Constrained Decoding for Cross-lingual Label Projection

Duong Minh Le, Yang Chen, Alan Ritter, Wei Xu
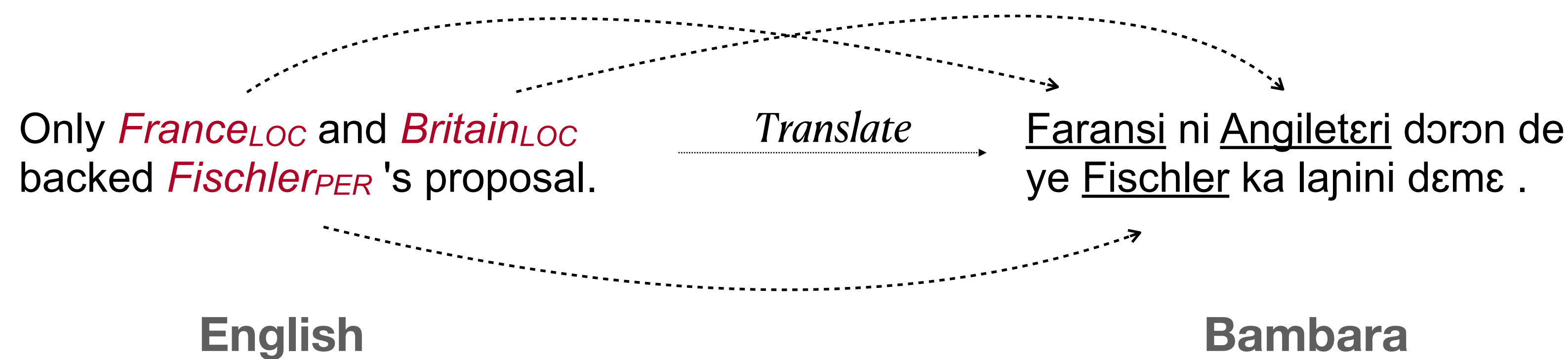
Georgia Tech

Introduction

- **Cross-lingual Label Projection** aims to map annotated label spans (e.g., named entities) of an input sentence onto a corresponding translated text

  Only *France*$_{LOC}$ and *Britain*$_{LOC}$ backed *Fischler*$_{PER}$'s proposal. — *Translate* → Faransi ni Angileteri dɔrɔn de ye Fischler ka laɲini dɛmɛ .
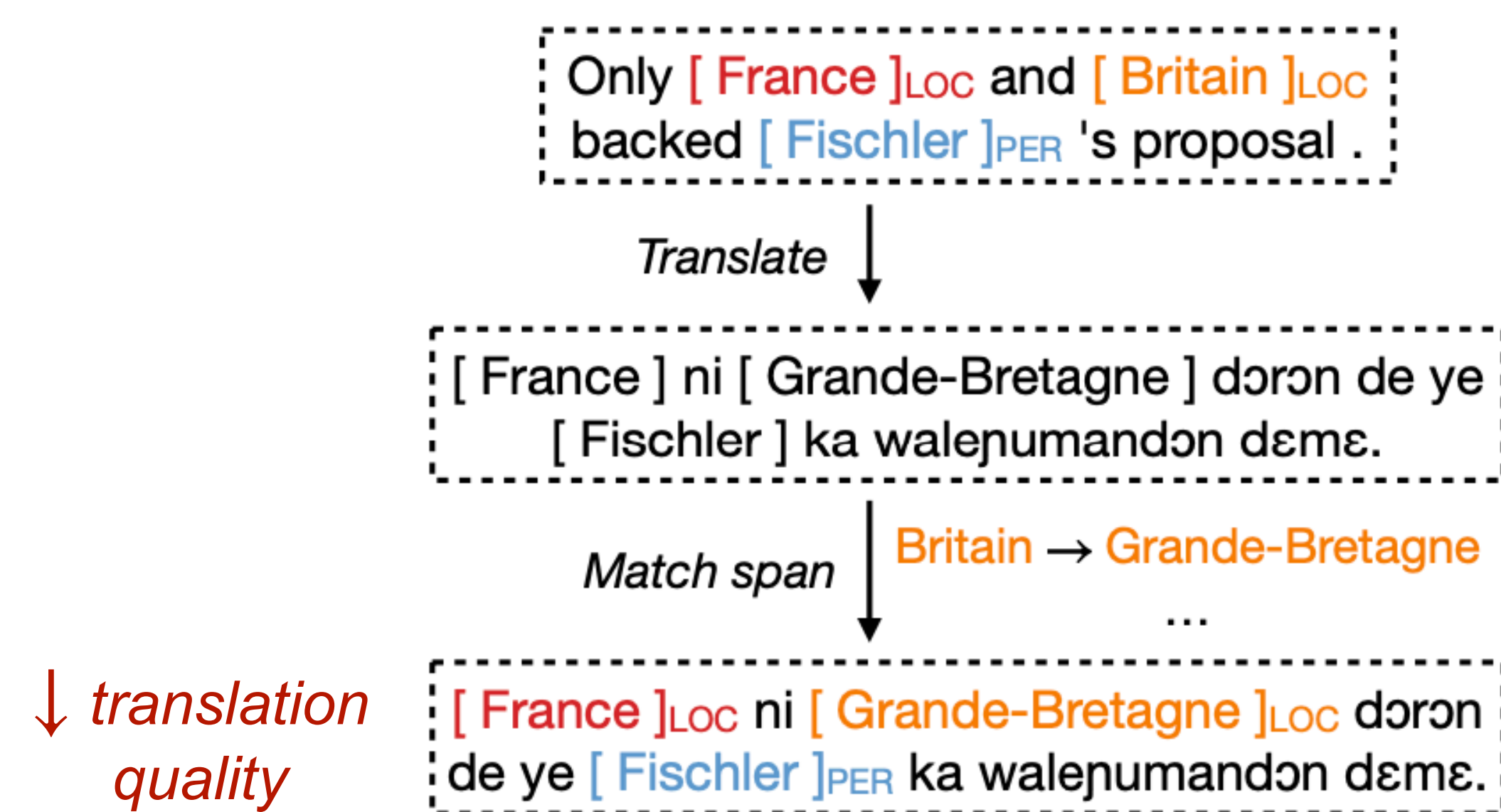
  English   **Bambara**

- For NLP tasks that involve span-level annotations, exploiting translation and *label projection* is an effective method to improve the performance of zero-shot cross-lingual transfer:
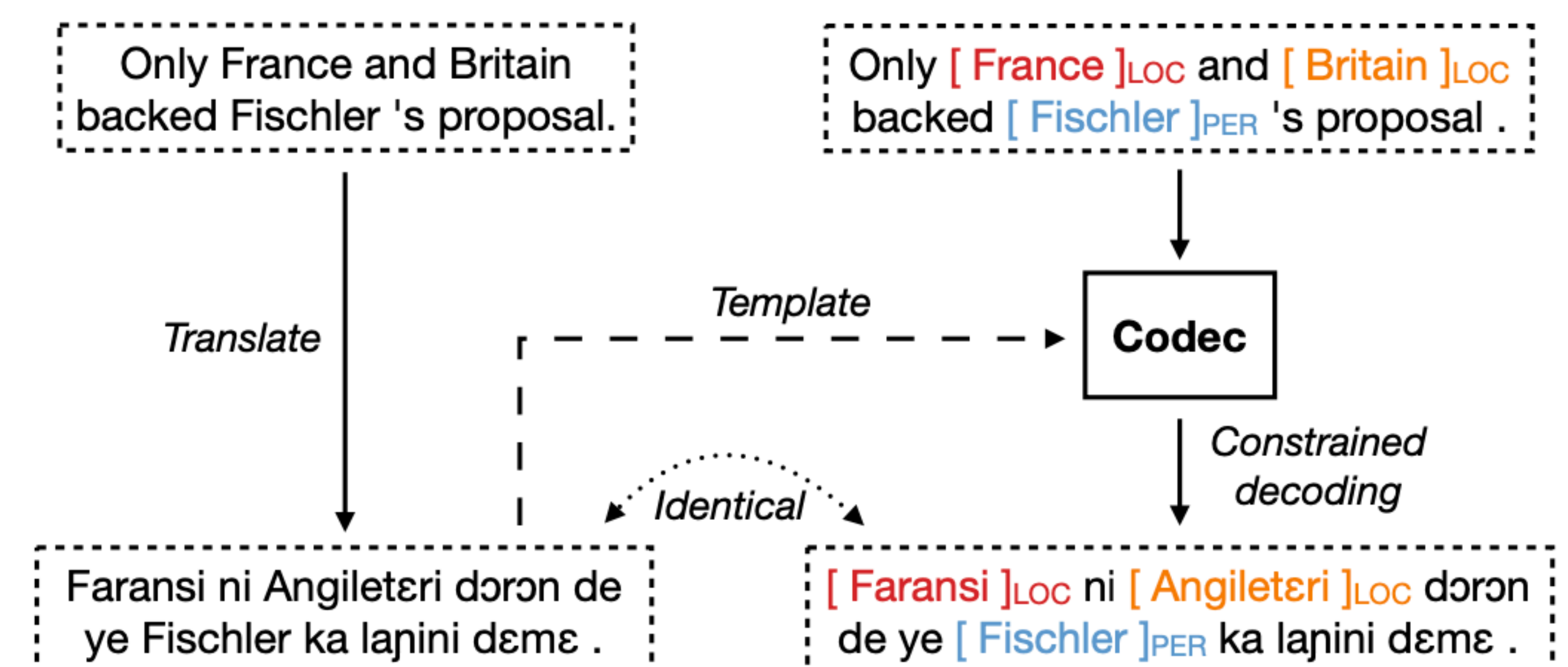
  ▸ Translate-train: translating training data that is available in a high-resource language (e.g., English) together with the gold labels into low-resource languages

  ▸ Translate-test: translating test data in low-resource languages to a high-source language to run inference on, then *projecting* the predicted span-level labels back onto the original test data

Limitation of the previous work

Previous state-of-the-art approach: Translate w/ markers

Only [France]$_{LOC}$ and [Britain]$_{LOC}$ backed [Fischler]$_{PER}$'s proposal .

Translate ↓

[France] ni [Grande-Bretagne] dɔrɔn de ye [Fischler] ka waleɲumandon dɛmɛ.

Match span ↓   Britain → Grande-Bretagne ...

↓ *translation quality*

[France]$_{LOC}$ ni [Grande-Bretagne]$_{LOC}$ dɔrɔn de ye [Fischler]$_{PER}$ ka waleɲumandon dɛmɛ.

Our approach (Codec): Translate w/o markers then insert markers with constrained decoding

Only France and Britain backed Fischler 's proposal.

Only [France]$_{LOC}$ and [Britain]$_{LOC}$ backed [Fischler]$_{PER}$'s proposal .

Translate → *Template* → **Codec** → *Constrained decoding*

Identical

Faransi ni Angileteri dɔrɔn de ye Fischler ka laɲini dɛmɛ .

[Faransi]$_{LOC}$ ni [Angileteri]$_{LOC}$ dɔrɔn de ye [Fischler]$_{PER}$ ka laɲini dɛmɛ .
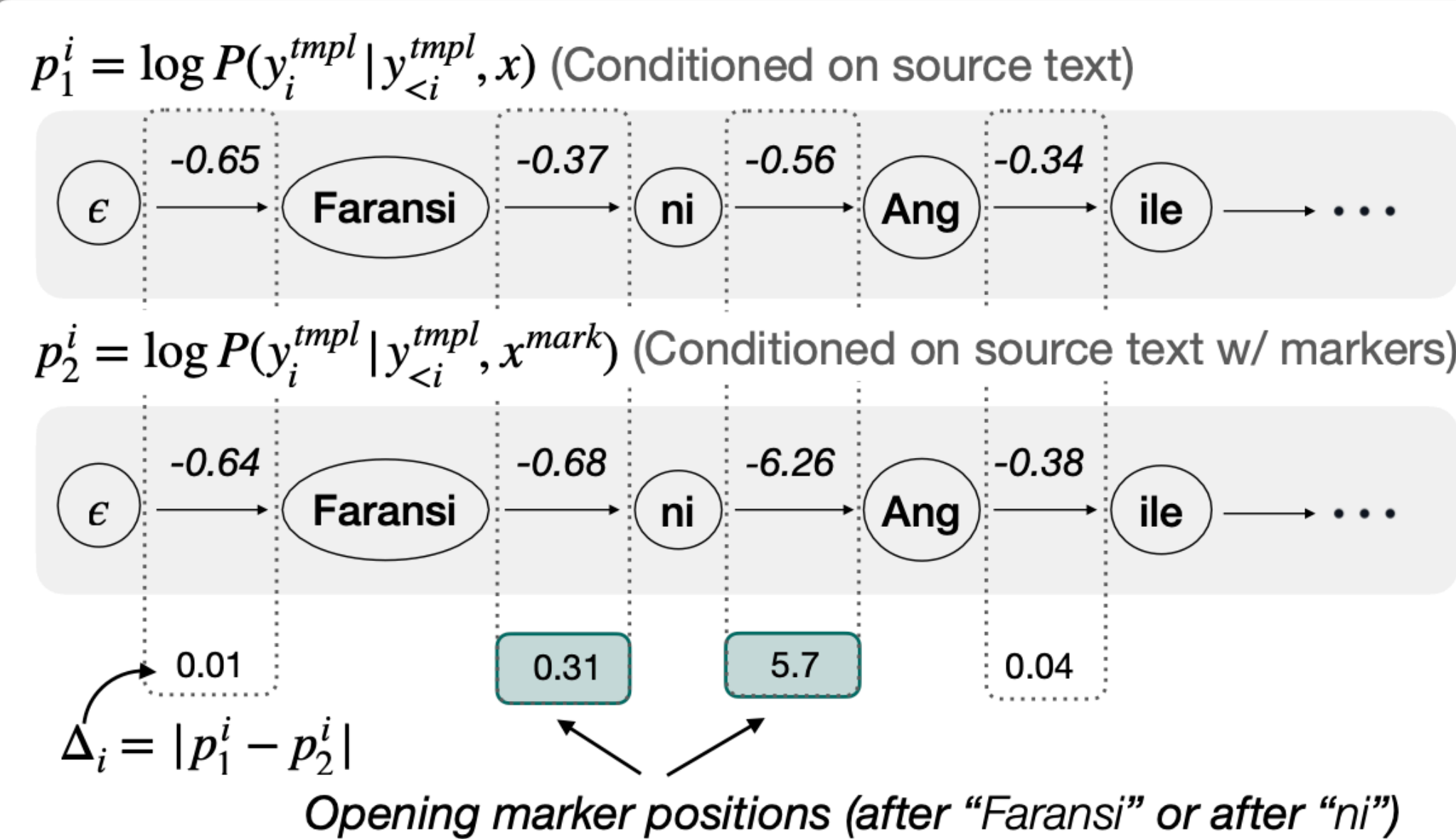
Codec

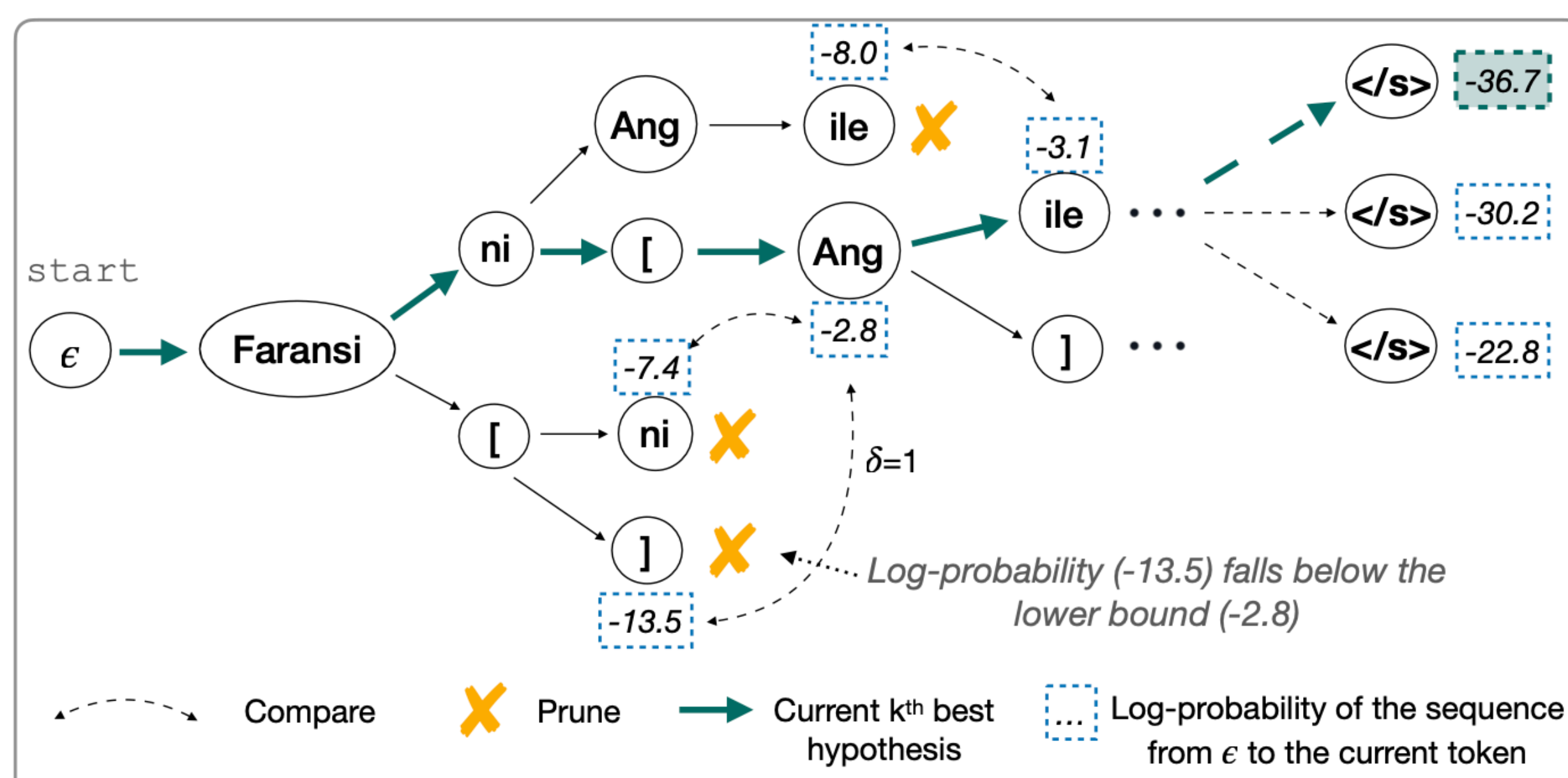**Input:** x = "Only France and Britain backed Fischler 's proposal ."

x^{mark} = "Only France and [Britain] backed Fischler 's proposal ."

y^{tmpl} = "Faransi ni Angileteri dɔrɔn de ye Fischler ka laɲini dɛmɛ ."

Step 1: Prune opening marker positions

$p_1^i = \log P(y_i^{tmpl} | y_{<i}^{tmpl}, x)$ (Conditioned on source text)

ε −0.65 → Faransi −0.37 → ni −0.56 → Ang −0.34 → ile → ...

$p_2^i = \log P(y_i^{tmpl} | y_{<i}^{tmpl}, x^{mark})$ (Conditioned on source text w/ markers)

ε −0.64 → Faransi −0.68 → ni −6.26 → Ang −0.38 → ile → ...

$\Delta_i = |p_1^i - p_2^i|$

0.01   0.31   5.7   0.04

Opening marker positions (after "Faransi" or after "ni")

Step 2: Search for k hypotheses with highest probabilities

start ε → Faransi → ni → [→ Ang → ile → ... → </s>

Ang → ile −8.0 ✗

</s> −36.7

</s> −30.2

</s> −22.8

−3.1 ... −2.8

[→ ni ✗ −7.4

] ✗

δ=1

Log-probability (−13.5) falls below the lower bound (−2.8) −13.5

⟶ *Compare*   ✗ *Prune*   → *Current kth best hypothesis*   ⬚ *Log-probability of the sequence from ε to the current token*

Step 3: Re-ranking

Notations:
- y^i: the ith hypothesis
- e^{src}: the label span in x^{mark}
- e_i^{tgt}: the label span found in y^i

Ranking scores:
- $s_i^{hyp} = \log P(y^i | x^{mark})$
- $s_i^{span} = \log P(e^{src} | e_i^{tgt})$

Experiments

Cross-lingual NER results on the test set of MasakhaNER2.0

(Δ_{FT}: against FT_{En}, †: GPT-4 is evaluated on 200 test examples for each language)

Lang.	GPT-4[†]	FT$_{En}$	Translate-train			Translate-test	
			Awes-align	EasyProject	CODEC (Δ_{FT})	Awes-align	CODEC (Δ_{FT})
Bambara	46.8	37.1	45.0	45.8	45.8 (+8.7)	50.0	**55.6** (+18.5)
Ewe	75.5	75.3	78.3	78.5	**79.1** (+3.8)	72.5	**79.1** (+3.8)
Fon	19.4	49.6	59.3	61.4	**65.5** (+15.9)	62.8	61.4 (+11.8)
Hausa	70.7	71.7	72.7	72.2	72.4 (+0.7)	70.0	**73.7** (+2.0)
Igbo	51.7	59.3	63.5	65.6	70.9 (+11.6)	**77.2**	72.8 (+13.5)
Kinyarwanda	59.1	66.4	63.2	71.0	71.2 (+4.8)	64.9	**78.0** (+11.6)
Luganda	73.7	75.3	77.7	76.7	77.2 (+1.9)	**82.4**	82.3 (+7.0)
Luo	**55.2**	35.8	46.5	50.2	49.6 (+13.8)	52.6	52.9 (+17.1)
Mossi	44.2	45.0	52.2	53.1	**55.6** (+10.6)	48.4	50.4 (+5.4)
Chichewa	75.8	**79.5**	75.1	75.3	76.8 (-2.7)	78.0	76.8 (-2.7)
chiShona	66.8	35.2	69.5	55.9	72.4 (+37.2)	67.0	**78.4** (+43.2)
Kiswahili	82.6	**87.7**	82.4	83.6	83.1 (-4.6)	80.2	81.5 (-6.2)
Setswana	62.0	64.8	73.8	74.0	74.7 (+9.9)	**81.4**	80.3 (+15.5)
Akan/Twi	52.9	50.1	62.7	65.3	64.6 (+14.5)	72.6	**73.5** (+23.4)
Wolof	62.6	44.2	54.5	58.9	63.1 (+18.9)	58.1	**67.2** (+23.0)
isiXhosa	69.5	24.0	61.7	**71.1**	70.4 (+46.4)	52.7	69.2 (+45.2)
Yoruba	**58.2**	36.0	38.1	36.8	41.4 (+5.4)	49.1	58.0 (+22.0)
isiZulu	60.2	43.9	68.9	73.0	**74.8** (+30.9)	64.1	**76.9** (+33.0)
AVG	60.4	54.5	63.6	64.9	67.1 (+12.7)	65.8	**70.4** (+16.0)

MANUAL ASSESSMENT

Manual analysis of Codec outputs for cross-lingual NER from English to Vietnamese/Chinese

Language	MT	Translate-train		
		Correct	Translation	CODEC
Vietnamese	GMT	94	2	4
	Oracle	96	0	4
Chinese	GMT	80	6	14
	Oracle	83	0	16

Language	MT	Translate-test			
		Correct	NER	Translation	CODEC
Vietnamese	GMT	51	33	14	13
	Oracle	85	0	0	15
Chinese	GMT	34	36	24	16
	Oracle	69	0	0	31