Karlsruhe Institute of Technology

Programming in Natural Language with *fu*SE:

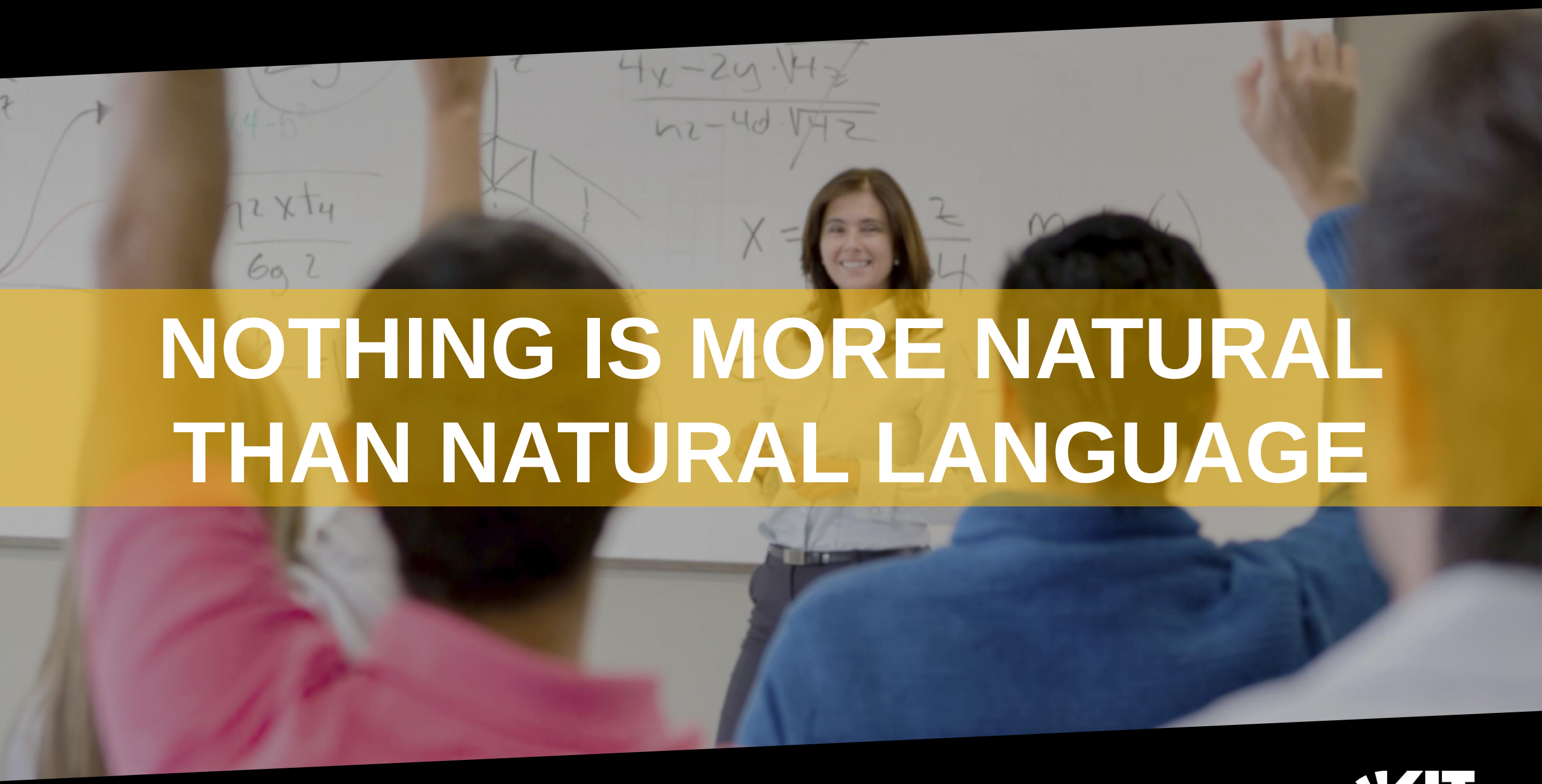
Synthesizing Methods from Spoken Utterances Using Deep Natural Language Understanding

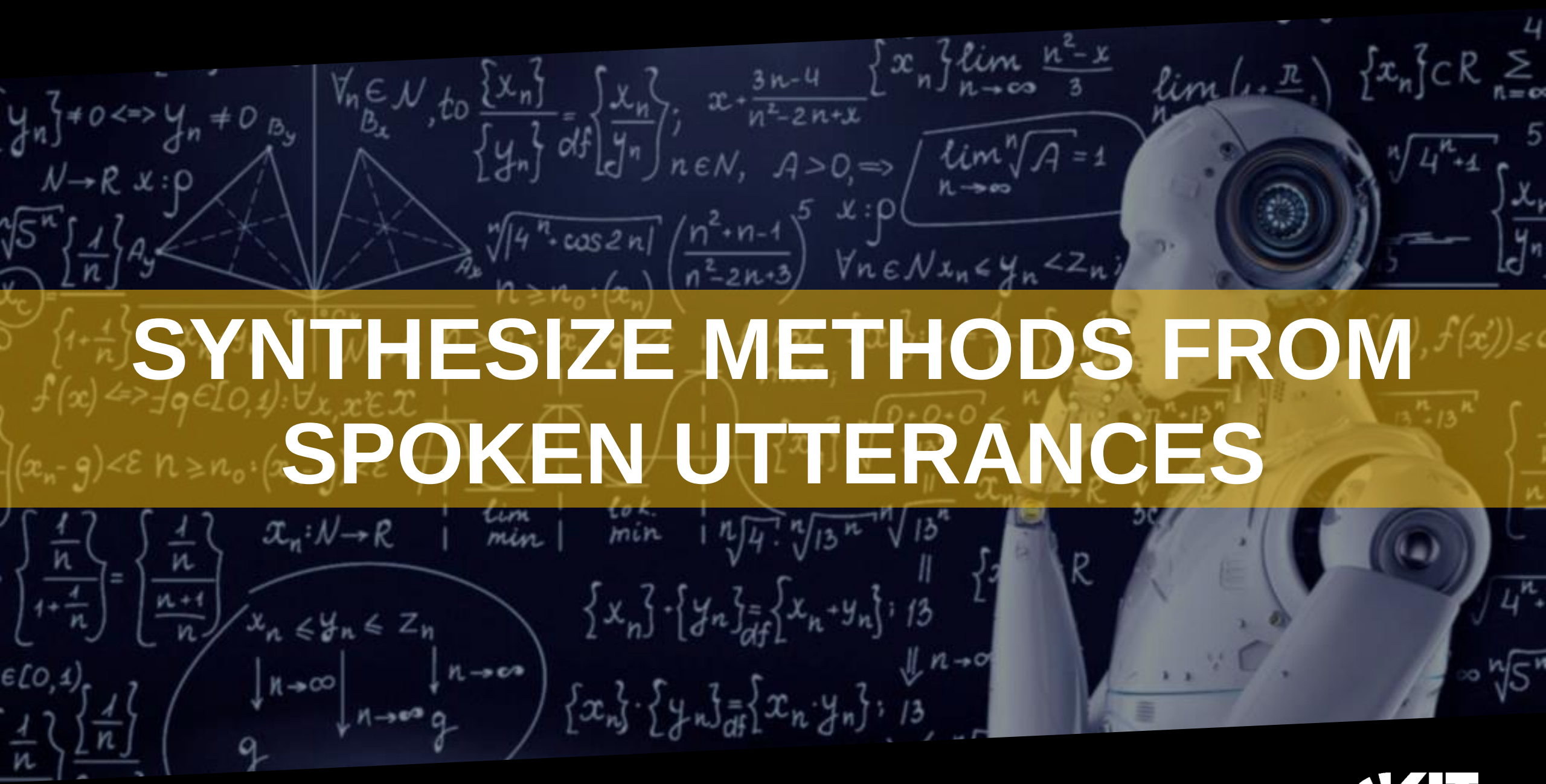
Sebastian Weigelt, Vanessa Steurer, Tobias Hey, and Walter F. Tichy

KIT – Department of Informatics – Institute for Program Structures and Data Organization (IPD).

INTELLIGENT SYSTEMS ARE ON THE RISE


TEACH AI YOURSELF


NOTHING IS MORE NATURAL
THAN NATURAL LANGUAGE



SYNTHESIZE METHODS FROM SPOKEN UTTERANCES

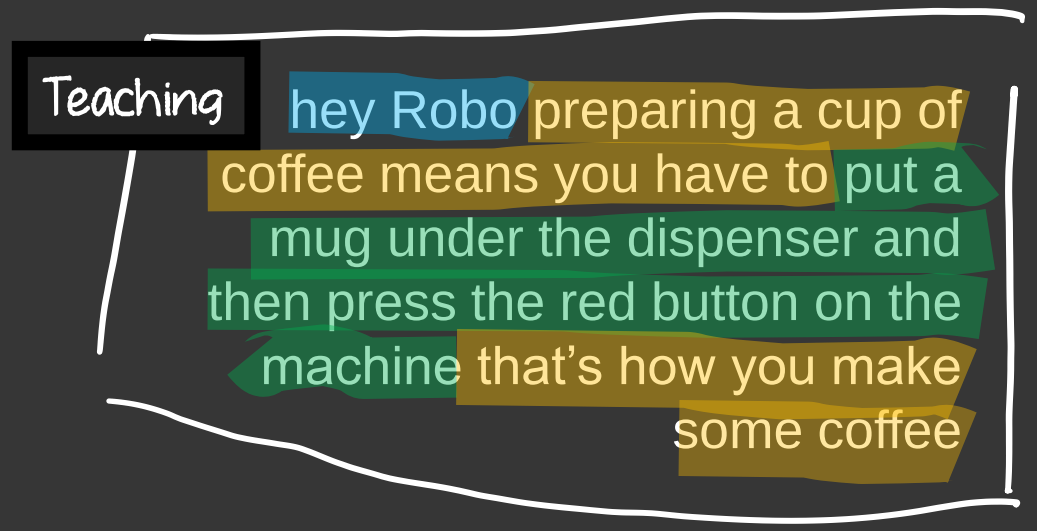
Task Definition

Synthesize Methods from Spoken Utterances

Given a natural language description,

we aim to classify whether it...

1. is a teaching effort or not,
2. extract the semantic structure and



Task Definition

Synthesize Methods from Spoken Utterances

Given a natural language description,

we aim to classify whether it...

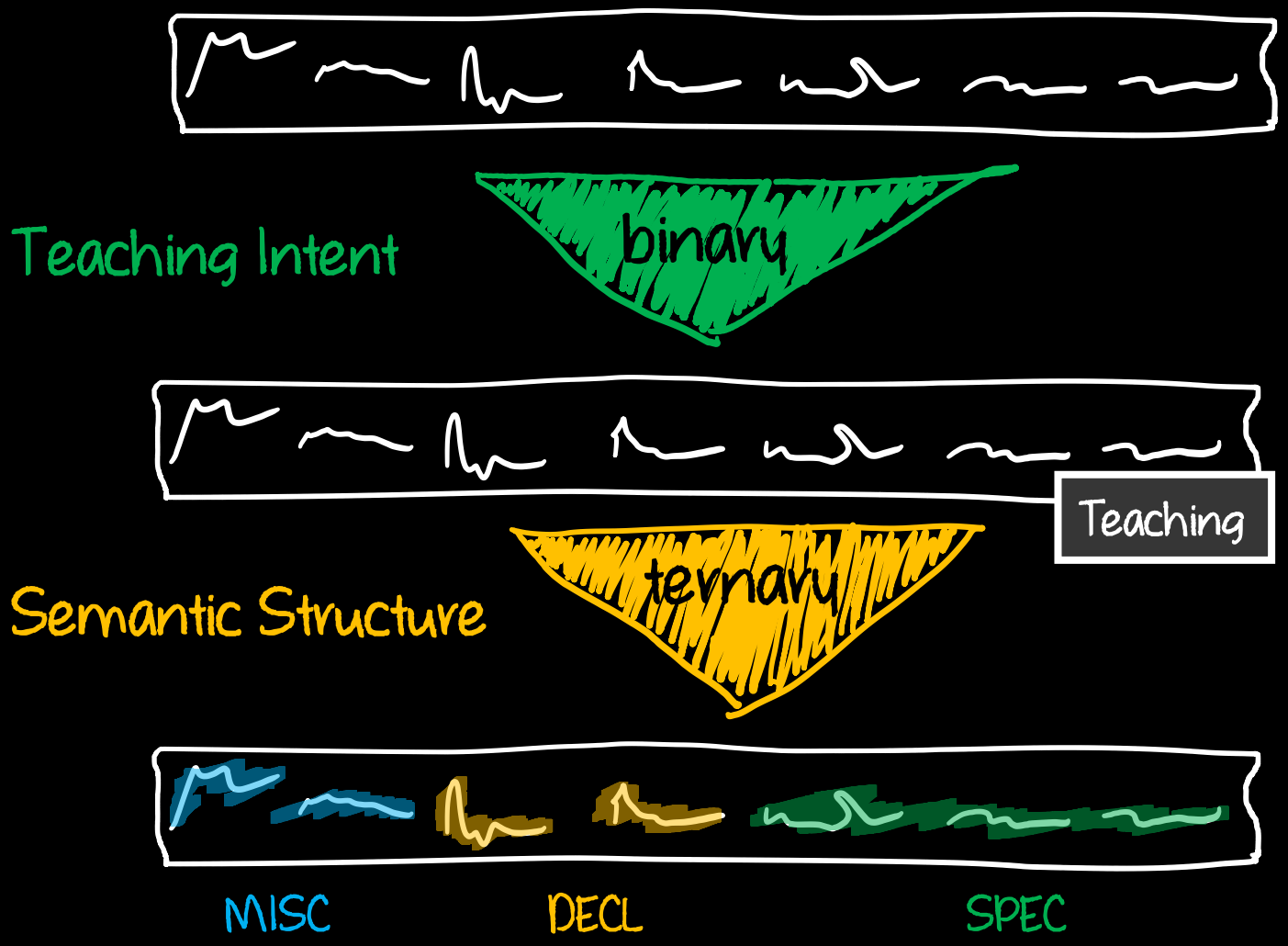
1. is a teaching effort or not,
2. extract the semantic structure and
3. synthesize the method signature and body

Teaching

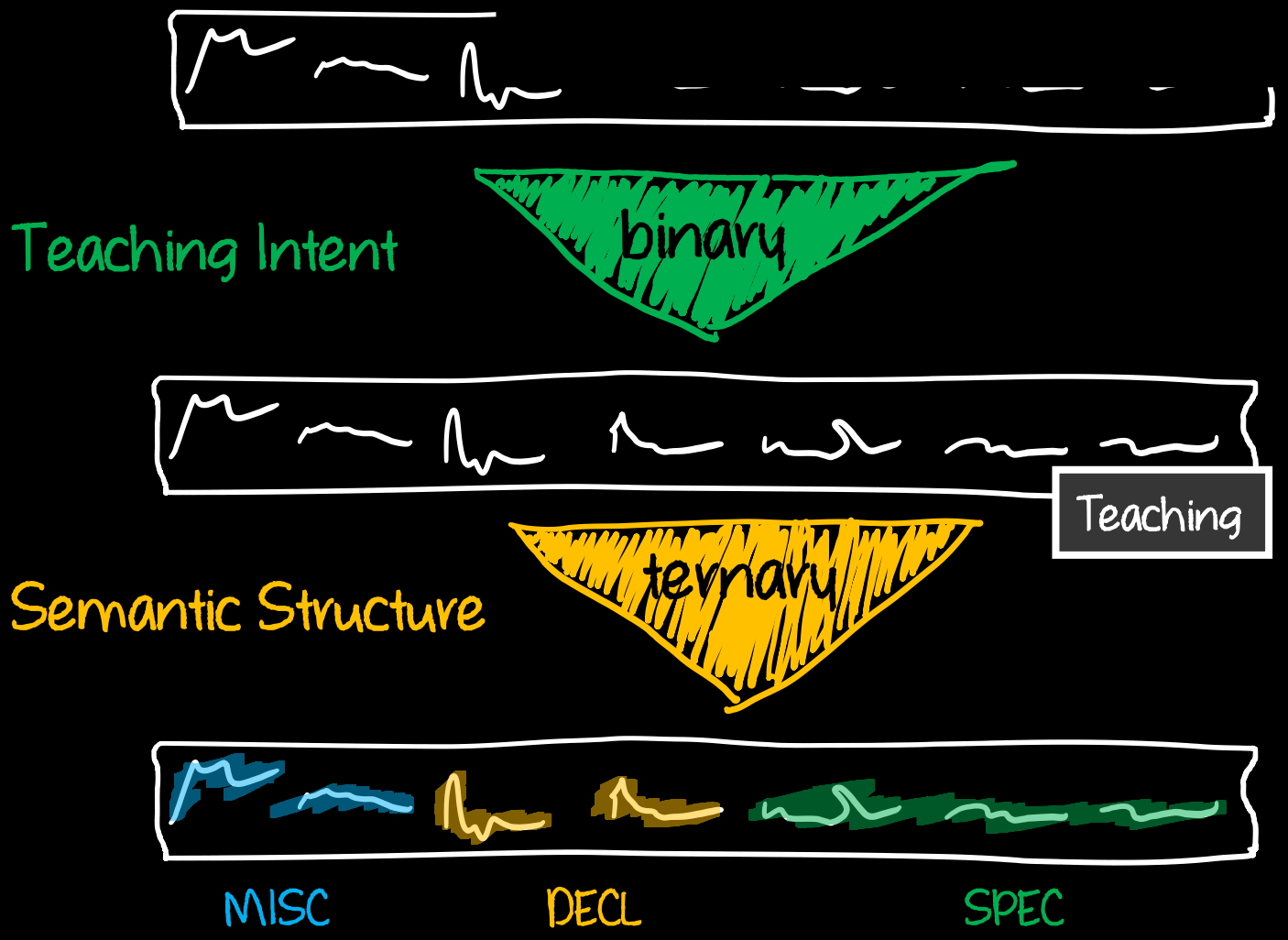
hey Robo preparing a cup of coffee means you have to put a mug under the dispenser and then press the red button on the machine that's how you make some coffee

```
procedure preapareCoffee()
    put(CoffeeCup, CoffeeMachine
        .Dispenser)
    press(CoffeeMachine.RedButton)
```

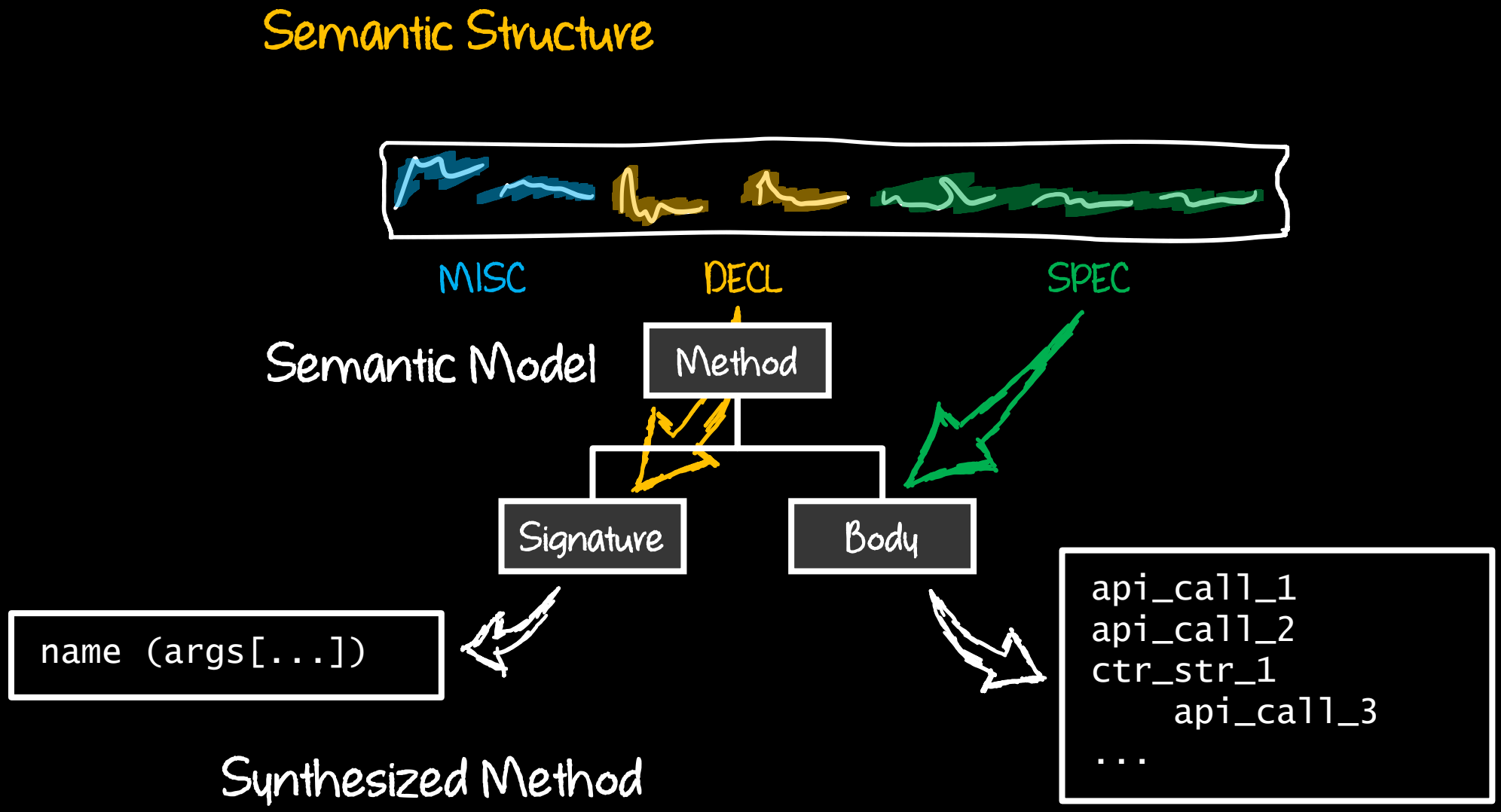
Approach – Overview



Approach – Overview



Approach – Overview



Dataset

Overview

Source: online user study

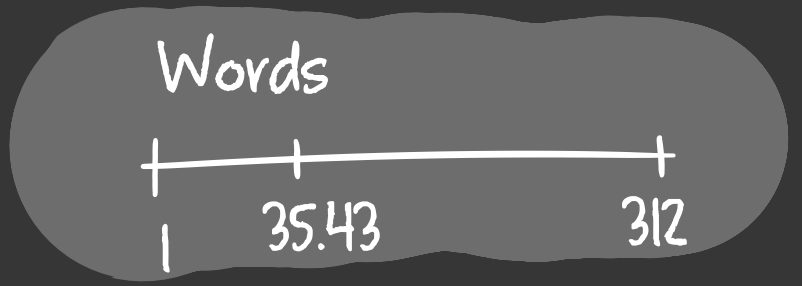
Task: teach a robot a skill using nothing but natural language

Setting: humanoid robot in a Kitchen

Scenarios: greeting someone
preparing coffee
serving drinks
setting a table for two

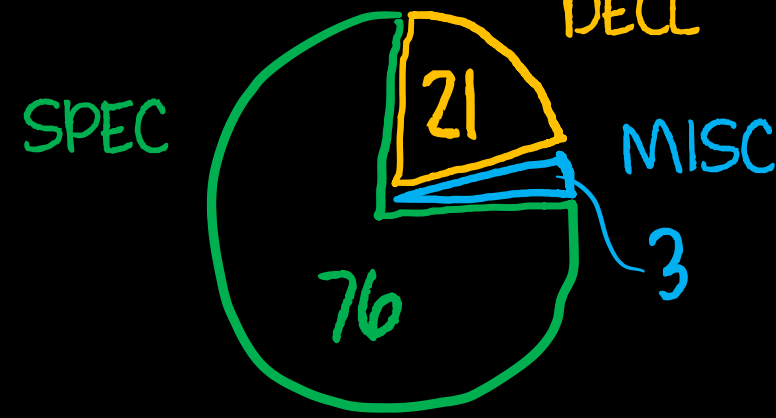
Dataset
http://dx.doi.org/10.21227/zecn-6c61

Numbers



Labels

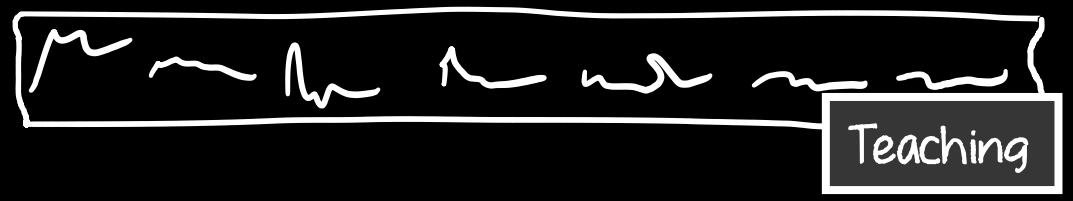
ternary (2nd classification)



binary (1st classification)

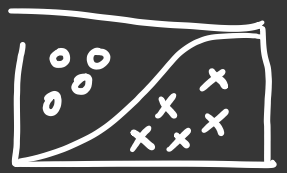
Approach – First-Level Classification: Overview

Task: is there a teaching intent or not?

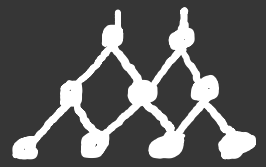


Sequence-To-Single-Label

Classifier



Classic



Neural Networks

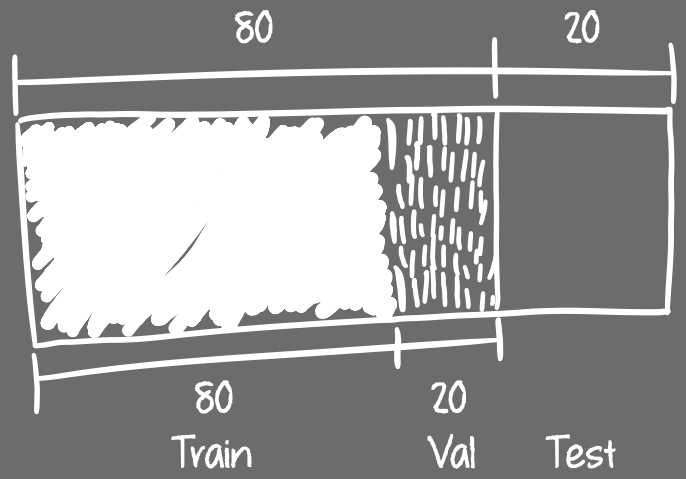
BERT

Challenge

teaching intent often stated implicitly

"You have to place the cup under the dispenser and press the red button to make coffee."

Random



Data Split

Train & Val

Test

Scenario-based

Approach – First-Level Classification: Results

Baseline (ZeroR)

Approach – First-Level Classification: Results

	Random
Baseline (ZeroR)	.573

Approach – First-Level Classification: Results

	Random
Baseline (ZeroR)	.573
Decision Tree	.903
Logistic Regression	.947

Approach – First-Level Classification: Results

	Random
Baseline (ZeroR)	.573
Decision Tree	.903
Logistic Regression	.947
CNN	.971
BiGRU	.959
BiLSTM	.959

Approach – First-Level Classification: Results

	Random
Baseline (ZeroR)	.573
Decision Tree	.903
Logistic Regression	.947
CNN	.971
BiGRU	.959
BiLSTM	.959
BERT, 10 epochs	.982
BERT, 300 epochs	.982

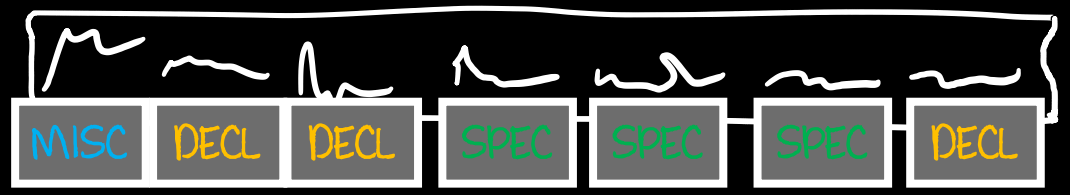
Approach – First-Level Classification: Results

	Random	Scenario
Baseline (ZeroR)	.573	.547
Decision Tree	.903	.719
Logistic Regression	.947	.719
CNN	.971	.874
BiGRU	.959	.932
BiLSTM	.959	.919
BERT, 10 epochs	.982	.973
BERT, 300 epochs	.982	.977

-.228

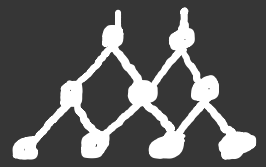
Approach – Second-Level Classification: Overview

Task: extract the semantic structure!



Sequence-To-Sequence

Classifier



Neural Networks

BERT

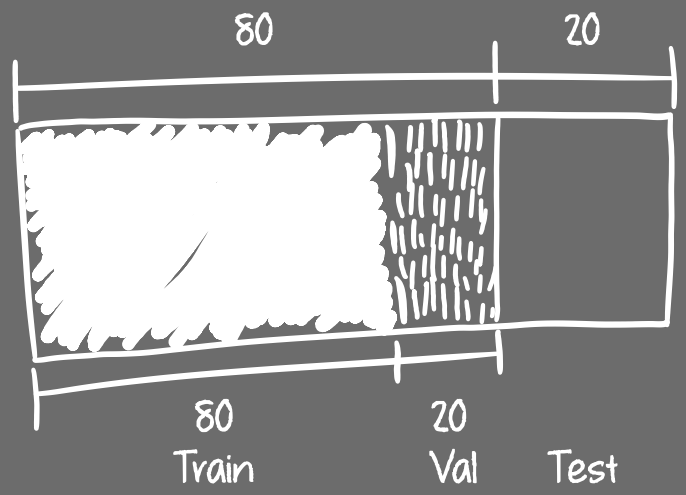
Challenge

non-continuous semantic parts

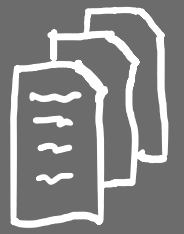
and varying structure

"hey robo look into the persons eyes to greet a person wave your robot hand and say hello this is how you greet someone that's it"

Random

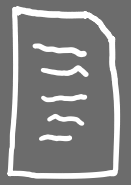


Data Split



Train & Val

Test



Scenario-based

Approach – Second-Level Classification

	Random	Scenario
Baseline (ZeroR)	.759	.757

Approach – Second-Level Classification

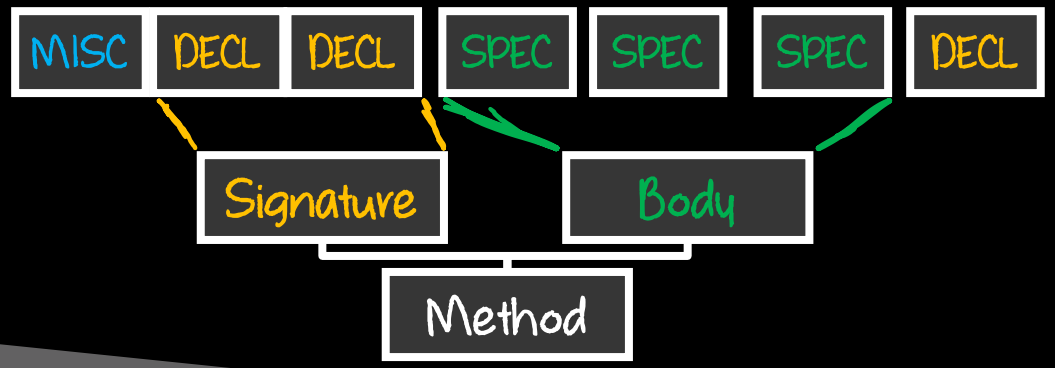
	Random	Scenario
Baseline (ZeroR)	.759	.757
BiLSTM	.985	.976
BiGRU	.988	.975

Approach – Second-Level Classification

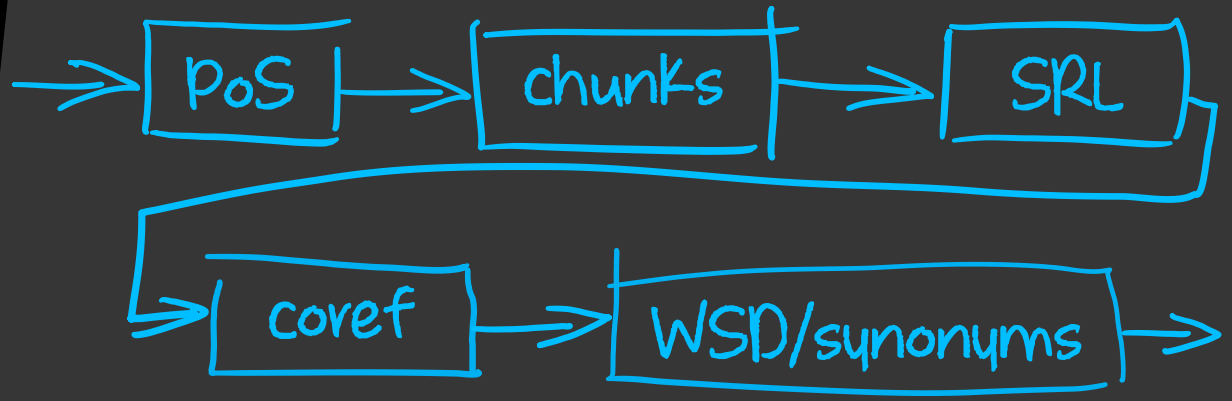
	Random	Scenario
Baseline (ZeroR)	.759	.757
BiLSTM	.985	.976
BiGRU	.988	.975
BERT, 10 epochs	.985	.972
BERT, 300 epochs	.983	.973

Approach – Third Stage: Method Synthesis

Task: Synthesize methods!



Pre-Processing



Approach

1. extract actions

```
make: coffee
put: cup, table
```

2. synthesize signature

```
prepare(what:
    Drinkable)
```

3. map actions to API calls

```
robo.place(CoffeeMug,
    KitchenTable)
```

Approach – Third Stage: 3 Mapping Actions to API calls

1 map single elements

lemmatize

remove stopwords

find synonyms

permutate

resolve coref.

MAP!

pressing, the button

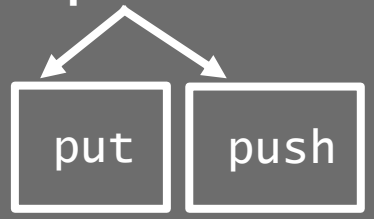
press, the button

press, button

[press, push],
[button, release]

...

push



[pressing] [the button of the coffee machine]

press	button of coffee machine

API::functions

- put
- push
- pour

API::arguments

- Dishwasher.Button
- CoffeeMachine.Button
- CoffeeMachine.Dispenser

Approach – Third Stage: 3 Mapping Actions to API calls

2 compose call candidates

determine formal parameters

check types (argument candidates)

fill in arguments

API::functions

- put
- push
- pour

API::arguments

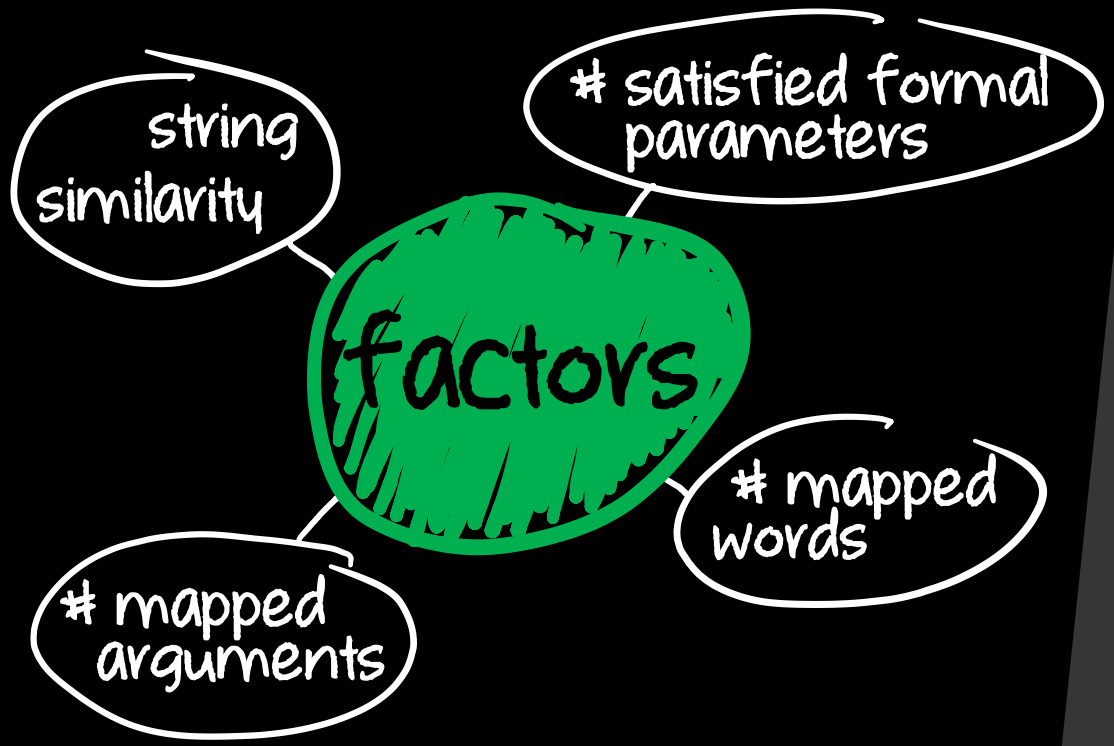
- Dishwasher.Button — Pushable
- CoffeeMachine.Button — Pushable
- CoffeeMachine.Dispenser — Object

```
put([redacted])
push(what::Pushable)
pour([redacted])
```

```
push(Dishwasher.Button)
push(CoffeeMachine.Button)
```

Approach – Third Stage: 3 Mapping Actions to API calls

3 rank candidates



one argument

pressing the button of the coffee machine

none

```
press()
```

one

```
push(Dishwasher.Button)
```

synonym for press

one

```
push(CoffeeMachine.Button)
```

result

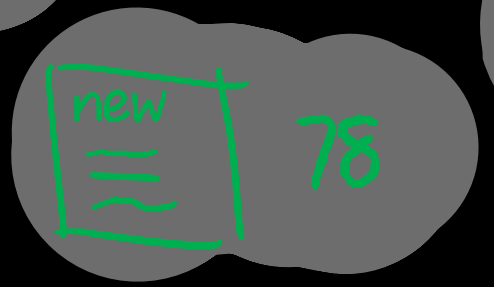
```
0.8    push(CoffeeMachine.Button)
0.5    push(Dishwasher.Button)
0.4    press()
...    ...
```

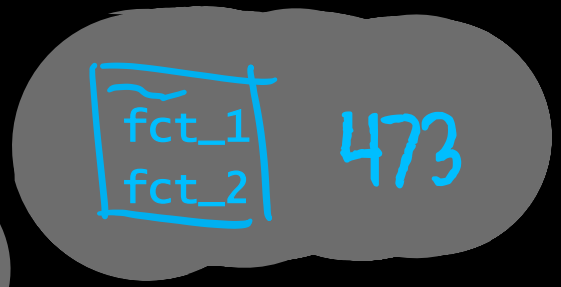
Evaluation

Dataset

2 new scenarios

- start dishwasher
- prepare cereals





Results

Signature (synthesize name & params)

Accuracy: 0.85

```
new startDishwasher()
```

VS

```
new startDishwasherHiRobo()
```

inaccurate

Body (map actions to API calls)

API calls

Pre	0.67
Rec	0.72
F_1	0.69

```
take (milk, fridge)
```

VS

```
take (milk, fridge)
```

single elements

Pre	0.87
Rec	0.93
F_1	0.91

Conclusion

Objective: Synthesizing methods from spoken utterances

Approach:

1. teaching intent classification
2. semantic structure classification
3. signature and body synthesis

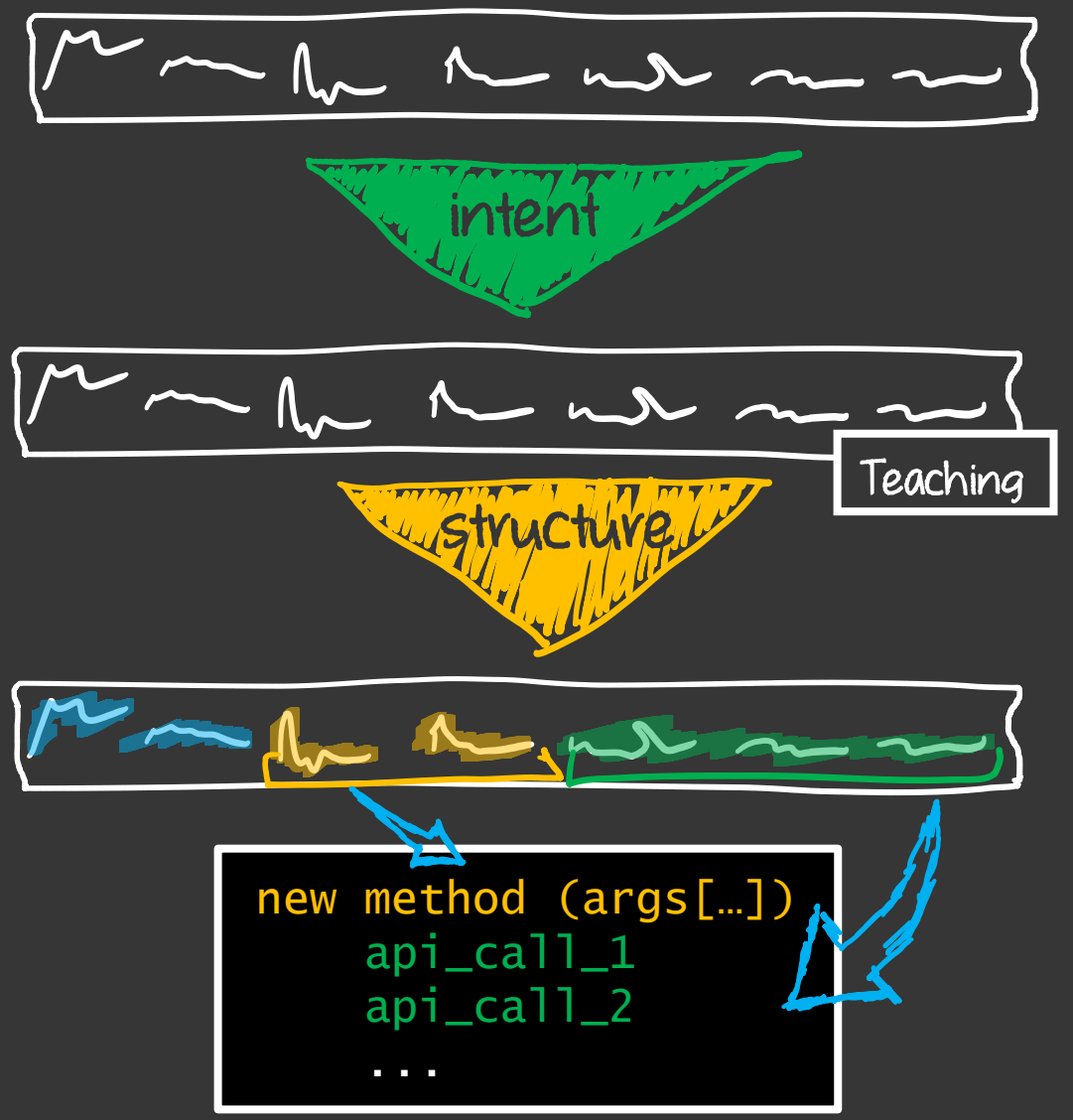
Results (accuracies):

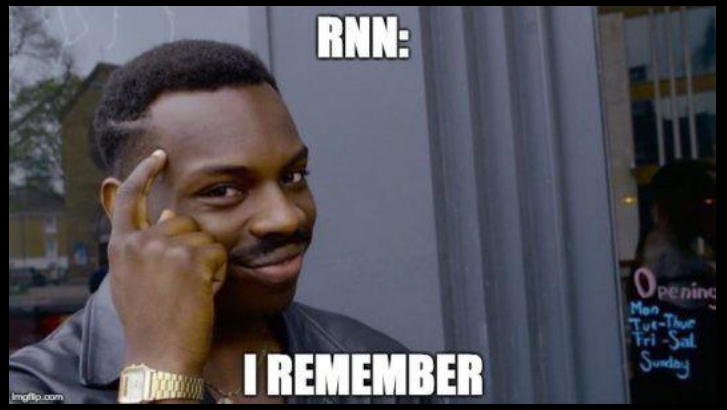
teaching intent .98

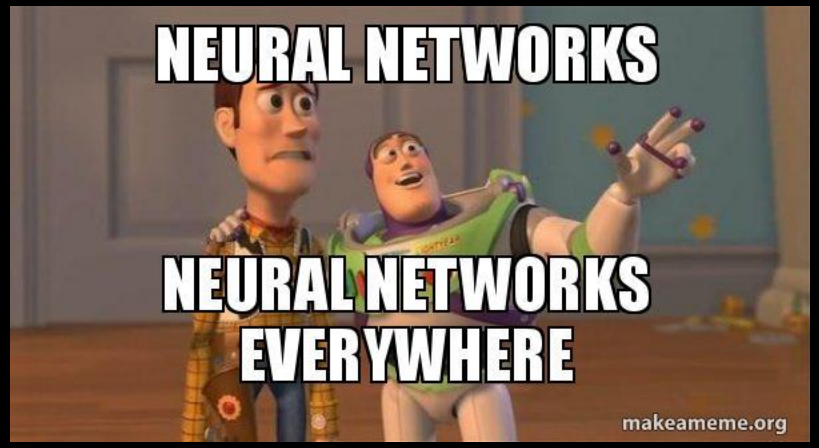
semantic structure .98

synthesis

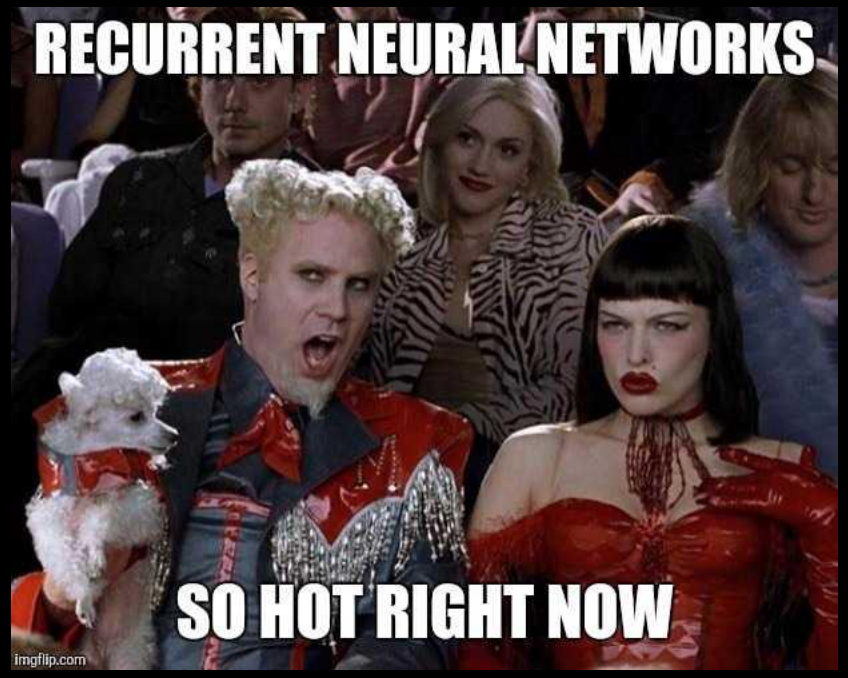
signature .85 body (F_1) .69




RNN:
I REMEMBER


NEURAL NETWORKS
NEURAL NETWORKS EVERYWHERE
makeameme.org


AND THEN I TOLD THEM
IT UNDERSTANDS NATURAL LANGUAGE
makeameme.org


RECURRENT NEURAL NETWORKS
SO HOT RIGHT NOW
imgflip.com


Questions?
meme-arsenal.ru


ONE DOES NOT SIMPLY
SAY THANK YOU WITHOUT A MEME
memes.com

Appendix – NN Configurations

types	architectures	additional layers	number of units	epochs	batch sizes	dropout values	learning rates
ANN		Flatten (Flat), Global max pooling 1D (GMax), Dense (D), Dropout(DO)	10, 20, 32, 40, 50, 64, 100, 128, 150, 250 256, 512	binary: 300, 500, 1000 ternary: 50, 100 300	binary: 50, 100, 300, 400 ternary: 32, 64, 100, 256, 300	0.1, 0.2, 0.3	0.001, 0.0005
CNN		Max pooling 1D (Max), Global max pooling 1D (GMax), Dense (D), Dropout(DO)					
RNN	LSTM GRU BiLSTM BiGRU	Dense (D), Dropout (DO)					