LREC 2026 Palma
CLiC Centre de Llenguatge i Computació
UNIVERSITAT DE BARCELONA

BIDIRECTIONAL CHINESE AND ENGLISH PASSIVE SENTENCES DATASET FOR MACHINE TRANSLATION

Xinyue Ma, Pol Pastells, Mariona Taulé Delor, Mireia Farrús

Departament de Filologia Catalana i Lingüística General, Universitat de Barcelona

Paper in PDF:

INTRODUCTION

Passive constructions in Chinese and English differ in form, frequency and distribution.

English passive > Chinese passive
8 times more frequent
English passive → Neutral context
Chinese passive → Negative context

In translation, human translators carefully choose the appropriate voice and construction to convey equivalent information into the target text. However, NMT models and LLMs are still no equals to humans in this task.

To evaluate MT model performance on passive sentence translation, we propose a bidirectional multi-domain parallel dataset of 16,850 Chinese ***bei 被*** passives and 57,115 English ***be*** passives, annotated automatically for translation strategies according to dependency information.

DATASET CONSTRUCTION

Data from five corpora

- CECPC-Core. *China English-Chinese Parallel Corpus-Core*
- Yiyan. *The Yiyan English-Chinese Parallel Corpus*
- BABEL. *The Babel English-Chinese Parallel Corpus*
- E-C Concord. *English-Chinese Parallel Concordancer*
- ClassicCL. *Chinese-English Parallel Corpus of Classic Chinese Literature 中国经典文学作品汉英平行语料库*

Registers

- A_PRESS
- B_OFFICIAL DOCUMENT
- C_ACADEMIC PROSE
- D_GENERAL PROSE
- E_LITERATURE

Subsets

- ZH → EN(be)
- ZH(bei) → EN
- EN(be) → ZH
- EN → ZH (bei)

Translation Strategies

Chinese Structures

- Passive
  - Syntactic passive
  - Lexical passive
- Active
  - Notional passive
  - Topic sentence
  - Light verb
  - Causative
  - Resultative
  - N/A

English Structures

- Passive
  - BE passive
  - GET passive
  - HAVE passive
  - BECOME passive
- Active
  - N/A

EXPERIMENTS AND RESULTS

Voice	Structure	ZH→EN(be) source text	EN(be)→ZH target text			EN→ZH(bei) target text	
			Human	OPUS	NLLB	OPUS	NLLB
Passive	Syntactic passive	8.9	13.8	26.7	26.5	43.2	40.3
	Lexical passive	2.3	3.3	3.7	1.9	3.0	2.3
Active	Notional passive	0.9	0.6	0.4	0.4	0.4	0.3
	Topic sentence	4.7	6.0	2.9	0.3	1.2	0.2
	Light verb	5.2	3.3	4.2	2.8	3.3	2.5
	Causative	1.4	3.4	2.2	1.3	1.9	1.0
	Resultative	5.3	6.6	3.0	2.2	3.2	2.3
	N/A	71.5	63.1	57.0	64.7	43.8	51.2

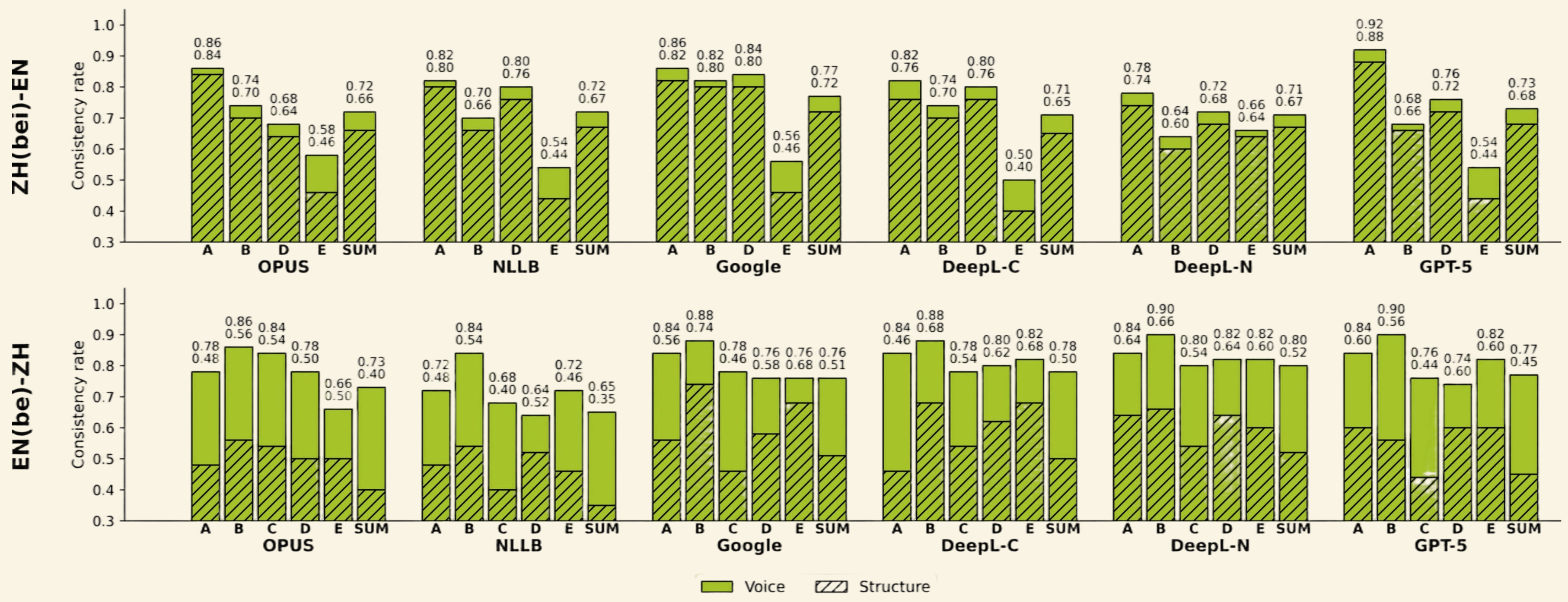
Proportion (in %) of structures in Chinese source and target text

Voice	Structure	EN→ZH(bei) source text	ZH(bei)→EN target text			ZH→EN(be) target text	
			Human	OPUS	NLLB	OPUS	NLLB
Passive	BE	65.0	42.4	74.2	73.8	44.1	44.8
	GET	0.9	0.7	1.0	0.2	0.3	0.2
	HAVE	0.2	0.4	0.1	0.1	0.0	0.1
	BECOME	0.3	0.0	0.0	0.0	0.0	0.0
Active	N/A	33.6	56.5	24.8	25.8	55.6	54.9

Proportion (in %) of structures in English source and target text

	Evaluation	NMT				LLM	
		OPUS	NLLB	Google	DeepL-C	DeepL-N	GPT-5
ZH(bei)→EN	BLEU	15	16.9	**22.9**	19.6	19	15.5
	chrF++	43.3	44.2	**51.4**	49.4	48.2	44
	COMET	74.4	75.4	**80.2**	78.3	79.2	79.1
	Label div.	3 / 5	2 / 5	4 / 5	**5 / 5**	4 / 5	**5 / 5**
EN(be)→ZH	BLEU	21	17.8	**29.7**	28.7	23.7	23
	chrF++	21.4	18.2	**28.7**	28.1	23.7	23.2
	COMET	78.1	74.1	**85.2**	84.2	84.6	84
	Strategy div.	**8 / 8**	7 / 8	**8 / 8**	7 / 8	**8 / 8**	**8 / 8**
	Label div.	17 / 18	10 / 18	16 / 18	13 / 18	**19 / 18**	16 / 18

Evaluation results using metrics and according to annotation.
div. stands for diversity in Label div. and Strategy div.



Consistency rate of model translations compared to human translations in the voice and the structure used, of each register and the whole test set.

Case Study

Now let the gas *be* rapidly *compressed* back to its initial volume, Vi.

a. Human:
接下来， 又 把 气体 快速
next also BA gas rapidly
压缩 回 初始的 体积 *Vi*。
compress back initial volume Vi

b. DeepL-N:
现在， 让 气体 被 迅速 压缩
now let gas BEI rapidly compress
回 初始 体积 *Vi*。
back initial volume Vi

难道不是适得其反，想要称霸的帝国主义却得到了被打倒的结果吗？

a. Human:
Didn’t the results turn out to be just the opposite of what they wanted? Didn’t the imperialists who aimed at domination *get struck down* themselves?

b. GPT-5:
Isn’t it ironic that the imperialists who sought hegemony ended up *being overthrown*?

CONCLUSION

- ◆ Compared to humans, Models are more influenced directly by the syntax of the source text and are ignorant of the general voice usage of the language.
- ◆ Commercial NMT models scored higher in metric evaluations, but LLMs showed better ability in using diverse alternative structures in translations.