



**Conversational
Brains**

Classifying feedback communicative functions

Carol Figueroa

ESR 14

Aix-Marseille University | Furhat Robotics AB

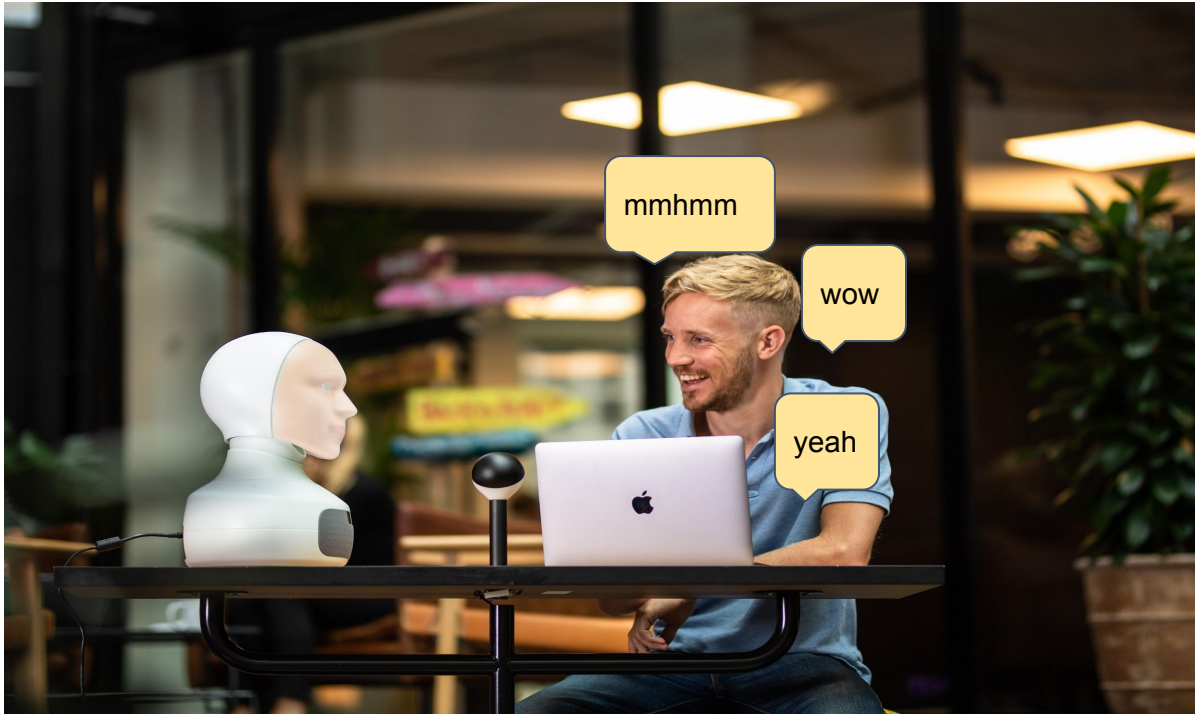
Magalie Ochs and Gabriel Skantze

carol@furhatrobotics.com

<http://www.cobra-network.eu/>



Use case: Recognize user's feedback



Use Case: Switchboard Corpus



- 2438 telephone conversations
- Each 3-10 minutes long
- 543 speakers from U.S.
- Audio and word level transcription included



Annotated Feedback in Switchboard













**Conversational
Brains**

Label	Function Description	Count
(C) Continue	Continue speaking. I hear you and I'm listening but not necessarily agreeing/disagreeing.	1024
(U) Non-understanding	I'm uncertain I understood/heard what you said.	63
(A) Agree	I agree with what you said.	435
(D) Disagree	I doubt what you said is true. I disagree with what you said.	46
(Y) Yes	I am giving a positive response/answer to your yes/no question.	56
(N) No	I am giving a negative response/answer to your yes/no question.	114
(S) Sympathy	I'm expressing sympathy/pity/sorrow/concern/compassion to a negative statement.	82
(Ds) Disapproval	I am showing disapproval/disgust.	65
(MS) Mild Surprise	I am showing mild surprise, showing slight interest.	103
(SS) Strong Surprise	I am showing strong surprise; I am impressed.	191
(O) Other	Not a feedback. Filler, listener trying to take turn.	77

85,956 potential feedback.
2,256 annotations, 2,179 are feedback.



Function labels in Grounding Framework

Grounding Level	Positive	Negative
<i>Specific</i>	A - Agree 	D - Disagree 
	Y- Yes 	N - No 
	S - Sympathy 	Ds - Disapproval 
	SS - Strong Surprise 	
	MS - Mild Surprise 	
<i>Generic</i>		
Understanding		
Perception	C - Continue	U - Non-understanding
Contact		



- Lexical: one-hot-encoding
- Prosodic:
 - Duration
 - Mean Pitch
 - Pitch Slope
 - Pitch Range
 - Mean Intensity

Context: 4000 ms of the preceding utterance of the interlocutor

- Part-of-speech: (POS) tags from spacy. POS bigrams created and sorted by their term frequency-inverse document frequency (TF-IDF).
- Dialog Acts: Used DialogTag, a Python library. One-hot encoding.
- Sentence Embedding: SimCSE, an auto-encoding embedding technique based on contrastive learning.



- One-shot
- Few-shot
- Fine-tune

GPT-3 probability distribution:

'Ag' 74%, 'Contin' 1.7%, 'Yes' 21%, 'agree' 3%, 'yes' 0.3%

Create a vector:

Agree 77%, Continue 1.7%, Yes-response 21.3%, all other 0%

Table 2: *Prompt given to GPT-3.*

GPT-3 Prompt

The following is a list of dialog acts and their description in parentheses:

- Continuer (Backchannel)
- Misunderstand (Expressing non-understanding)
- Agree (Agreeing with a statement)
- Disagree (Disagreeing with a statement)
- Yes-answer (A positive answer to a yes/no question)
- No-answer (A negative answer to a yes/no question)
- Sympathy (Expressing empathy)
- Reproach (Expressing disapproval or disgust or disappointment)
- Interest (Expressing interest)
- Surprise (Expressing surprise)
- Other (thinking or interrupting conversation)

The following is a dialog between two persons.

The dialog acts are written in brackets.

A: i was moving the lawn yesterday

B: mhm [**continuer**]



- `svm.SVC(kernel='linear', class_weight='balanced')`
- GPT-3 zero-shot
- GPT-3 few-shot
- GPT-3 fine-tuned

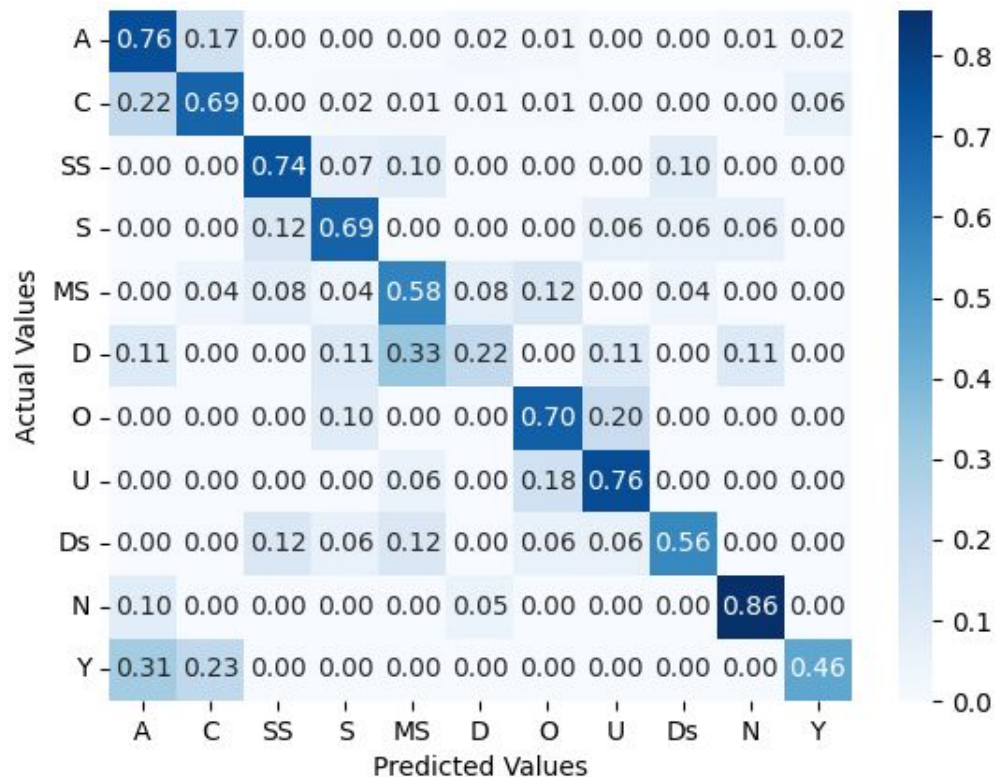
Table 3: *F1 weighted scores for different feature sets. *Uses GPT-3 (and not SVM) as the main classifier.*

Model #: Features	F-score
1: Lexical	0.63
Prosody	
2: Duration	0.10
3: Mean pitch	0.16
4: Pitch slope	0.24
5: Pitch range	0.18
6: Mean intensity	0.15
Context	
7: SimCSE	0.32
8: Dialog act (DA)	0.14
9: Part-of-speech (POS)	0.09
GPT-3	
10: Zero-shot majority*	0.61
11: Few-shot majority*	0.65
12: Fine-tuned*	0.80
13: Zero-shot as features (ZS)	0.61
14: Few-shot as features (FS)	0.63
Combinations	
15: Prosody (all)	0.37
16: Lexical + Prosody (LexPro)	0.63
17: Lexical + GPT-3 (ZS)	0.68
18: Lexical + GPT-3 (FS)	0.69
19: Lexical + SimCSE	0.72
20: LexPro + SimCSE + DA + GPT-3 (FS)	0.76
Majority-class baseline	0.28
Inter-annotator agreement	0.74

Model 19: Lexical + SimCSE



Conversational
Brains



Trained: 1804

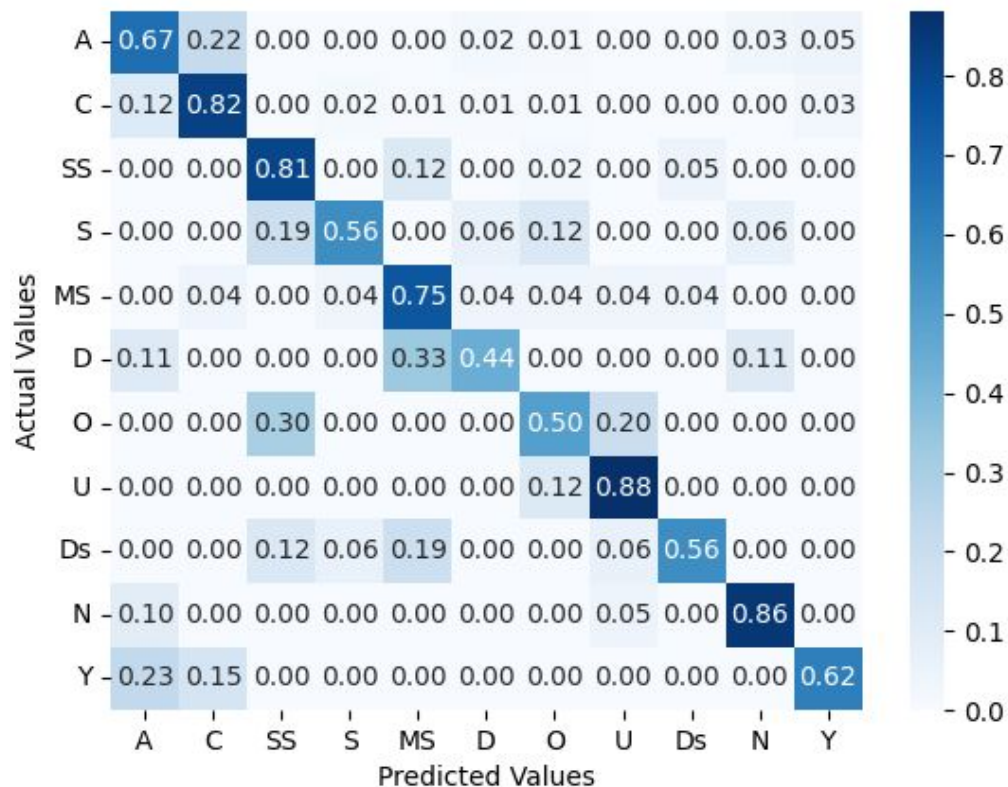
Tested: 452



Model 20: LexPro + SimCSE + DA + GPT-3 (FS)



**Conversational
Brains**



Model 12: Fine-tuned GPT-3

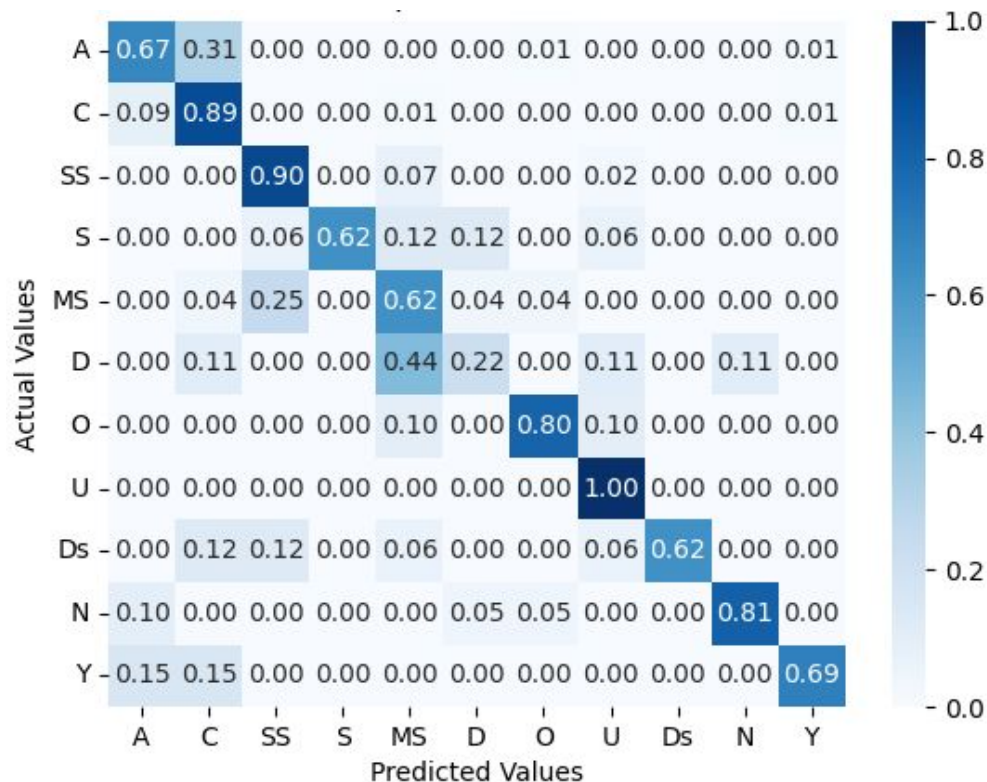


Table 4: *Distribution of automatically annotated tokens in the full Switchboard corpus, using Model 19. F=Female, M=Male.*

Function	Total	Total %	F %	M %
(C) Continue	38,475	46.0	48.8	42.6
(U) Non-understanding	279	0.3	0.3	0.4
(A) Agree	22,374	26.7	23.8	30.3
(D) Disagree	940	1.1	1.0	1.3
(Y) Yes-response	4,045	4.8	4.5	5.3
(N) No-response	673	0.8	0.8	0.8
(S) Sympathy	1,693	2.0	2.8	1.1
(MS) Mild surprise	2,222	2.7	2.7	2.6
(SS) Strong surprise	2,832	3.4	4.4	2.2
(Ds) Disapproval	573	0.7	0.7	0.7
(O) Other	9,594	11.5	10.4	12.7



- Just using lexical features and SimCSE gives fairly good performance, on par with inter-annotator agreement.
- Using GPT-3 in a zero-shot or few-shot fashion does not contribute much.
- A fine-tuned GPT-3 model outperforms all other models.



Thank you for your time!

