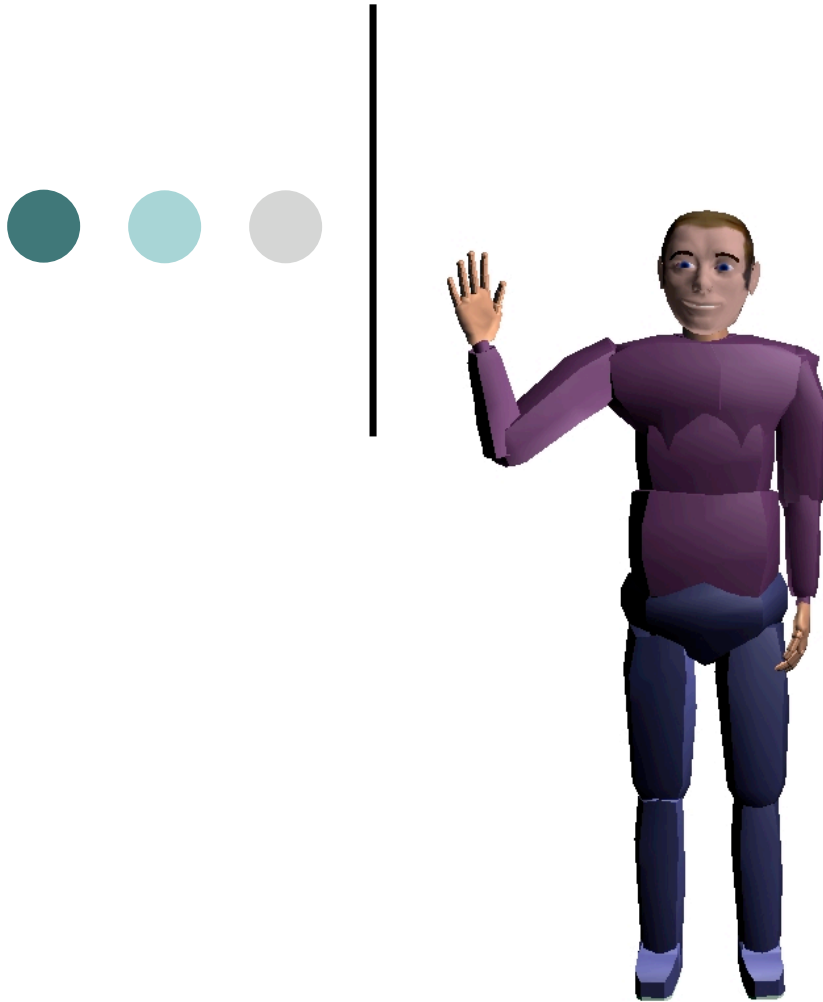
Zusammenfassung der Ergebnisse

- Probleme beim verbalen Enkodieren:
 - höherer Komplementarität
 - geringere Redundanz
- Einfluss kommunikativer Ziele:
 - höhere Redundanz bei Instruktionen
 - höhere Komplementarität bei Landmarken
- Informationsstatus von Entitäten:
 - etwas geringere Redundanz bei der Einführung von Entitäten
 - Höhere Redundanz bei gegebenen Entitäten
- Kein Einfluss von verbalem Feedback (Zwischenfragen, Verständnissignale)



Ausblick: Erweiterung einer NLG-Pipeline





Vielen Dank für die
Aufmerksamkeit!