

CoQAR: Question Rewriting on CoQA.

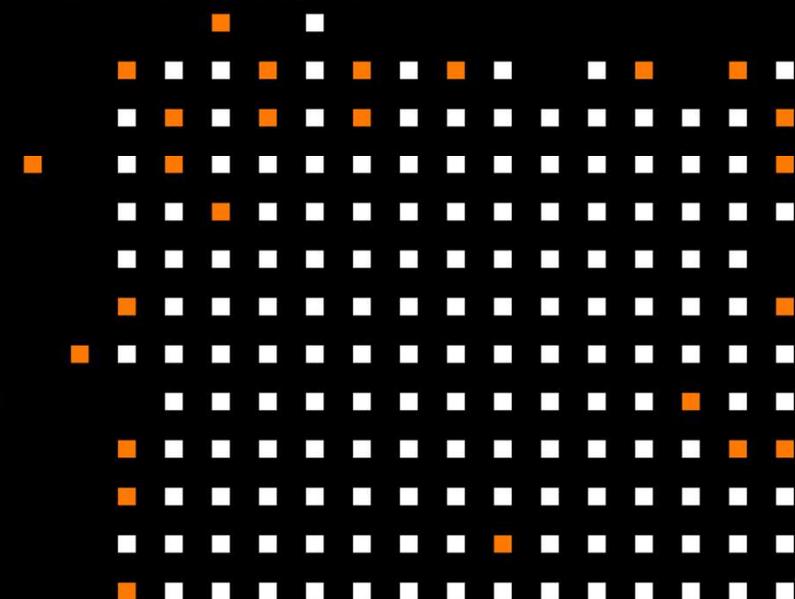
Quentin Brabant, Gwenole Lécorné and
Lina Rojas



NL4XAI School

Nancy, April 19th – 21st 2022

To be presented in LREC 2022



Outline:

1. Research at Orange.
2. Conversational QA and the corpus CoQA
3. What is Question Rewriting?
4. CoQAR: CoQA with Question Rewriting
5. CoQAR for Question Paraphrasing
6. CoQAR for Question Rewriting
7. Question Rewriting for (Conversational) QA
8. Conclusions



Orange : a world-leading telecommunications operator

142,000
employees



8th
largest global
telecoms brand

26 countries
in Europe, Africa
and the Middle-East

#1
for digital convergence in
Europe



Orange Research



1st

patents applicant
among European
operators on its fields of
activity

200

new inventions
protected by patent on
average per year

775

researchers in France and abroad

including

132

PhD and
post-doctoral

1

Scientific Council with 8
leading figures from the scientific

Orange is engaged in 17 sustainable development goals



What is Conversational QA?

Paragraph

Jessica went to sit in her rocking chair. Today was her birthday and she was turning 80. Her granddaughter Annie was coming over in the afternoon and Jessica was very excited to see her. Her daughter Melanie and Melanie's husband Josh were coming as well. Jessica had . . .



Conversation

Q₁: Who had a birthday?

R₁: Jessica went to sit in her rocking chair. Today was her birthday and she was turning 80.

Q₂: How old would she be?

[1] D. C. C. D. M. Siva Reddy, "CoQA: A Conversational Question Answering Challenge," Transactions of the Association for Computational Linguistics 7, 2019. Site :

<https://stanfordnlp.github.io/coqa/>

Conversational QA, the corpus CoQA

- Short Answers:

- A_1, \dots, A_n

- Extractive Answers:

- R_1, \dots, R_n



Jessica went to sit in her rocking chair. Today was her birthday and she was turning 80. Her granddaughter Annie was coming over in the afternoon and Jessica was very excited to see her. Her daughter Melanie and Melanie's husband Josh were coming as well. Jessica had . . .

Q₁: Who had a birthday?

A₁: Jessica

R₁: Jessica went to sit in her rocking chair. Today was her birthday and she was turning 80.

Q₂: How old would she be?

A₂: 80

R₂: she was turning 80

Q₃: Did she plan to have any visitors?

A₃: Yes

R₃: Her granddaughter Annie was coming over

Q₄: How many?

A₄: Three

R₄: Her granddaughter Annie was coming over in the afternoon and Jessica was very excited to see her. Her daughter Melanie and Melanie's husband Josh were coming as well.

Q₅: Who?

A₅: Annie, Melanie and Josh

R₅: Her granddaughter Annie was coming over in the afternoon and Jessica was very excited to see her. Her daughter Melanie and Melanie's husband Josh were coming as well.

CoQAR: Question Rewriting on CoQA

- **Annotate Questions** with out-of-context paraphrases
- **Why?**
 - to have more **diverse** data
 - to successfully train **generative-models** on Conversational QA (generate the question given the context) or for question rewriting.
 - to **reuse** classical QA systems.
 - annotations will be released as **Open Source** with baseline generative-models.



What is Question Rewriting ?

From In-Context
to Out-of-Context
Questions

passage	This is the story of a young girl and her dog. The young girl and her dog set out a trip into the woods one day. Upon entering the woods the girl and her dog found that the woods were dark and cold [...].
question	What is the story about?
rewritings	What is the subject of this story? Who are the two main characters in the story? Who is this story centered on?
answer	A girl and a dog.
answer span	This is the story of a young girl and her dog.
question	What were they doing?
rewritings	What were the girl and her dog up to? What did the girl and her dog decide to do? What was the activity of the girl and the dog for the day?
answer	Set on on a trip
answer span	The young girl and her dog set out a trip
question	where?
rewritings	Where did the girl and her dog go on a trip? What location did the girl and her dog journey to? What place did the girl and her dog go on that day?
answer	the woods
answer span	set out a trip into the woods



Question Rewriting Annotations

- ✓ ~~X~~ Crowdfworkers
- Three English native-speaker annotators.

Number of questions per rewritings

	Number of rewritings				
	0	1	2	3	total
train	365	108	31,378	13,210	45,061
dev	9	0	37	7,937	7,983

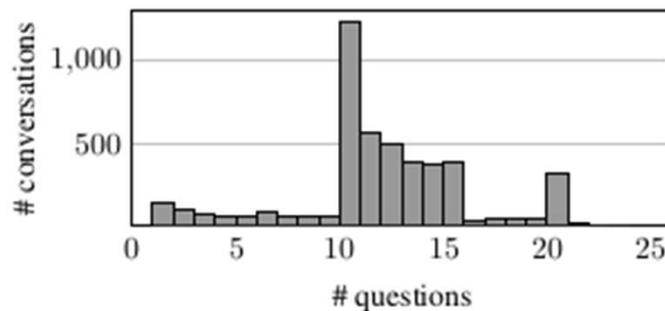
- 4.5 K conversations
- 53K question-answer pairs

Inter-Annotator Agreement:
SacreBLEU score of 32.67
and a BERT-score of 90.22

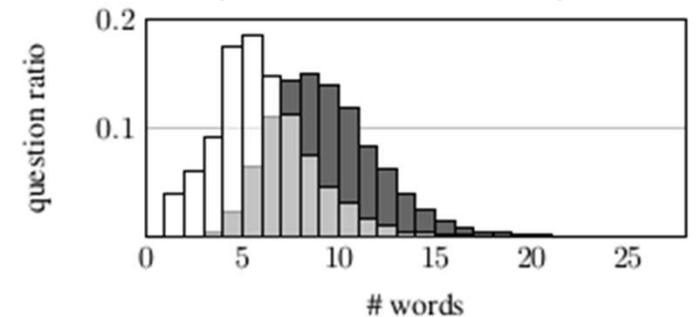


Diverse surface forms while
preserving the meaning

Histogram of conversation length

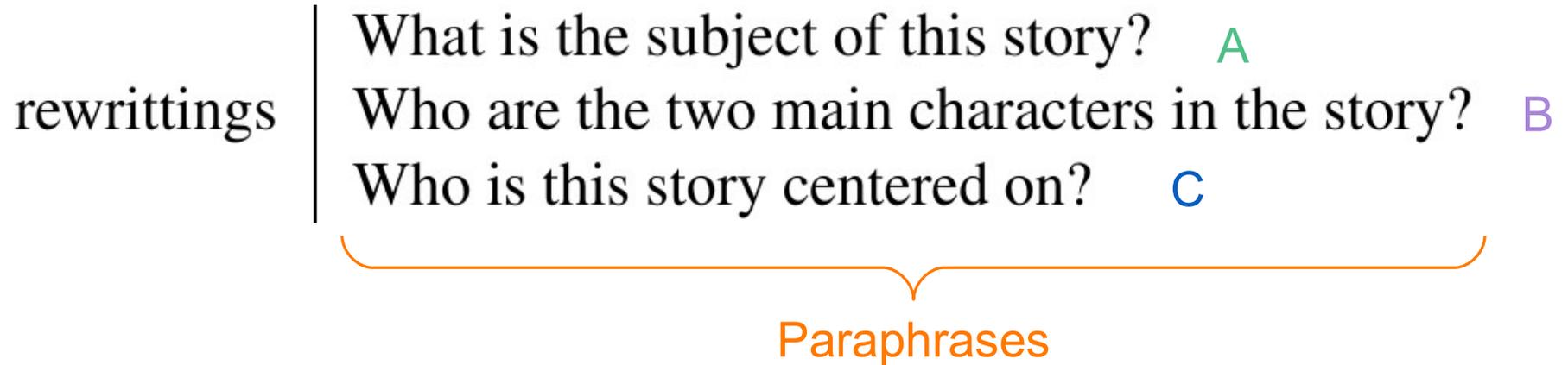


Histogram of question length



Original (white), Rewritings (dark grey).

CoQAR for Question Paraphrasing



- Apprentissage supervisé classique
- → Fine-tuning on BART

A → B B → A
A → C C → A
B → C C → B

CoQAR for Question Paraphrasing

3 versions of the model:

- BART(CoQAR), trained on CoQAR
- BART(QQP), trained on QQP
- BART(CoQAR+QQP), trained on both

Test set	Model	BLEU	METEOR
CoQAR	Naive (copy)	0.694	0.492
	BART(CoQAR)	0.705	0.537 OK
	BART(QQP)	0.673	0.464
	BART(CoQAR+QQP)	0.737	0.526
QQP	Naive (copy)	0.737	0.634
	BART(CoQAR)	0.626	0.445 ???
	BART(QQP)	0.695	0.619
	BART(CoQAR+QQP)	0.692	0.611

Table 3: BLEU and METEOR scores on the test of CoQAR and QQP for various models.

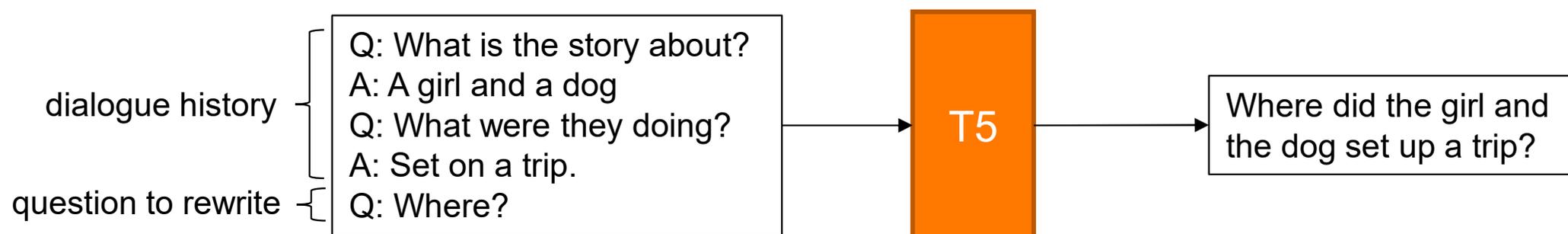
Test set	Model	Meaning preservation		Linguistic correctness		
		MOS	(Std dev.)	MOS	(Std dev.)	
CoQAR	Reference	3.82	(1.04)	4.46	(0.83)	Consistent with METEOR
	BART(CoQAR)	3.97	(1.15)	4.54	(0.75)	
QQP	Reference	3.32	(1.28)	4.33	(1.01)	QQP not a good paraphrase corpus
	BART(QQP)	3.64	(1.12)	4.37	(0.91)	
	BART(CoQAR+QQP)	3.65	(1.21)	4.51	(0.66)	
CANARD	BART(CoQAR)	4.15	(1.04)	4.41	(0.94)	CoQAR generalizes well

Table 4: Results of the human evaluation of QP.



CoQAR for Question Rewriting

Model:



Training:

- standard supervised learning
- examples from CoQAR and CANARD (a similar dataset)



CoQAR for Question Rewriting: Input-output pairs extraction

question	What is the story about?
rewritings	What is the subject of this story? Who are the two main characters in the story? Who is this story centered on?
answer	A girl and a dog.
answer span	This is the story of a young girl and her dog.
question	What were they doing?
rewritings	What were the girl and her dog up to? What did the girl and her dog decide to do? What was the activity of the girl and the dog for the day?
answer	Set on on a trip
answer span	The young girl and her dog set out a trip
question	where?
rewritings	Where did the girl and her dog go on a trip? What location did the girl and her dog journey to? What place did the girl and her dog go on that day?

Input	Output
Q: What is the story about? A: A girl and a dog. Q: What were they doing?	What were the girl and her dog up to?
Q: What is the story about? A: A girl and a dog. Q: What were they doing?	What did the girl and the dog decide to do?
Q: What is the story about? A: A girl and a dog. Q: What were they doing?	What was the activity of the girl and the dog for the day?
Q: What is the story about? A: A girl and a dog. Q: What were they doing? A: Set on a trip. Q: Where?	Where did the girl and her dog go on a trip?
(idem)	What location did the girl and her dog journey to?
(idem)	What place did the girl and her dog go on that day?



CoQAR for Question Rewriting : evaluation of models

3 versions of the model:

- T5(CoQAR), trained on CoQAR
- T5(CANARD), trained on CANARD
- T5(CoQAR+CANARD), trained on both

Test set	Model / train set	BLEU	METEOR
CoQAR	T5(CoQAR)	0.38	0.58
	T5(CANARD)	0.32	0.53
	T5(CoQAR+CANARD)	0.39	0.59
CANARD	T5(CoQAR)	0.31	0.57
	T5(CANARD)	0.47	0.69
	T5(CoQAR+CANARD)	0.44	0.66

Table 5: BLEU and METEOR scores obtained by the Question Rewriting models.

Test set	Model	Meaning preservation		Linguistic correctness	
		MOS	(Std dev.)	MOS	(Std dev.)
CoQAR	Human rewriting	4.5	(0.86)	4.86	(0.45)
	T5(CoQAR)	3.82	(1.42)	4.66	(0.82)
CANARD	Human rewriting	4.60	(0.96)	4.7	(0.89)
	T5(CANARD)	3.92	(1.34)	4.43	(1.08)
	T5(CoQAR+CANARD)	3.96	(1.47)	4.76	(0.77)

Table 6: Results of the human evaluation of QR.

Question Rewriting for Conversational QA

QR Mechanism	F1	EM
None(question+context)	68.13	49.63
Human rewriting	63.26	45.10
T5(CoQAR+CANARD)	63.30	44.97

Task: Extract the answer in the paragraph (span, unknown, yes, no)

Model : RoBERTa fine tuned

Dataset: CoQAR

RoBERTa seems to correctly solve coreference and ellipsis in short in-context questions/answers rather than when a complex long out-of-context question is given.



Question Rewriting for QA

Test set	QR mechanism	F1	EM	unknown
CoQAR	None (question+context)	35.89	8.22	0.5
	Human rewritings	42.21	9.23	0.4
	T5(CoQAR)	41.56	9.09	0.4
	T5(CANARD)	39.84	8.84	0.4
	T5(CoQAR+CANARD)	41.80	9.17	0.4
CANARD	None (question+context)	27.46	17.63	16.9
	Human rewritings	28.02	16.08	15.3
	T5(CANARD)	27.21	15.76	15.7
	T5(CoQAR)	27.06	16.44	14.9
	T5(CoQAR+CANARD)	27.18	16.03	15.3

Task: Extract the answer in the paragraph

Model : Hugging-face DistilBERT for SQUAD

Tested on : CoQAR and CANARD

Solving the context with both Human or T5 rewritings does improve QA predictions!

Conclusions

- Questions in 4.5 conversations of CoQA were annotated with out-of-context questions.
- Rich paraphrases in terms of diversity in the surface form.
- CoQAR seems more adapted to the task of Question Paraphrasing than QQP.
- Human evaluation shows that out-of-context paraphrases of CoQAR are as good as those of CANARD in terms of linguistic correctness and semantic similarity.
- Adding data from CoQAR to QQP or to CANARD during training does improve linguistic correctness.
- Question Rewriting performed either by humans or by models, does not improve the performance of conversational QA; it does enable the usage of non-conversational QA in CQA settings.