

Question classification in German dialogues

Eleni Adamantidou, Alexandros Paramythis

CONTEXTITY

Overview

Goal: Identify and classify questions in German dialogues (spoken or written)

Approach: Reuse / adapt dialogue-act classification approaches from English

Status: In-progress OS project, all input welcome!

Pipeline

Not annotated datasets - without punctuation

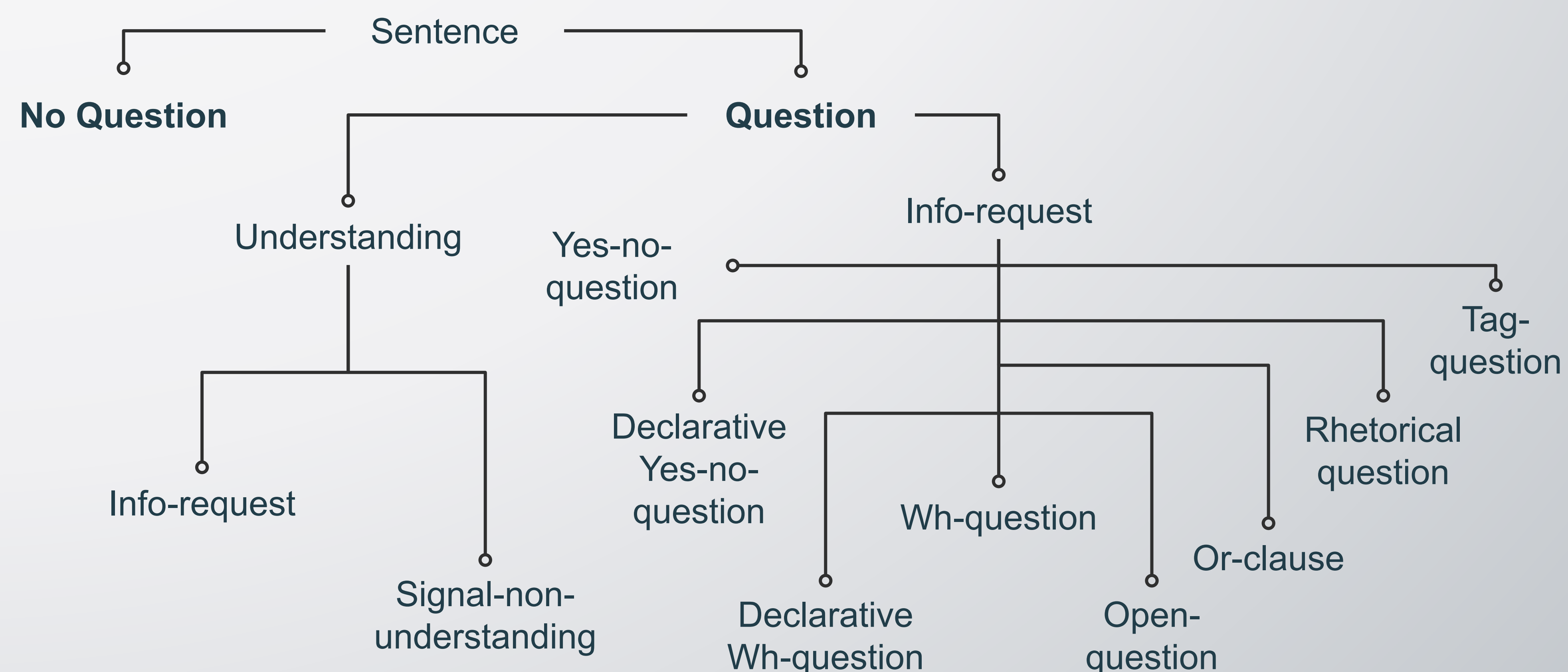
- Dialog structures
- Folk

Not annotated datasets - with punctuation

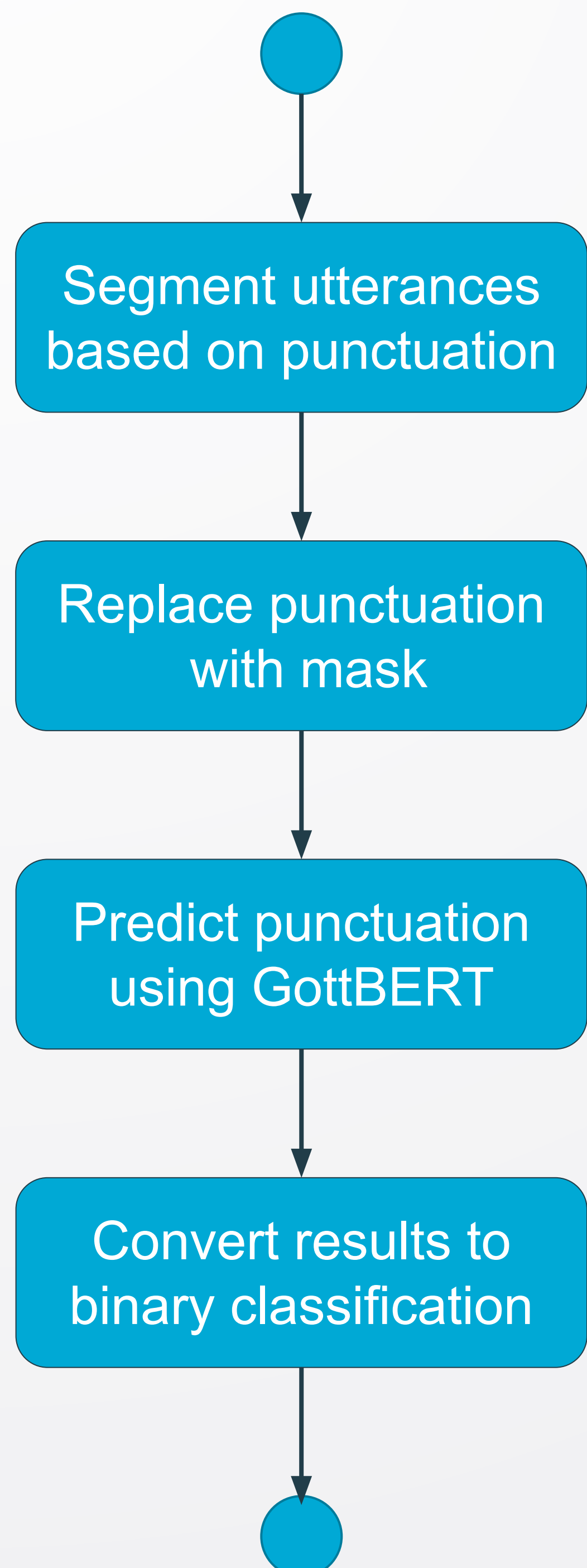
- Dortmund
- GermanQuAD

Fully annotated datasets

- Verbmobil



1st step: Is this a question, GottBERT?



Initial Utterances
Wie gefallen dir die Lieder von Bühne Huber, Soda und so weiter?
recht gut, kommen leider nicht an meine heran.....
Was hältst du von Verona???
toll, toll, toll!
was hältst du von katja (die, welche die fotos macht)?
eine superfrau!! kannst du mir ihre telefon-nummer geben?

Masked Utterances	Punct.	Simplified.	Pred. 1	Pred. 2
Wie gefallen dir die Lieder von Bühne Huber, Soda und so weiter <mask>	?	?	?	.
recht gut, kommen leider nicht an meine heran <mask>	!
Was hältst du von Verona <mask>	???	?	?	
toll, toll, toll <mask>	!	!		,
was hältst du von katja (die, welche die fotos macht) <mask>	?	?		</s>
eine superfrau <mask>	!!	!	</s>	
kannst du mir ihre telefon-nummer geben <mask>	?	?	?	.

Masked Utterances	Ground truth	Prediction
Wie gefallen dir die Lieder von Bühne Huber, Soda und so weiter <mask>	q	q
recht gut, kommen leider nicht an meine heran <mask>	nq	nq
Was hältst du von Verona <mask>	q	q
toll, toll, toll <mask>	nq	nq
was hältst du von katja (die, welche die fotos macht) <mask>	q	nq
eine superfrau <mask>	nq	nq
kannst du mir ihre telefon-nummer geben <mask>	q	q

Results - Confusion Matrix

		Actual class	
		Question	No question
Predicted class	Question	22% 7'892	2% 593
	No question	13% 4'677	63% 22'343

- Results shown here only for present (as opposed to assumed) punctuation
- Matrix refers to sentences with ≥ 3 tokens
- Only marginal improvement for ≥ 4 tokens
- Including assumed punctuation and all sentence sizes yields $\sim 90\%$ accuracy
- False negatives ambivalent also for humans

Ongoing work and latest results

<https://github.com/Contextity/question-classifier-german>