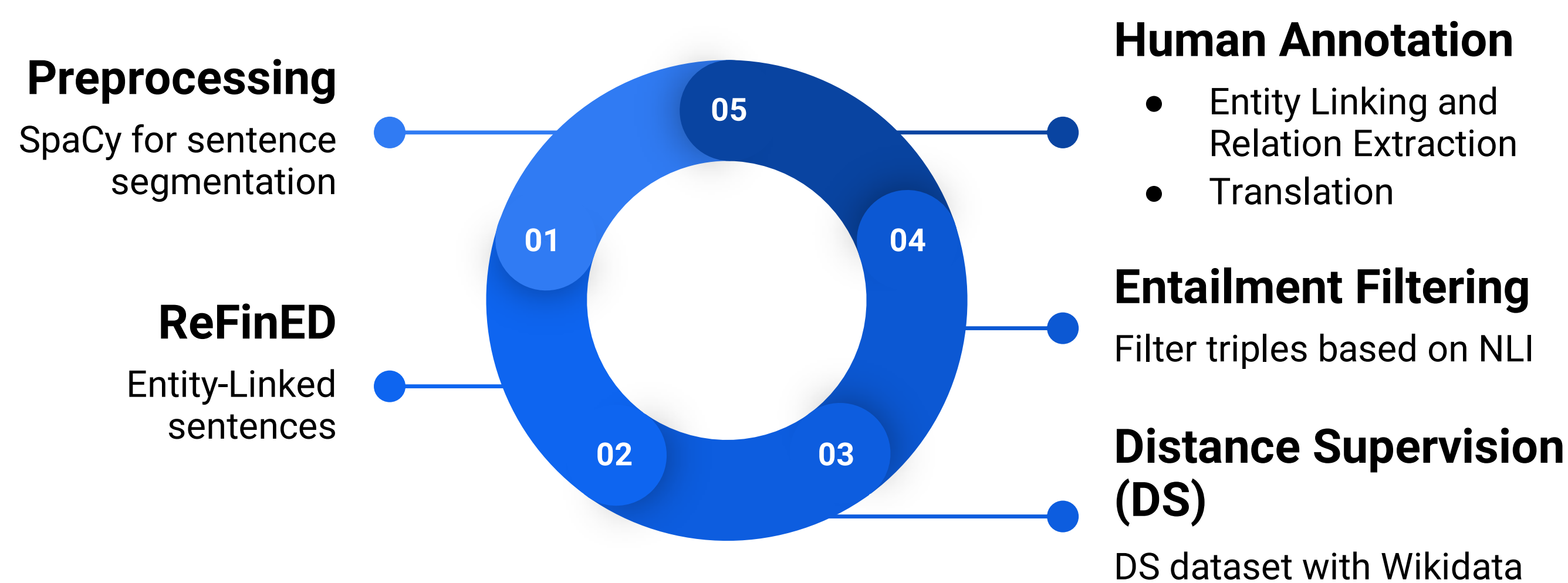


## Introduction

- Information Extraction (IE) is essential for many NLP tasks (KB construction, QA, etc.)
  - Existing IE datasets mostly based on Wikipedia, not applicable for the general web domain where the text is noisy or does not contain fact triples
  - Generative models trained those datasets tend to hallucinate and produce a high rate of false positive
- We present **WebIE**, the first large-scale, entity-linked IE dataset collected from C4, with a subset annotated by humans and translated to four other languages, and includes negative examples to better reflect the data on the web
- We propose several training strategies including jointly training Entity-Linking as an auxiliary task for generative IE, and benchmark with different decoding methods e.g. constraint Trie
- Experimental results show the necessity for training on WebIE for faithful and robust generative IE on the web

## WebIE Collection

- Source: 1M document randomly sampled from the most frequent 200 domains in AllenAI-C4
- Human annotation via crowdsourcing on a subset of WebIE, translation to French, Spanish, Portuguese, and Hindi
- Add negative examples (50% of the instances)



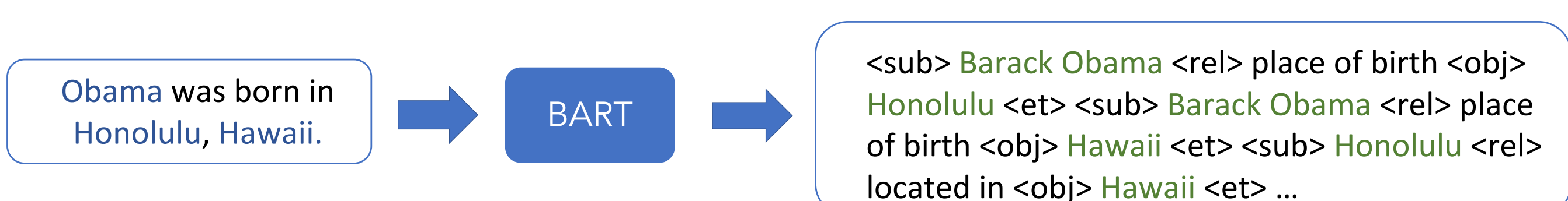
- Statistics of WebIE:

Web Domains	200	Total Triples	1.91M
Relation Types	661	Human Annotated Triples	21.1K
Sentences	1.65M	Negative Instances	50%
train/val/test split	90/5/5	Languages (test set)	5

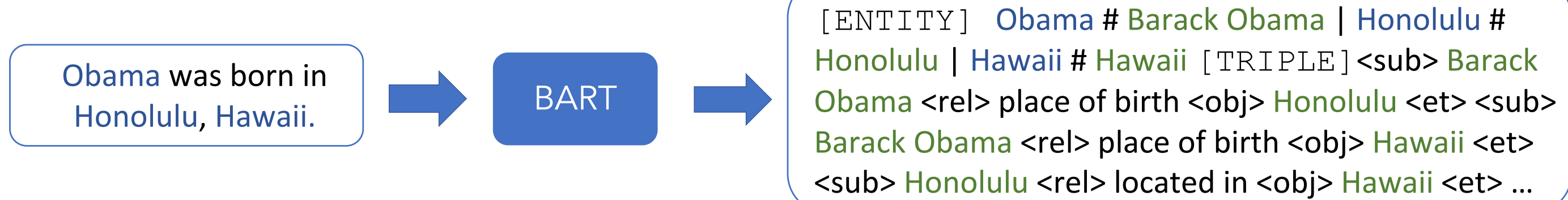
## Training Strategies

- Primarily benchmark generative models on WebIE
- Standard training with BART (and mBART) and joint training with Entity-Linking as an auxiliary task

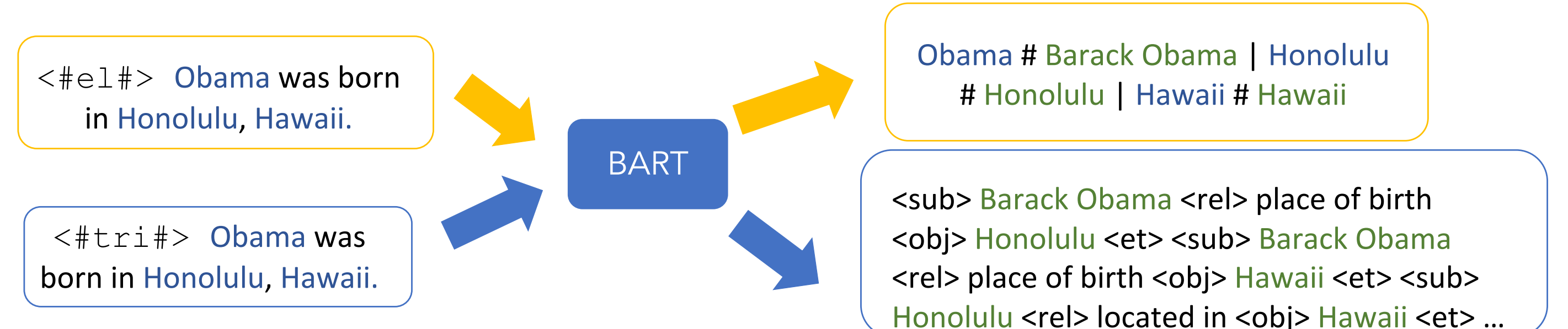
### BART



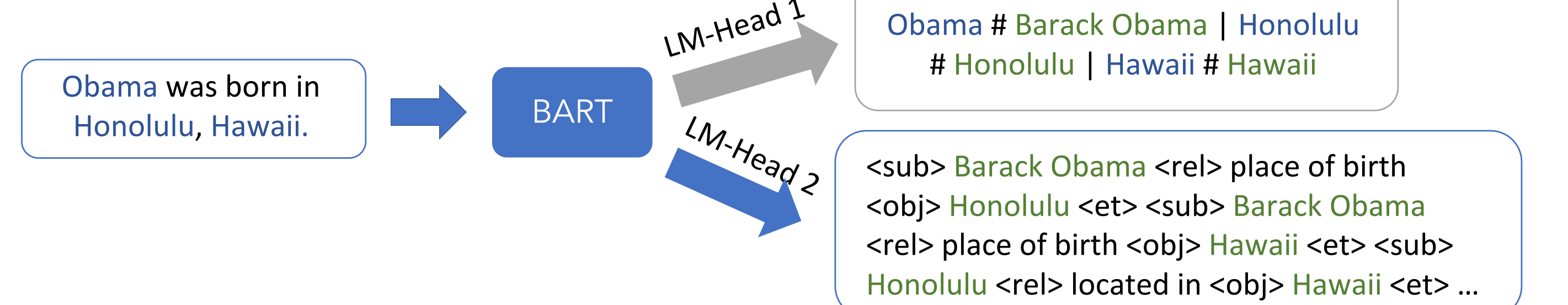
### + ENTITY-PROMPT



### + ARTIFICIAL-PROMPT



### + 2LM-HEADS



## Main Results

- Benchmark beam search and constraint decoding with prefix Trie.
- Multilingual WebIE to evaluate crosslingual transfer performance

Model	WebIE (anno. test)				REBEL			Wiki-NRE		
	Precision	Recall	F1	Acc.-Neg.	Precision	Recall	F1	Precision	Recall	F1
Bart <sub>rand</sub> (r)	11.82	15.63	13.46	0.00	66.89	70.37	68.58	27.61	66.73	39.06
Bart <sub>plm</sub> (r)	15.98	34.92	21.93	0.00	66.28	76.78	71.14	25.39	77.45	38.24
Bart <sub>rand</sub> (w)	52.95	46.60	49.57	95.04	27.47	23.13	25.12	18.98	43.75	26.48
Bart <sub>plm</sub> (w)	57.00	65.91	61.13	94.18	35.81	43.00	39.08	24.30	78.01	37.06
Bart <sub>rand</sub> (r+w)	51.89	54.28	53.06	93.71	66.87	72.24	69.45	29.02	82.35	42.91
Bart <sub>plm</sub> (r+w)	55.22	71.25	62.22	82.59	66.42	78.29	71.87	29.25	86.38	43.70

Table 1. Results on WebIE, REBEL, WikiNRE (latter two are based on Wikipedia). The blue shade indicates zero-shot performance.

Model	REBEL			WebIE (anno)		
	Precision	Recall	F1	Precision	Recall	F1
Bart <sub>rand</sub>	66.89	70.37	68.58	52.95	46.60	49.57
Entity-prompt	67.91	67.54	67.72	51.90	54.28	<b>53.06</b>
Artificial-prompt	66.41	70.72	68.50	53.86	48.18	50.86
2lm-heads	67.05	70.88	<b>68.91</b>	51.07	49.59	50.32

Table 2. Entity linking as an auxiliary task with constraint Trie decoding.

Language	Precision	Recall	F1	Empty-Pos.%	Accuracy-Neg.
English	60.29	64.29	62.22	2.63	96.11
French	46.52	40.26	43.16	12.63	96.64
Spanish	45.13	38.89	41.78	12.80	96.97
Portuguese	44.15	36.61	40.02	14.82	97.22
Hindi	4.23	1.67	2.40	67.55	98.64

Table 3. Performance on mWebIE with mBART. Results for non-English are zero-shot. Empty-Pos(itive)% shows false negatives, revealing zero-shot performance has a high rate of empty results for positive examples.

## Conclusions

- Models trained on WebIE are more generalisable
  - WebIE is necessary for faithful Information Extraction on the web
  - WebIE serves as a complementary dataset for existing Information Extraction datasets based on Wikipedia
  - Entity-linking as auxiliary task improves generative Information Extraction results and offers more explainability
- Zero-shot crosslingual transfer still challenging especially for non-Latin script languages i.e. Hindi