

# VOICE Awards: A German Human-Machine Dialog Corpus

Tatjana Scheffler   Roland Roller   Norbert Reithinger

DFKI GmbH  
firstname.lastname@dfki.de

October 10, 2009  
AAACL 2009



## 1 VOICE Awards Human-Machine Dialog Corpus

- A New Corpus
- Corpus Evaluation

## 2 Corpus-Assisted Creation of SDS Resources

- Domain Vocabulary
- Utterance Templates
- Dialog Automata

## 3 Summary



# VOICE Awards Human-Machine Dialog Corpus

- 1 VOICE Awards Human-Machine Dialog Corpus
  - A New Corpus
  - Corpus Evaluation
- 2 Corpus-Assisted Creation of SDS Resources
- 3 Summary



# Content of the VOICE Awards Corpus

**VOICE Awards competition:** dialogs of human users with state-of-the-art commercial spoken dialog systems; task-oriented, two-speaker dialogs with occasional overlapping speech

---

number of systems	≈ 140
number of dialogs	≈ 2000
words	> 500,000

---

years	2005–2009
users	lay, expert
length of dialogs	3 mins. (.5–20)

---



# Corpus Construction

- manually transcribed
- hand-annotated (single pass) for
  - dialog acts (18 types)
  - task success
  - errors/miscommunication
  - repetitions

year	# dialogs	transcribed	annotated	completed?
2005	~223	34%	0%	no
2006	~536	87%	19%	no
2007	432	100%	100%	yes
2008	471	100%	100%	yes
2009	~285	0%	0%	no
<b>total</b>	<b>~1947</b>	<b>74%</b>	<b>52%</b>	<b>early 2010</b>



# Annotation Schema: Dialog Acts

DA:	<b>social</b>	hello, bye, thank, sorry
	<b>request</b>	open-question
		request-info
		alternative-question
		yes-no-question
	<b>confirm</b>	explicit-confirm implicit-confirm
<b>metacommunication</b>	instruction	
	repeat-please	
	(request-instruction)	
<b>answer</b>	provide-info	
	accept, reject	
<b>other</b>	noise	
	other-da	



# Annotation Schema: Miscommunication

## Miscommunication:

<b>error</b>	not-understand misunderstand state-error no-input bad-input
<b>other miscommunication</b>	self-correct system-command other-error



# Annotation Schema: Task Success, Repetitions

## Task Success:

<b>success</b>	task-completed subtask-completed
<b>failure</b>	system-abort user-abort escalated abort-subtask other-failure

## Repetitions:

**repeat-prompt**  
**repeat-answer**





# Evaluation of Dialog Act Annotations

- re-annotated 69 dialogs (2375 dialog act segments)
- 2 annotators agree on segmentation 81% of the time
- $\kappa=0.89$  for dialog act labels (on matching segments only)
- dialog act schema is domain-specific and allows for very robust annotation
- confusion matrix reveals only a few categories lead to mismatches between annotators (e.g., alternative\_question and instruction)
- recovery from some errors through DA type hierarchy



# Dialog Act Annotation: Confusion Matrix

=== Confusion Matrix ===

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	<-- classified as
45	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	a = hello
2	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b = bye
0	0	18	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	c = thank
0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	d = sorry
0	0	0	0	67	0	0	0	0	0	0	1	0	0	0	0	0	0	0	e = open_question
0	0	0	0	19	88	2	0	0	0	0	4	15	0	1	0	0	0	0	f = request_info
0	0	0	0	0	11	155	2	0	0	0	73	0	0	6	0	0	0	0	g = alternative_question
0	0	0	0	0	2	1	11	0	0	0	1	0	0	0	0	0	0	0	h = yes_no_question
0	0	0	0	0	0	0	0	28	0	0	0	0	0	1	0	0	0	0	i = explicit_confirm
0	0	0	0	0	0	3	0	31	20	0	0	0	0	1	0	0	0	0	j = implicit_confirm
0	0	0	0	0	2	11	0	0	0	0	109	1	0	7	0	0	0	0	k = instruction
0	0	0	0	0	2	0	0	0	0	0	0	81	0	2	0	0	0	1	l = repeat_please
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	m = request_instruction
0	1	0	20	0	1	9	1	26	0	9	0	0	669	0	0	0	0	0	n = provide_info
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	36	0	0	0	o = accept
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0	0	p = reject
0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	40	0	q = noise
0	1	0	0	0	0	2	0	0	0	0	0	0	0	3	0	0	1	2	r = other_da



# Corpus-Assisted Creation of SDS Resources

- 1 VOICE Awards Human-Machine Dialog Corpus
- 2 Corpus-Assisted Creation of SDS Resources
  - Domain Vocabulary
  - Utterance Templates
  - Dialog Automata
- 3 Summary



# Domain Vocabulary

- automatic extraction of significant vocabulary by
  - 24 topic domains (bank, phone, cinema, traffic, ...)
  - 8 interaction types (data entry, ordering, game, ...)
- use for
  - ontology
  - ASR/NLU
- chi-square test, TF-IDF
- good results even with small data sets  
(2-3 systems per domain)
- in addition, extracted SDS-typical vocabulary



## Banking:

<b>term</b>	<b>English</b>	$\chi^2$
Kontostand	account balance	56.6
Kontonummer	account number	54.5
möchten	would like	44.1
Umsätze	transactions	40.7
Konto	account	40.2
Überweisung	wire transfer	32.9
Cent	Cent	29.1
minus	negative	28.1
Ziffer	digit	27.6
Geburtsdatum	birth date	26.0
Hauptmenü	main menu	23.9
Bankleitzahl	routing number	22.9
Servicewunsch	service request	21.8
beträgt	amounts to	21.3
Gutschrift	credit	20.8



# Utterance Templates

- user utterance templates
  - for generation
  - supermajority (>80%) of real user utterances consists of just one word (or name/number)
- system utterance templates
  - for natural language understanding
  - use POS-tagger and chunker to construct templates with empty slots from the corpus
  - exploring alignment-based learning

...

Otherwise please say \*XX\* or \*XX\*.

Please state your \*XX\* now.

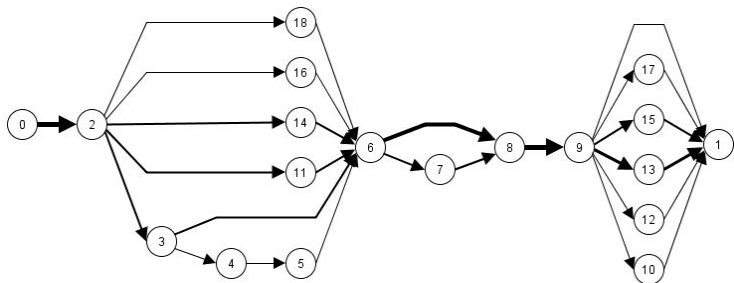
Please choose \*XX\*, \*XX\* or \*XX\*.

...



# Dialog Automata – System Knowledge

- automatically extract a system model from the corpus
- visualize the interaction
- goal: derive aggregate user behavior from the corpus



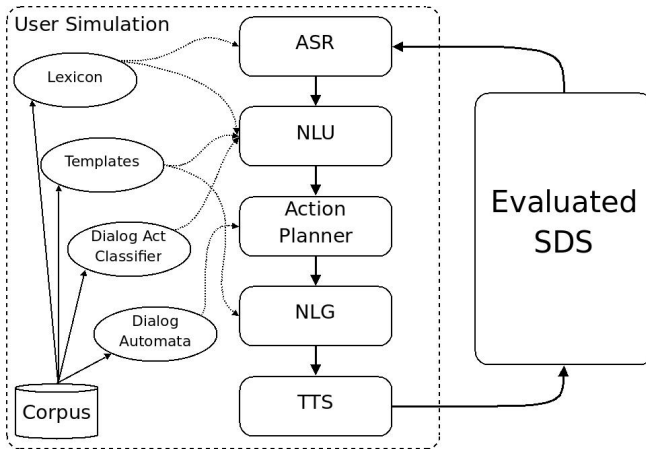
# Summary

- 1 VOICE Awards Human-Machine Dialog Corpus
- 2 Corpus-Assisted Creation of SDS Resources
- 3 Summary**





# Resources for a Dialog System



# Summary and Future Work

## Summary

- new German human-machine dialog corpus
- automatically acquire semi-general knowledge bases
  - domain and ontology terms
  - utterance templates
  - system state information / automata

## Current Work

- improve template extraction and NLU
- extract typical user behavior from the corpus
- further automatic annotations (slots, word-level timing)



# THANK YOU!

Comments, questions, corpus requests:  
**[tatjana.scheffler@dfki.de](mailto:tatjana.scheffler@dfki.de)**

The project SpeechEval is funded by the Investitionsbank Berlin  
under ProFIT grant #10140648.



This project is sponsored by the  
European Union  
(European Regional Development Fund)

