IndirectQA: Understanding Indirect Answers to Implicit Polar Questions in French and Spanish

LREC-COLING 2024



Christin Müller Barbara Plank 📤 🧐

MaiNLP lab, Center for Information and Language Processing, LMU Munich, Germany
 ^(a) Munich Center for Machine Learning (MCML), Munich, Germany
 ^(c) Department of Computer Science, IT University of Copenhagen, Denmark
 Christin.muller@googlemail.com, b.plank@lmu.de



1. Introduction

- Answers to yes/no-questions are often indirect
 - Normal in natural conversation
 - Difficult for dialogue systems to interpret

Example:

Question: "An island? By the sea?" Answer: "Islands are generally by the sea."

2. Related Work

- Current research focuses on datasets in English
 - Circa: Louis et al., 2020, <u>"I'd rather just go to bed": Understanding Indirect</u> <u>Answers</u>.
 - First large-scale corpus to interpret indirect answers
 - Friends-QIA: Damgaard et al., 2021, <u>"I'll be there for you": The One with</u> <u>Understanding Indirect Answers</u>.
 - More natural data extracted from existing TV series
 - SwDA-IA: Sanagavarapu et al., 2022, <u>Disentangling Indirect Answers to Yes-No</u> <u>Questions in Real Conversations</u>.
 - Included context, since ground truth depends on it

3. Contribution – IndirectQA

- a new, cross-lingual evaluation dataset with parallel data in English, French, and Spanish
- data extracted from the OpenSubtitles Corpus (Lison and Tiedemann, 2016)
- 1,053 question-indirect answer pairs extracted

4. IndirectQA

- Yes/no-questions and indirect answers from OpenSubtitles corpus (Lison et al., 2016), two different genres (comedy and crime, drama, mystery)
- Pros:
 - Parallel, aligned data
 - Variety of languages
 - Human translated
- Cons:
 - No turns
 - No cross-lingual alignment

```
You set it up weeks in advance and then you cancel three times.
       <time id="T34E" value="00:04:17,483" />
198
      </s>
199
      <s id="46">
200
       <time id="T35S" value="00:04:17,567" />
201
     Shouldn't they leave you alone and give you some space out of respect for our situation?
      </s>
203
     <s id="47">
204
    Isn't that what we need right now?
205
       <time id="T36E" value="00:04:23,907" />
206
      </s>
207
      <s id="48">
208
       <time id="T37S" value="00:04:23.990" />
209
    No, we need people who know what we need, because we don't.
210
       <time id="T37E" value="00:04:27,535" />
211
      \langle s \rangle
212
      <s id="49">
213
       <time id="T38S" value="00:04:27,618" />
214
    That's why they're called friends.
215
       <time id="T38E" value="00:04:29.662" />
216
      </s>
217
      <s id="50">
218
       <time id="T39S" value="00:04:29,746" />
219
    Well, I'm not sure that's why they're called friends.
220
       <time id="T39E" value="00:04:32.332" />
221
      </s>
222
      <s id="51">
223
       <time id="T40S" value="00:04:32,416" />
224
225 It's a German word, I think.
       <time id="T40E" value="00:04:34,792" />
226
227
      </s>
      <s id="52">
228
       <time id="T41S" value="00:04:34,876" />
229
    I can't recall the origin, I took a course ...
230
```

Language	Number	Year	Movie-ID	Doc-ID	Genre	Sentence-ID (Start Question)	Question	Answer-ID (Start Answer)	Answer
English	0	2015	3591512	6461786	Crime, Drama, Mystery	<s id="22"></s>	The coast?	<s id="23"></s>	Audrey's been telling me about it from her magazines - haven't you, Audrey?
	1	2015	3591512	6461786	Crime, Drama, Mystery	<s id="23"></s>	Audrey's been telling me about it from her magazines - haven't you, Audrey?	<s id="24"></s>	Some Hollywood film star was supposed to have bought it, but no, it's Mr and Mrs Owen.
	2	2015	3591512	6461786	Crime, Drama, Mystery	<s id="27"></s>	An island? By the sea?	<s id="29"></s>	Islands are generally by the sea.
French	0	2015	3591512	6442019	Crime, Drama, Mystery	<s id="19"></s>	La côte ?	<s id="20"></s>	Audrey l'a lu dans ces magazines, n'est-ce pas, Audrey ?
	1	2015	015 3591512 6442019 Crime, Drama, Mystery		<s id="20"></s>	Audrey l'a lu dans ces magazines, n'est-ce pas, Audrey ?	<s id="21"></s>	Une star d'Hollywood était censé l'avoir acheté, mais non, M. et Mme Owen l'ont fait.	
	2	2015	3591512	6442019	Crime, Drama, Mystery	<s id="24"></s>	Une île ? Près de la mer ?	<s id="26"></s>	Les îles sont généralement près de la mer.
Spanish	0	2015	3591512	6439216	Crime, Drama, Mystery	<s id="19"></s>	¿La costa?	<s id="20"></s>	Audrey me ha estado hablando de ella por las revistas, ¿a que sí, Audrey?
	1	2015	3591512	6439216	Crime, Drama, Mystery	<s id="20"></s>	Audrey me ha estado hablando de ella por las revistas, ¿a que sí, Audrey?	<s id="22"></s>	Se suponía que una estrella de Hollywood iba a comprarla, pero no, el señor y la señora Owen.
	2	2015	3591512	6439216	Crime, Drama, Mystery	<s id="25"></s>	¿Una isla? ¿En el mar?	<s id="27"></s>	Las islas generalmente están en el mar.

4. IndirectQA – Annotation

/

Strict Label Set	Relaxed Label Set				
Yes	Yes				
No	No				
Probably yes / sometimes yes	-				
Yes, subject to some conditions	Yes, subject to some conditions				
Probably no	-				
In the middle, neither yes nor no	In the middle, neither yes nor no				
Other	Other				
N/A	N/A				
Labelset for Annotation by Louis et al. (2020)					
	Lacking context				

4. IndirectQA – Distribution

Language	Extracted #	Final # QIA pairs	Label 1	Label 2	Label 3	Label 4	Label 5	Label 6
English	615	all (615)	35.61	13.01	1.95	12.36	24.39	12.52
	292	Crime (292)	35.62	13.36	1.37	15.75	21.92	11.99
	323	Comedy (323)	35.60	12.69	2.48	9.29	26.93	13.00
French	239/205	all (438)	34.93	14.16	1.60	11.42	24.20	13.47
Spanish	239/205	all (438)	35.16	14.16	1.60	11.19	24.20	13.70

Table 2: Final label distribution of the IndirectQA dataset (in percentages). Total amount (#) of QIA pairs extracted for the genres comedy and crime, drama, mystery (abbrev. crime), indicated as x/x, respectively. All refers to the total of QIA pairs extracted (both genres combined). Due to data cleaning, the numbers in the Final QIA pairs row vary; those are the QIA pairs used for evaluation. Label numbers correspond to the descriptions provided in Section 3.2.

5. Method and Model Setup

- Cross-lingual Transfer Learning
 - Intermediate Task Training: additional pretraining on labeled data (Phang et al., 2020)
 - Sequential setup
 - mBERT: good option for zero-shot cross-lingual transfer tasks, inlcuding classification (Pires et al., 2019)
- Implementation (both intermediate task training and baseline models):
 - Classification task with MaChAmp Toolkit (van der Goot et al., 2021)
 - Version 0.4, 20 epochs, mBERT encoder



Model Overview using MaChAmp. Figure adapted from Phang et al. (2020) Intermediate task data: MNLI (Inference Dataset, Williams et al., 2018)

6. Models

- Baseline Model
 - FRIENDS-QIA + IndirectQA combined
- Intermediate Task Training Models
 - MNLI
 - XNLI (FR and ES)

6. Models – Overview and Results

Model	Model Name	Training Data		Test	Acc.	F1
English		FRIENDS-QIA + IndirectQA (en)		en	61.43	37.68
Baseline	FRIENDS-QIA + IndirectQA	FRIENDS-QIA + IndirectQA (en)	fr	44.06	33.64	
				es	44.06	33.95
Model	Model Name	Intermediate Task	Target Task	Test	Acc.	F1
Intermediate	MNLI-FX	MNLI (en)	FRIENDS-QIA + IndirectQA (en)	fr	50.00	40.06
		MNLI (en)	FRIENDS-QIA + IndirectQA (en)	es	42.92	36.44
	XNLI-FR-FX	MNLI (en) + XNLI (fr)	FRIENDS-QIA + IndirectQA (en)	fr	44.06	35.97
		MNLI (en) + XNLI (fr)	FRIENDS-QIA + IndirectQA (en)	es	42.01	34.55
	XNLI-ES-FX	MNLI (en) + XNLI (es)	FRIENDS-QIA + IndirectQA (en)	fr	45.21	38.43
		MNLI (en) + XNLI (es)	FRIENDS-QIA + IndirectQA (en)	es	42.92	36.92

Table 3: Performance scores (Acc.=Accuracy, F1=Macro F1) of the baseline and intermediate task training models evaluated on French and Spanish data (EN given as reference). Green cells indicate improvement over the respective baseline, red drop below the baseline scores, no color is identical to the baseline score.

6. Models – Results per Genre

Lang	Genre	Acc.	F1	Genre	Acc.	F1
fr	Comedy	38.72	28.95	Crime	50.25	38.98
es	Comedy	36.17	27.71	Crime	53.20	41.13

Table 4: Results per genre, comedy vs crime (crime, drama, mystery).

6. Models – Results per Label

Model	Model Name	Test Data	F1/ Label	F1/ Label	F1/ Label	F1/ Label	F1/ Label	F1/ Label
Baseline			1/Yes	2/ No	3/ Yes (condition)	4/ Neither	5/ Other	6/ Context
FRIENDS-QIA + IndirectQA	FRIENDS-QIA + IndirectQA	en	75.15	62.20	20.00	43.70	25.00	00.00
		fr	61.00	44.59	00.00	28.25	32.39	35.61
		es	65.54	42.60	00.00	21.62	35.46	38.46
Intermediate Task Training								
MNLI/ XNLI	MNLI-FX	en	72.29	52.77	32.00	47.28	19.57	00.00
		fr	66.47	44.93	14.29	36.99	38.71	38.96
		es	59.04	45.86	33.33	22.82	34.39	23.19
	XNLI-FR-FX	en	70.55	55.72	26.09	48.44	30.11	00.00
		fr	58.69	41.72	22.22	24.79	35.06	33.33
		es	54.72	47.46	20.00	23.02	36.13	25.97
	XNLI-ES-FX	en	74.46	57.98	30.77	47.67	19.75	00.00
		fr	60.98	44.93	22.22	28.57	34.90	38.96
		es	57.04	44.83	16.67	25.84	38.67	38.46

Table 6: F1-scores per label for the intermediate task training models. Green cells indicate higher performance than the baseline.

7. Discussion and Conclusion

- Still a quite challenging task
- Subtitles as a source are somewhat difficult to work with
- Contribution: IndirectQA, a new, cross-lingual evaluation dataset of 1,053 question-indirect answer pairs with parallel data in English, French, and Spanish (available at https://github.com/mainlp/indirectQA)
- Presentable results regarding the MNLI/XNLI setup

8. References

Cathrine Damgaard, Paulina Toborek, Trine Eriksen, and Barbara Plank. 2021. <u>"I'll be there for you": The One with Understanding Indirect</u> <u>Answers</u>. In *Proceedings of the 2nd Workshop on Computational Approaches to Discourse*, pages 1–11, Punta Cana, Dominican Republic and Online. Association for Computational Linguistics.

Pierre Lison and Jörg Tiedemann. 2016. <u>OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles</u>. In *Proceedings* of the Tenth International Conference on Language Resources and Evaluation (LREC'16), pages 923–929, Portorož, Slovenia. European Language Resources Association (ELRA).

Annie Louis, Dan Roth, and Filip Radlinski. 2020. <u>"I'd rather just go to bed": Understanding Indirect Answers</u>. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7411–7425, Online. Association for Computational Linguistics.

Jason Phang, lacer Calixto, Phu Mon Htut, Yada Pruksachatkun, Haokun Liu, Clara Vania, Katharina Kann, and Samuel R. Bowman. 2020. <u>English Intermediate-Task Training Improves Zero-Shot Cross-Lingual Transfer Too</u>. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 557–575, Suzhou, China. Association for Computational Linguistics. Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. <u>How Multilingual is Multilingual BERT?</u>. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4996–5001, Florence, Italy. Association for Computational Linguistics.

Krishna Sanagavarapu, Jathin Singaraju, Anusha Kakileti, Anirudh Kaza, Aaron Mathews, Helen Li, Nathan Brito, and Eduardo Blanco. 2022. <u>Disentangling Indirect Answers to Yes-No Questions in Real Conversations</u>. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4677–4695, Seattle, United States. Association for Computational Linguistics.

Rob van der Goot, Ahmet Üstün, Alan Ramponi, Ibrahim Sharaf, and Barbara Plank. 2021. <u>Massive Choice, Ample Tasks (MaChAmp): A</u> <u>Toolkit for Multi-task Learning in NLP</u>. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations, pages 176–197, Online. Association for Computational Linguistics.

Adina Williams, Nikita Nangia, and Samuel Bowman. 2018. <u>A Broad-Coverage Challenge Corpus for Sentence Understanding through</u> <u>Inference</u>. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1112–1122, New Orleans, Louisiana. Association for Computational Linguistics.

Thank you!