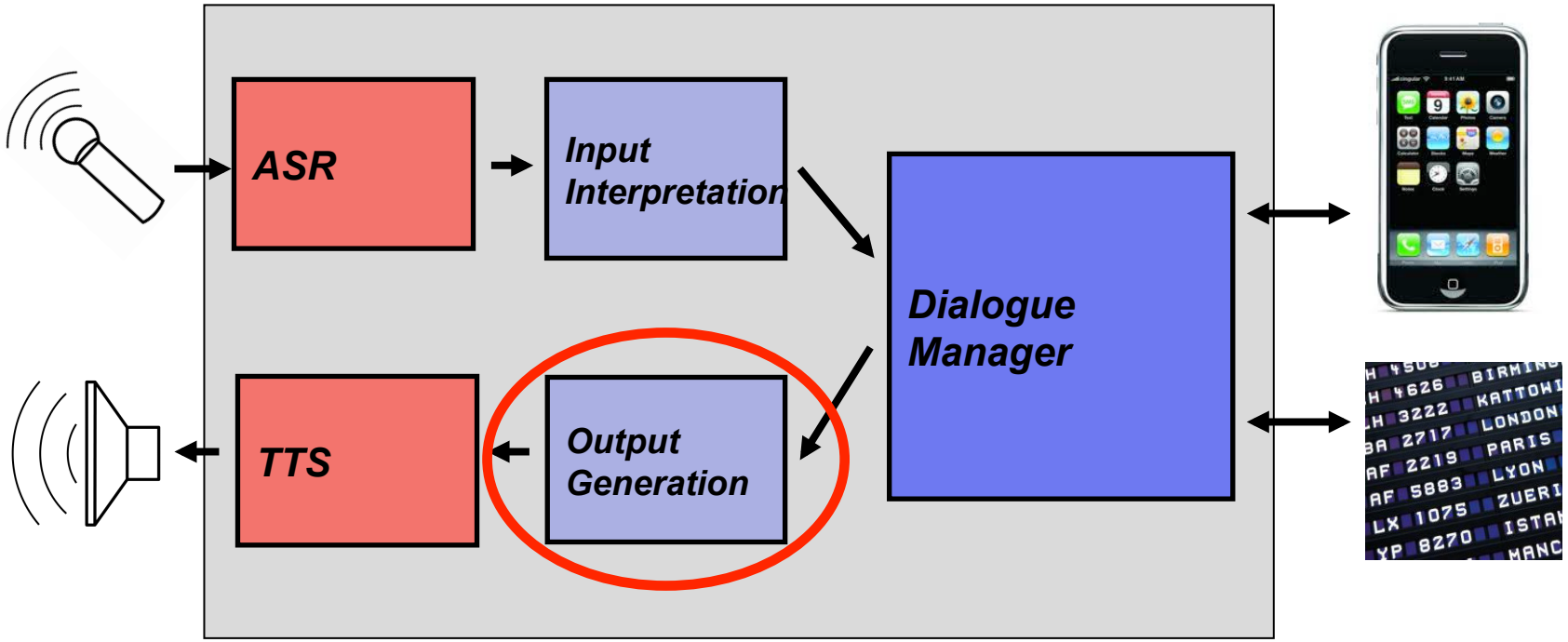
Language Technology II: Natural Language Dialogue

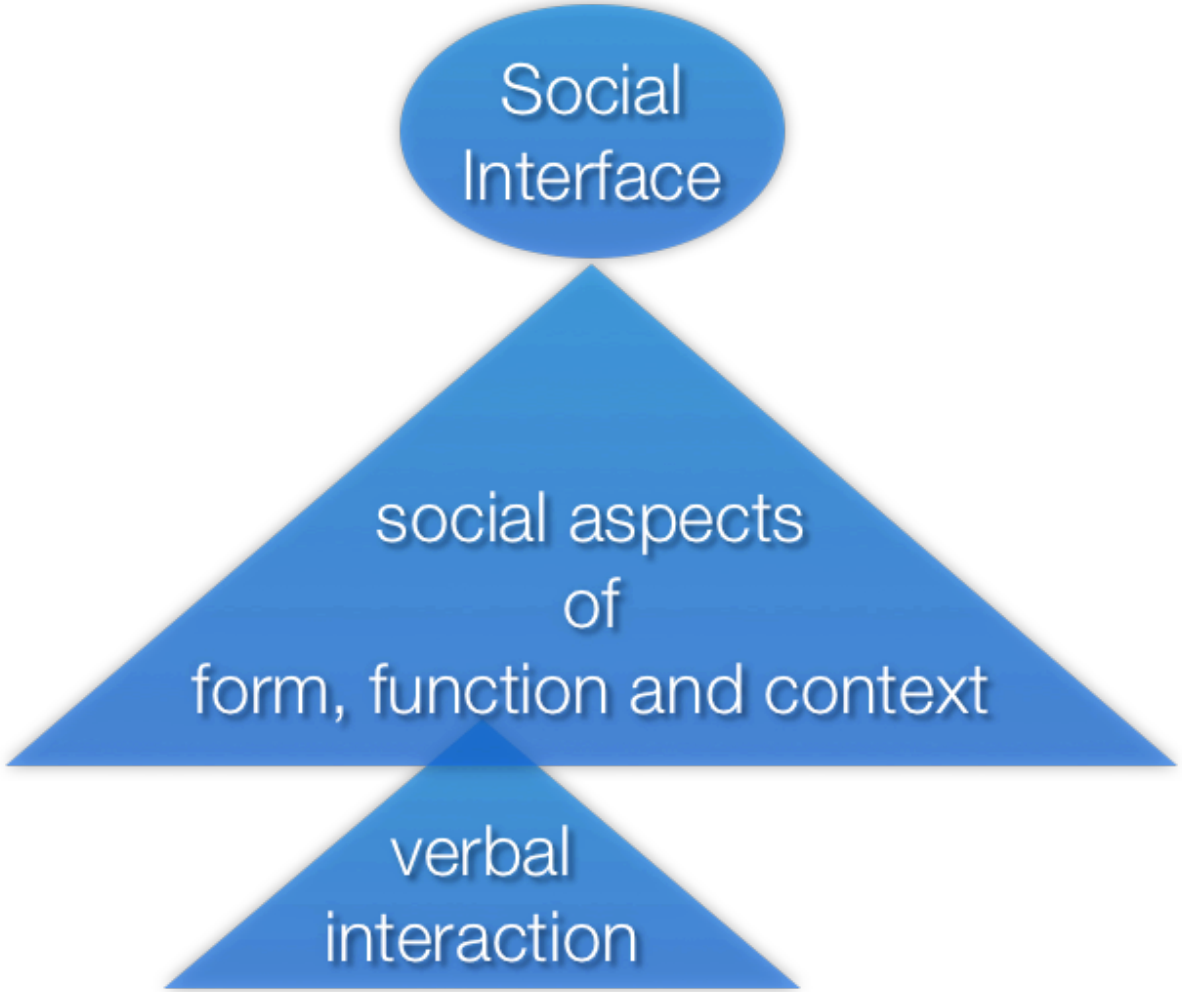
Verbal Output Generation in Dialogue Systems

Ivana Kruijff-Korbayová
ivana.kruijff@dfki.de

Dialog System: Basic Architecture



Social Qualities of Verbal System Output



Social Qualities of Verbal System Output

- Variation of surface realization form
- Agentivity:
  - Explicit reference to self as an agent
  - Explicit reference to any interaction participant as agent
- Familiarity display
  - Explicit reference to common ground
- Expressivity
  - Explicit reference to emotions and attitudes
- Alignment
  - Use of the same forms as the other

Agentivity
(personal vs. impersonal style)

Agentivity

- Explicit reference to self as an agent by use of agentive form, i.e., active voice, first person singular (I-form)
- Nass&Brave 2005:
  - experiments with speech interfaces with synthetic vs. recorded speech using agentive vs. non-agentive forms in product recommendations
  - finding: non-agentive form preferred for synthetic voices
  - possible explanation: system with synthetic voice does not have sufficient claim to (rational) agency
  - lesson: importance of consistency w.r.t. personality, gender, ontology (e.g., human-machine) ... and social role

Agentive Style and Entrainment

- Brennan&Ohaeri 1994:
  - experiments with a wizarded text-based dialogue system using agentive vs. non-agentive style
  - finding: users of a dialogue system more than twice as likely to use second person pronominal reference, indirect requests and politeness marking when the system used agentive style
  - lesson: users adopt style used by the system (entrainment)

TALK Project: SAMMIE System

- Multimodal interface to in-car MP3 player

- Playback control, search&browse DB, search, create&edit playlists
- Mixed initiative dialogue, unrestricted use of modalities
- Collaborative problem solving
- Multimodal turn-planning and NLG (German, English)

U: Show me albums by Michael Bublé .

S: I have these 3 albums. [+display]

U: Which songs are on this one?

S: The album Caught in the Act contains these songs.

U: Play the first one.

Output Variation in SAMMIE

- Personal vs. impersonal style
- Telegraphic vs. full utterance form
- Reduced vs. full referring expressions
- Lexical choice
- Presence vs. absence of adverbs

Output Variation in SAMMIE

- Agentivity: personal vs. impersonal style, e.g.,
  - Search result
    I found 23 albums. / You (We) have 20 albums.
    There are 23 albums.
  - Song addition
    I added the song "99 Luftballons" to Playlist 2.
    The song "99 Luftballons" has been added to Playlist 2.
  - Song playback
    I am playing the song "Feeling Good" by Michael Bublé.
    The song "Feeling Good" by Michle Bublé is playing.
  - Non-understanding
    I did not understand that.
    That has not been understood.
  - Clarification request
    Which of these 8 songs would you like to hear?
    Which of these 8 songs (is desired)?

Output Variation in SAMMIE

- Personal vs. impersonal style
- Telegraphic vs. full utterance form, e.g.,
  23 albums found vs. *I found 23 albums*.
- Reduced vs. full referring expressions, e.g.,
  the song vs. *the song "99 Luftballons"*
- Lexical choice, e.g.,
  song vs. *track vs. title*
- Presence vs. absence of adverbs, e.g,
  I will (now) play 99 Luftballons.

Sources of Output Variation Control

- Random selection
- Global (default) parameter settings
- Contextual information

Sources of Output Variation Control

- Random selection
- Global (default) parameter settings ~ style
- Contextual information

Evaluation Experiment

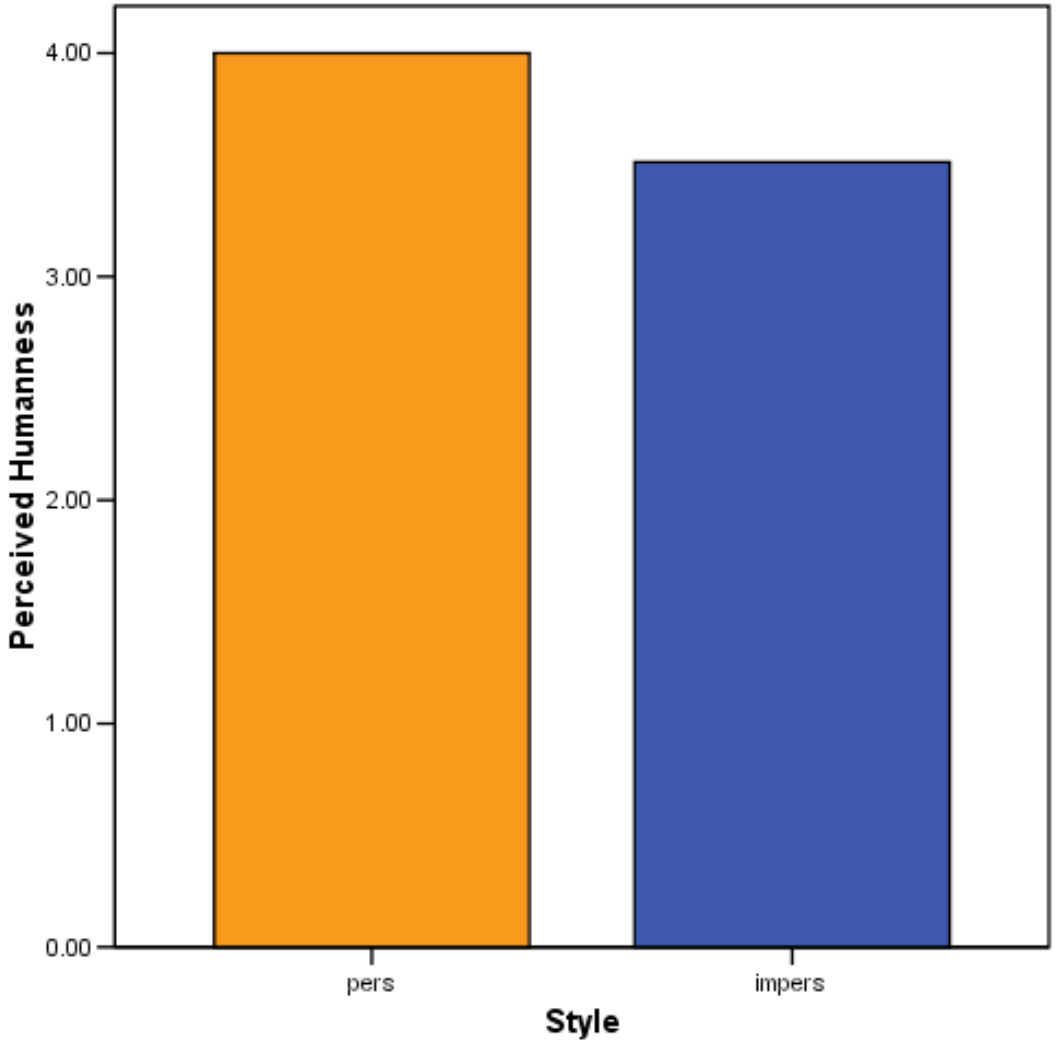
- *Personal vs. impersonal style*
- *28 subjects*
- *11 experimental tasks*
  - *Finding specific titles*
  - *Selecting tittles by constraints*
  - *Manipulating playlists*
  - *Free use*

Analysis:

- Questionnaire responses
  - General satisfaction
  - Ease of communication
  - Usability
  - Output clarity
  - Perceived humanness
  - Flexibility and creativity

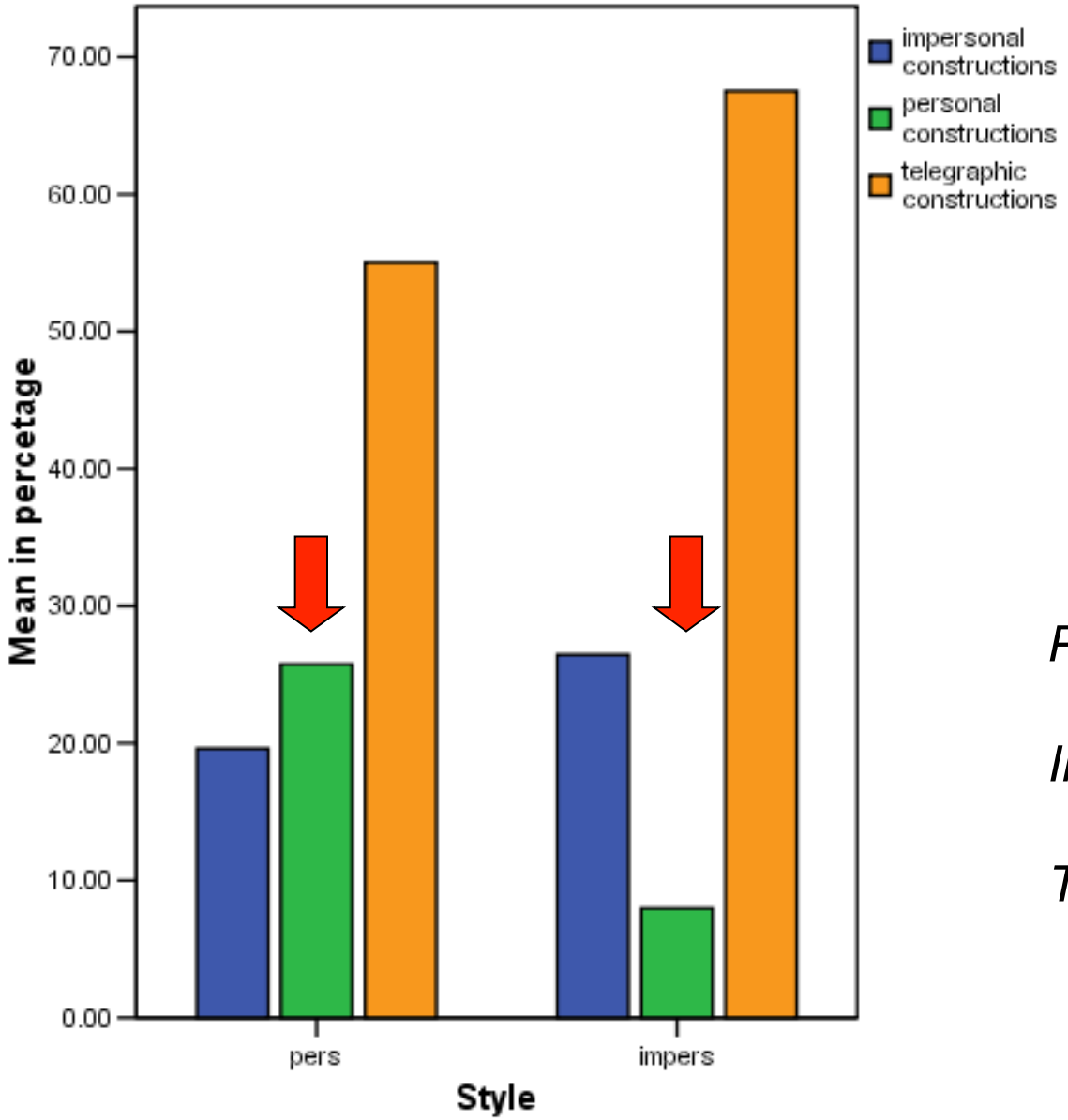
- Dialogue transcripts
  - Construction type
    - Personal
    - Impersonal
    - telegraphic
  - Personal pronouns
  - Politeness marking

Evaluation Results: Users' Attitudes



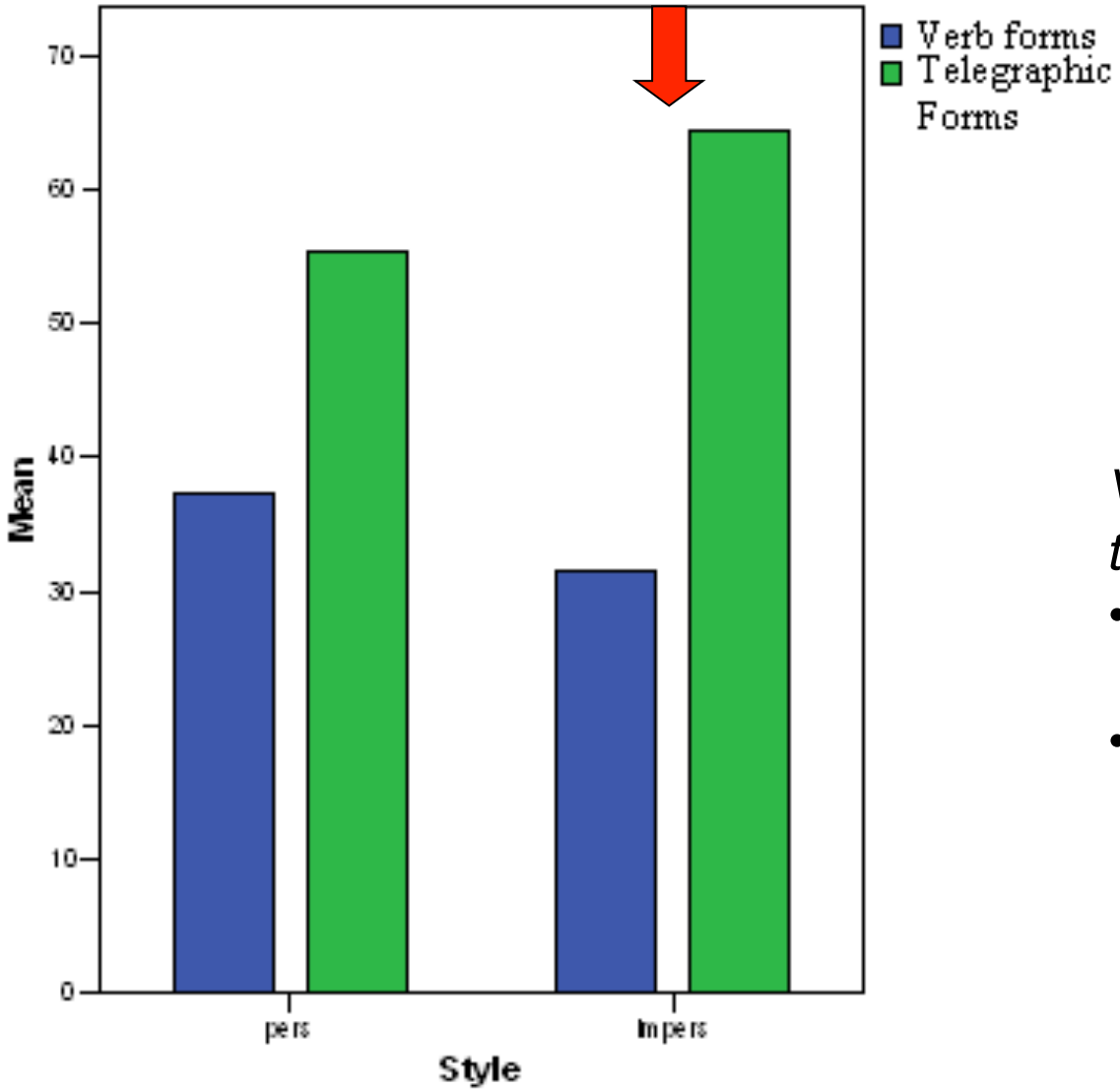
t(25)=1.64; p=.06

Evaluation Results: Users' Style



Personal constructions:
t(19)=1.8; p=.05
Impersonal constructions:
t(26)=1.0; p=.17
Telegraphic constructions:
t(26)=1.4; p=.09

Evaluation Results: Sentences vs. Fragments

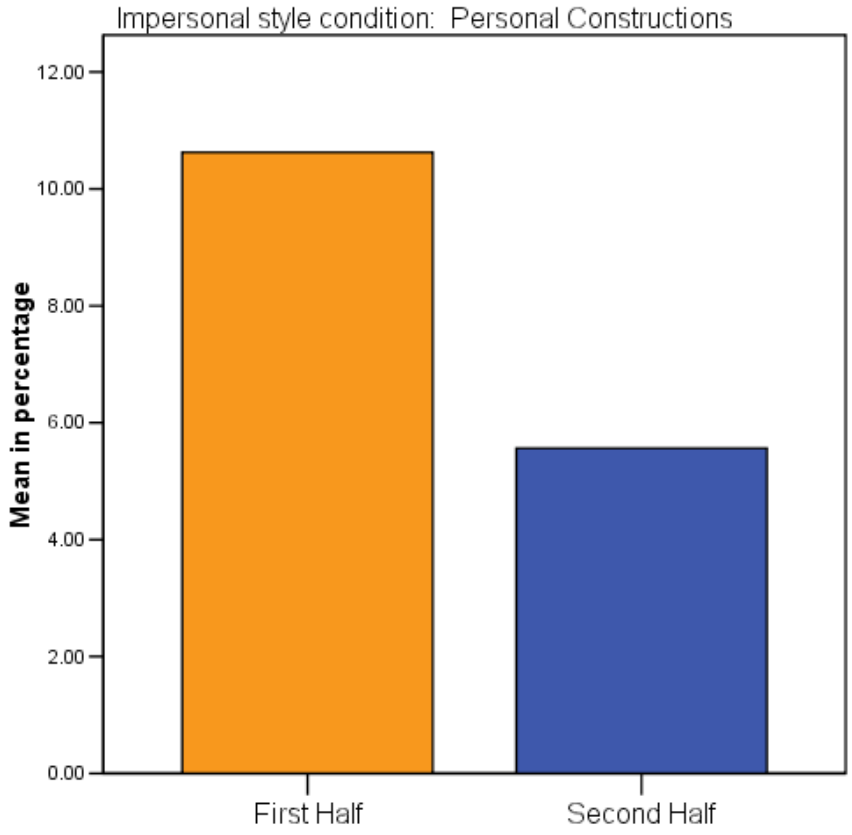


Verb-containing vs. telegraphic utterances:

- *impersonal style: t(13)=3.5; p=.00*
- *personal style: t(13)=.7; p=.25*

Evaluation Results: Alignment over Time

- Division of sessions into 2 halves
- Change from 1st to 2nd half in proportion of
  - Personal, impersonal and telegraphic constructions
  - Personal pronouns
  - Politeness marking
- Decrease in use of personal constructions in impersonal style condition;
- No other effect

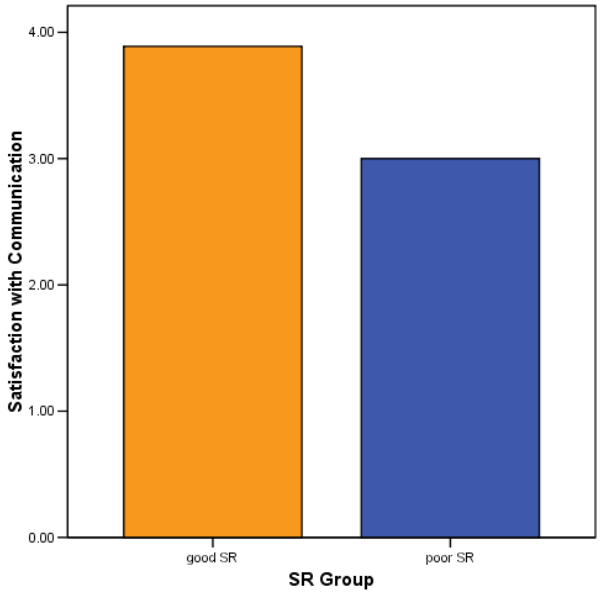


t(13)=2.5; p=.02

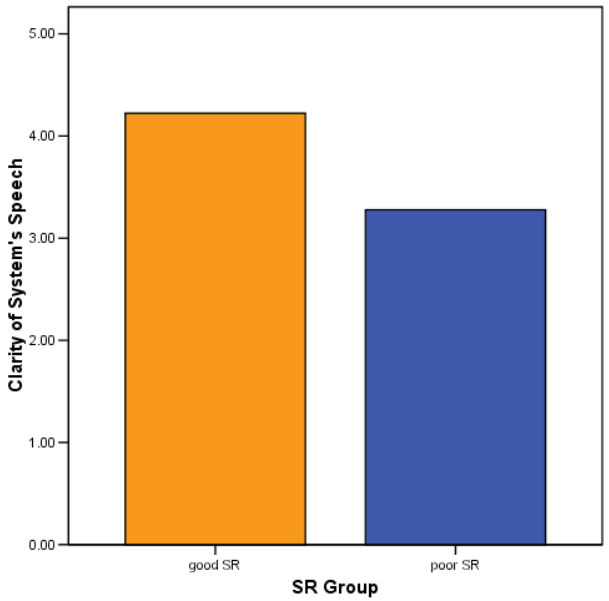
Evaluation Results: Influence of Speech recognition?

- Post-hoc analysis:
  Is there any difference in users' judgments of the system or in alignment behavior depending on speech recognition?
- 3 groups according to speech recognition performance
  - "good": < 30% utterances not understood (9 part.)
  - "average": 30-35% utterances not understood (10 part.)
  - "poor": > 35% utterances not understood (9 part.)

Speech Recognition and Users' Attitudes



t(16)=1.9; p=.04



t(16)=2.0; p=.03

Also for usability t(16)=1.71; p=.05 and perceived flexibility t(16)=1.61; p=.06

Speech Recognition Group		Satisfaction with Communication	Usability of the System	Perceived Flexibility of the System	Clarity of the System's Speech	Perceived Humanness of the System	Ease of Communication
good SR	Mean	3.8889	3.6444	3.2963	4.2222	3.7407	3.0889
	S. D.	.91287	.81104	1.12354	.97183	.54716	1.16237
poor SR	Mean	3.0000	3.0222	2.5556	3.2778	4.0000	2.8667
	S. D.	1.06719	.73106	.79931	1.03414	.66667	.93808

Evaluation Results: Summary

- More personal constructions in personal style condition;
  But not more impersonal ones in impersonal style
  and no difference w.r.t. telegraphic ones
- Significantly more telegraphic than verb-containing constructions in impersonal style; but no difference in personal style
- No difference in use of personal pronouns, politeness marking and speech recognition performance depending on style condition
- Decrease of personal constructions in impersonal style over time; but no other changes
- Better judgments of the system by users experiencing better speech recognition performance
- No influence of speech recognition performance on alignment

Conclusions and Open Issues

- Results consistent with earlier studies using non-interactive or simulated systems [Nass/Brave'05; Brennan/Ohaeri'94], but weaker
- Possible influencing factors
  - System interactivity
  - Domain/task
  - Cognitive load due to primary driving task
  - Speech recognition performance
  - Speech synthesis quality
- Definition of personal vs. impersonal style
- Neutral vs. de-agentivizing uses of constructions

Familiarity Display

Familiarity Display

- Explicit reference to common ground built up during an interaction and across multiple interactions

Familiarity Display

Familiarity display	Neutral display
Use of user's name:	
So, which answer do you choose, *Marco*?	So, which answer do you choose?
References to previous encounters and play experiences:	
I am happy to see you *again*.	I am happy to see you.
It was nice playing with you *last time*.	–
References to previous performance in an activity:	
Are you ready to play quiz *again*?	Are you ready to play quiz?
Today you were *again really good* at quiz.	Today you were really good at quiz.
Well done, you've done *better than last time*	Well done.
Reference to familiarity of a quiz question or a dance move:	
The next question should sound familiar.	The next question.
Let's try *again* this move: the spring step.	Let's try this move: the spring step.
Reference to familiarity of activity rules:	
Remember the magical pose?	Now the magical pose.

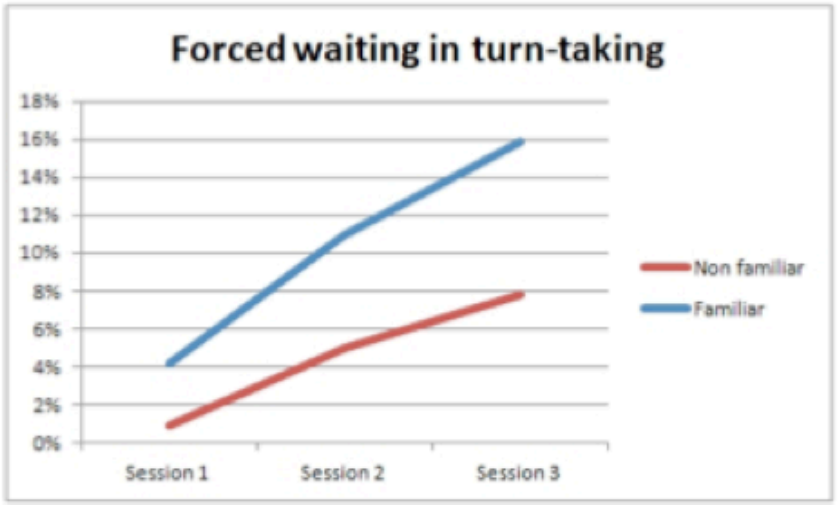
Familiarity Display

- Nalin et al. 2012, Aliz-E project:
  - experiment with a partly wizarded HRI system performing various activities with children over three sessions, with familiarity display vs. neutral w.r.t. familiarity

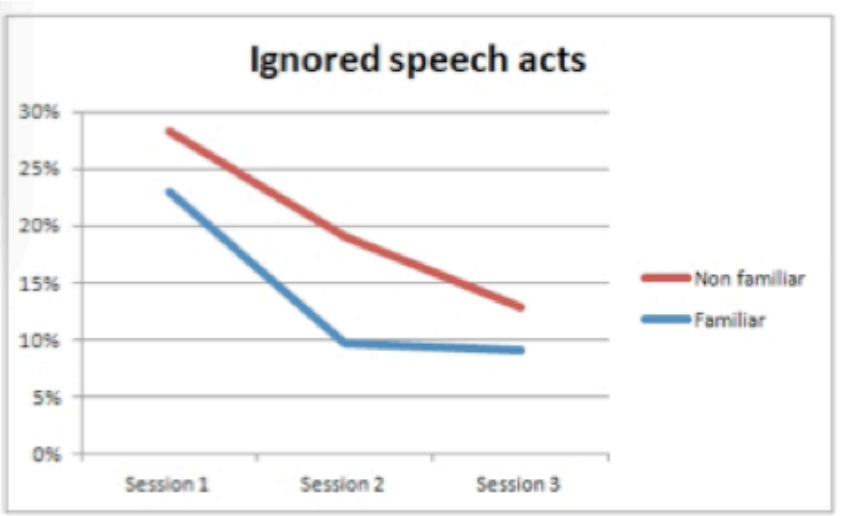


  - finding: adaptation of various aspects of verbal and non-verbal behavior, incl. speech timing, speed and tone, verbal input formulation, nodding and gestures
  - finding: more adaptation of verbal turn-taking behavior in the condition with familiarity display (waiting to speak, compliance)

Familiarity Display and Compliance



FD vs. ND condition ($F(1, 29)=4.375$, $p=0.047$)

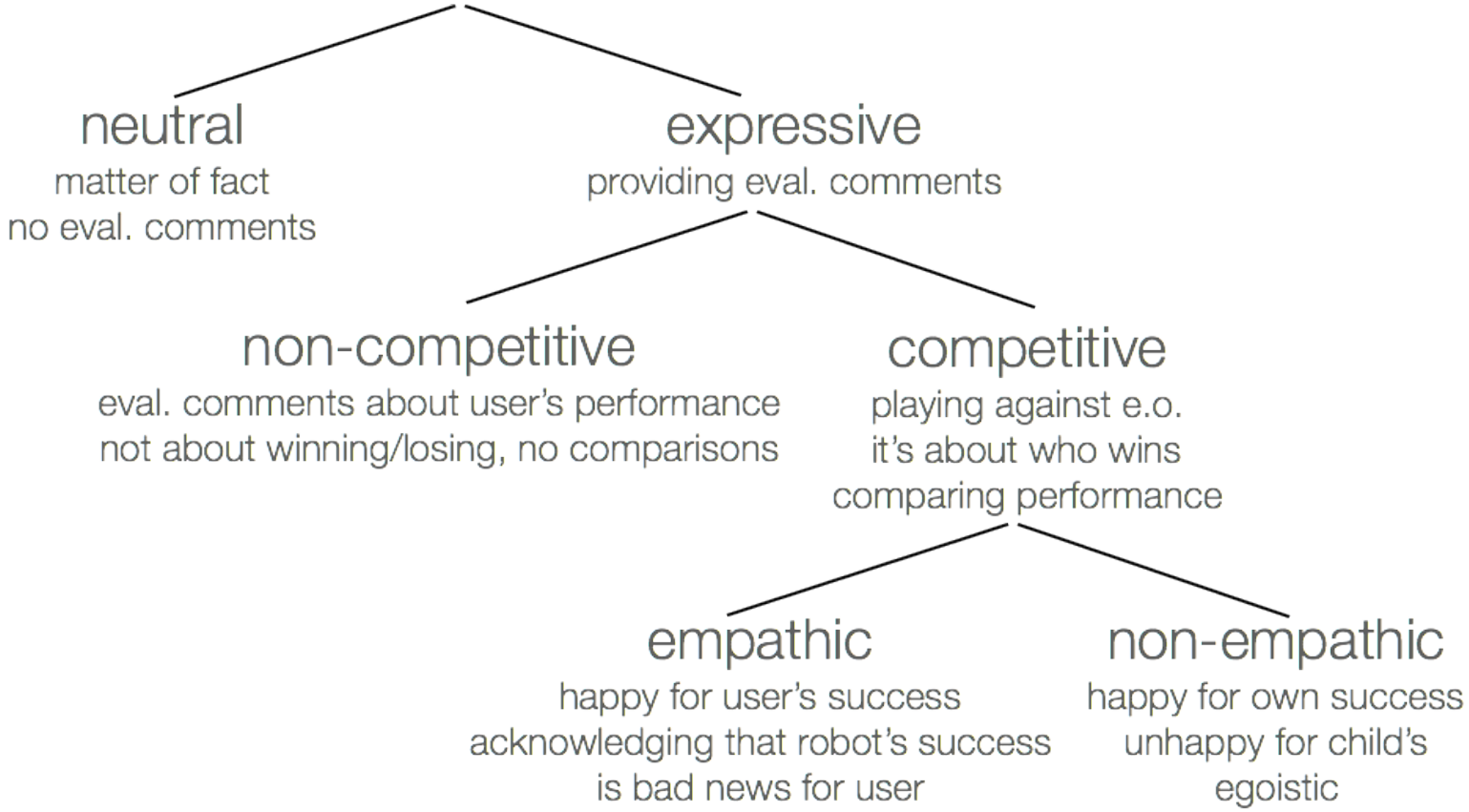


improvement across sessions ($F(1, 29)=10.608$, $p=0.001$)
FD vs. ND condition ($F(1, 29)=5.121$, $p=0.033$)

Conclusion: Explicit reference to common ground appears to positively influence commitment to interaction "success"

Expressivity

- Explicit reference to emotions and attitudes, e.g.: performance assessment in a game-like joint activity



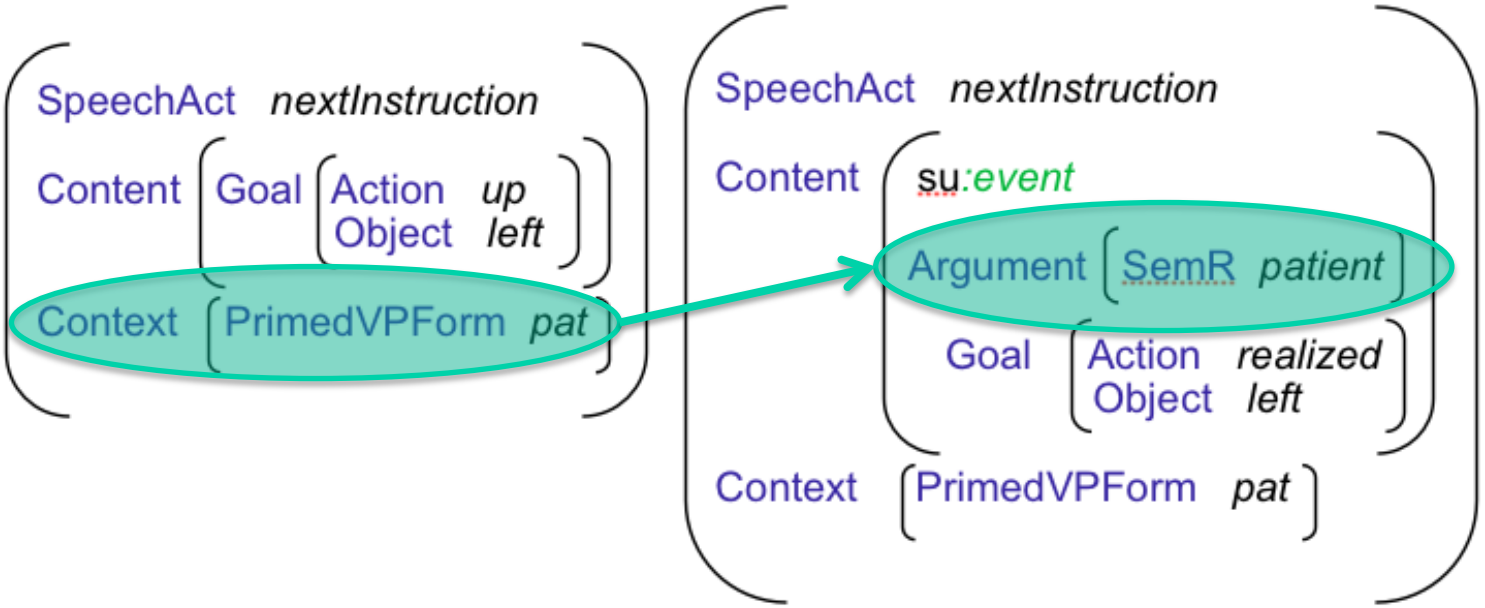
Lexical and Syntactic Alignment

Sources of Output Variation Control

- No control: random selection
- Global control:
  - default parameter settings
  - Parameter settings based on style
- Local control based on contextual information
  - Grounding status of content to be conveyed
    (cf. implicit grounding verification strategy)
  - Mimicking or adapting to user's style:
    = using the same surface realization forms as the other,
    based on linguistic features extracted from user's input
    ⇒ alignment/entrainment

Lexical and Syntactic Alignment

- Lexical and syntactic priming of system output by user input,
  e.g., U: *Right hand up* vs. U: *Raise the right arm*
  R: *Left hand up* vs. R: *Raise the left arm*
- Utterance planning:
  - Using primed alternatives to guide planning of output logical forms
  - Top-down planning: verb phrase, noun phrase

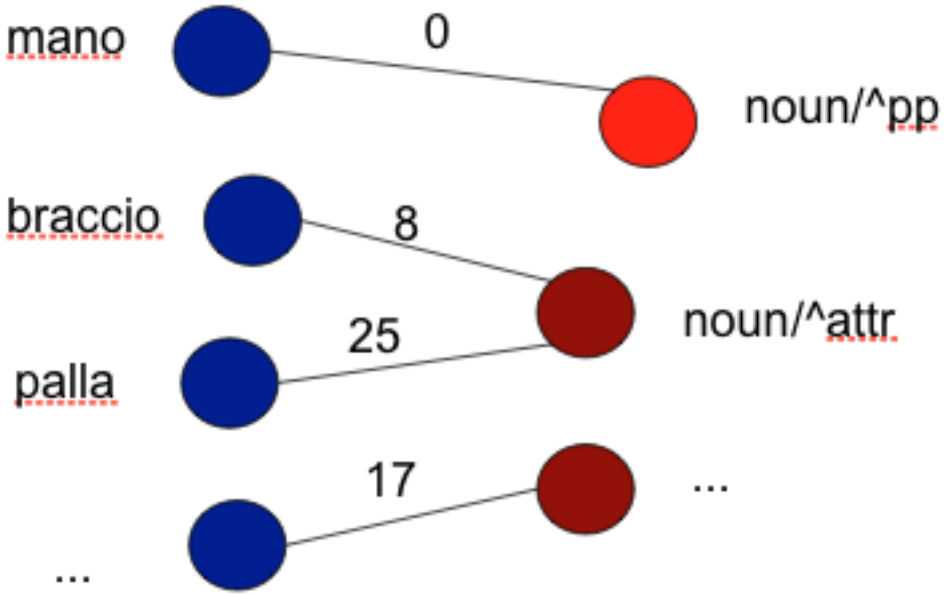


Lexical and Syntactic Alignment

- Using a memory model:
  a dictionary and an activation graph
- Activation updated after each user utterance
- Highly activated alternatives prime output planning

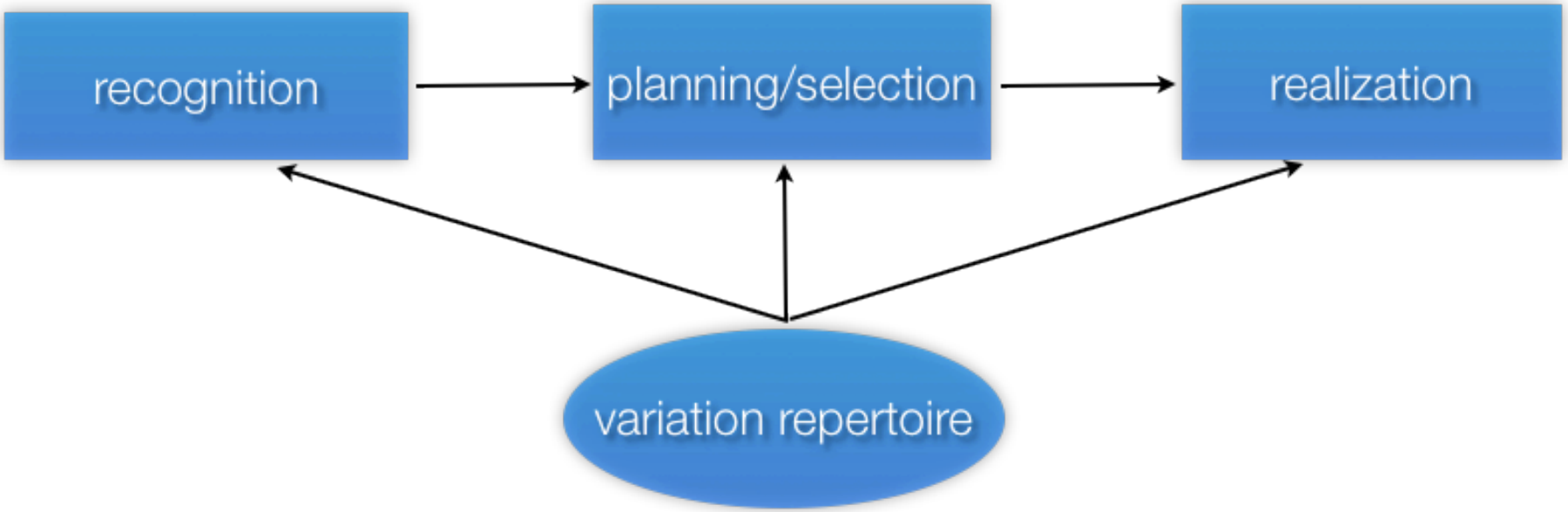
Child's utterance:

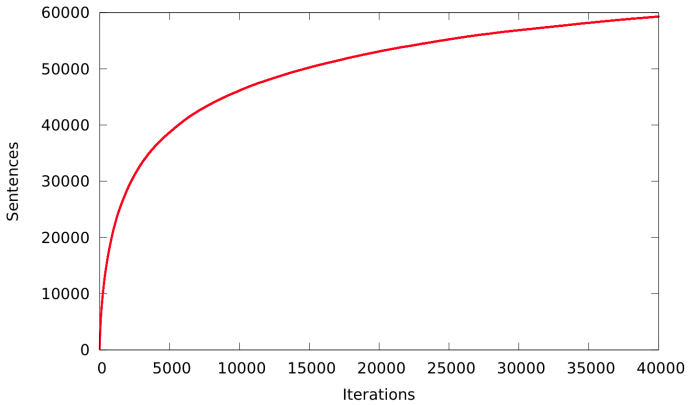
alzare	v\np/np
la	np/^n
mano	noun/^attr
destra	attr
...	



Generation of Varied System Output

System Output Variation





Aliz-E Quiz system 2012:
60 dialogue acts,
about 60k realization alternatives in total

System Output Variation

- Utterance planning rule example:

```
:canned
^ <DialogueAct>(!greeting ^ !provide  ^ !accept ^ !acknowledgement)
^ <Content>(!<About>name)
^ <Context>(<Familiarity>yes ^ <ChildName>#child) ^ <stringOutput>#s
^ <Control>(#ctrl: ^ !<familiarityDone>)
->
###s1 = concatenate(#s, ", ", #child),
###s2 = concatenate(#child, ", ", #s),
# ^ <stringOutput>random(###s1, ###s2, #s),
#ctrl ^  <familiarityDone>yes.
```

Summary

- Dialogue systems are perceived as social agents
- There are many dimensions of social qualities that human-computer interaction can/should reflect
  - Variation
  - Agentivity (personal vs. impersonal style)
  - Familiarity display
  - Expressivity
  - Alignment
- Also users adapt/entrain to system verbal behavior

Social Robots

- Duffy (2000):
  - societal robots: agents capable of interactive, communicative behavior
- Breazeal (2002):
  - sociable robots: communicate with humans, understand and relate to them in a personal way; humans understand them in social terms; socially intelligent in a human-like way
- Fong et al. (2003):
  - social robots: embodied agents in a society of robots or humans; recognize e.o., engage in social interactions, possess histories, explicitly communicate with and learn from e.o.
  - socially interactive robots: express and perceive emotions; communicate with high-level dialogue; learn and recognize models of other agents; can establish and maintain social relationships, using natural cues (gaze, gestures, etc.); exhibit distinctive personality and character; develop social competencies
- Bartneck & Forlizzi (2004)
  - social robots interact with humans by following their behavioral norms

References

- Bartneck, C., Forlizzi, J. (2004) A Design-Centred Framework for Social Human-Robot Interaction. In: Ro-Man 2004, Kurashiki, pp. 591–594.
- Breazeal, C. (2003): Towards sociable robots. Robotics and Autonomous Systems 42, 167–175,7.
- Breazeal, C. (2002): Designing Sociable Robots. MIT Press, Cambridge.
- Brennan, S. and Ohaeri, J.O. (1994). Effects of message style on user's attribution toward agents. In Pro-ceedings of CHI'94 Conference Companion Human Factors in Computing Systems, pages 281–282. ACM Press.
- Duffy, B.R. (2000) The Social Robot. Ph.D Thesis, Department of Computer Science, University College Dublin.Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. Robotics and Autonomous Systems, 42, 143-166.
- Hegel, F., Muhl, C., Wrede, B., Hielscher-Fastabend, M., & Sagerer, G. (2009). Understanding social robots. In *Advances in Computer-Human Interactions, 2009. ACHI'09. Second International Conferences on (pp. 169-174). IEEE.*
- *Kruijff-Korbayová, I. and Kukina, O. (2008) The effect of dialogue system output style variation on users' eval-uation judgements and input style. In Proceedings of SigDial'08, Columbus, Ohio.*
- *Nass, C. and Brave, S. (2005). Should voice interfaces say "I"? Recorded and synthetic voice interfaces' claims to humanity, chapter 10, pages 113–124. The MIT Press, Cambridge.*
- *Nalin, M., Baroni I., Kruijff-Korbayová, I., Canamero, L., Lewis, M.,Beck, A., Cuayáhuitl, H., Sanna, A. (2012) Children's adaptation in multi-session interaction with a humanoid robot. In Proceedings of the Ro-Man Conference, Paris, France.*