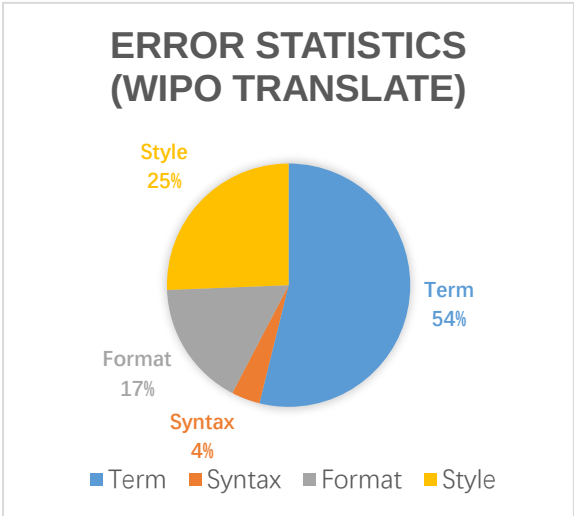
General and Specialized Machine Translation Engines:
A comparative study of Google Translate and WIPO Translate for Patent Documents

Project Purpose: To evaluate the performance of Google Translate and WIPO translate on patent document tasks, and to explore the further improvement of specialized machine translation.

Methodology: Selected patent texts in automobile, AI, medicine and chemistry four fields as translation samples. Tagged error types and collected error quantities (Table 4.1 & Fig.1). Used Bleu score to objectively evaluate the performances (Table 5.3-5.6). Analyzed the translations in detail on the level of terms, syntax, style, and format.

Table 4.1 Text 1 Error Types Tagging

Original Text	Standard Translation	WIPO Translate	Google Translate
Abstract: A COVID-19 risk prompt apparatus, method and system based on ultrasonic imaging. The method comprises: obtaining a recognition result by means of a pre-trained recognition model, wherein the recognition result comprises a B line and/or a lung consolidation area recognized from each lung area of a lung;	摘要：一种基于超声成像的新冠肺炎风险提示装置、方法和系统，通过预先训练的识别模型得到识别结果，识别结果包括从肺部的各肺区中识别出的B线和/或肺实变区域。	摘要：一种基于超声成像的 covid-19（漏译）风险提示装置，（标点）方法及系统。所述方法包括：通过预先训练的识别模型获取识别结果，所述（冗余）识别结果包括从肺（术语）的每个肺部区域（术语）识别出的 b 线（术语）和/或肺固结区域（术语）；	摘要：一种基于超声成像的 COVID-19（漏译）风险提示装置、方法和系统。该方法包括：通过预训练的识别模型得到识别结果，识别结果包括从肺（术语）的每个肺区域（术语）识别出的B线和/或肺实变区域。
and according to epidemiological information and symptom information of an inspected person and the recognition result output by the recognition model, determining a risk level of the inspected person suffering from the COVID-19.	根据被检测者的流行病学信息和症状信息，以及由识别模型输出的识别结果，确定被检测者患有新冠肺炎的风险等级。	根据被检者（术语）的流行病学信息和症状信息以及识别模型输出的识别结果（可读性），确定 covid-19（漏译）的被检者的风险等级（分词），（标点）	根据被检查人（术语）的流行病学信息和症状信息以及识别模型输出的识别结果（可读性），确定被检查人（术语）感染 COVID-19（术语）的风险等级。



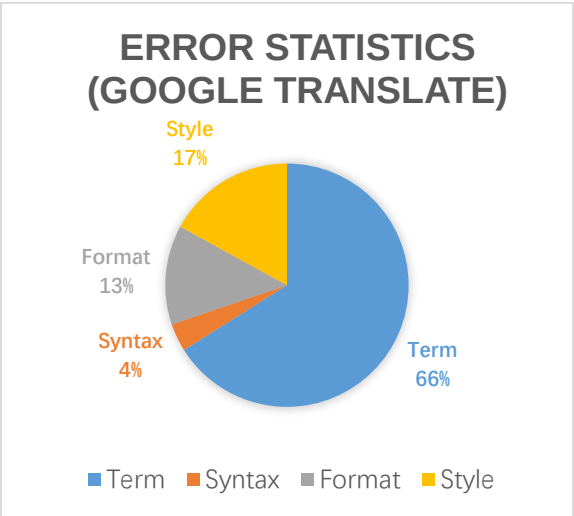


Fig.1 Error Statistics of Two Translate Engines

General and Specialized Machine Translation Engines: A comparative study of Google Translate and WIPO Translate for Patent Documents

Results: Google translate is superior to WIPO translate on the whole but makes more errors on the term level. Both engines perform better in the field of AI but worse in the fields of medicine and chemistry (Fig.2).

Table 5.3 Text 1 (Medicine) WIPO Translate & Google Translate (BLEU Score)

		1-gram	2-gram	3-gram	4-gram	Precision × brevity
WIPO	Individual	64.71	40.61	28.75	20.00	35.06×100.00
	Cumulative	64.71	51.26	42.27	35.06	
Google	Individual	63.48	47.40	38.10	30.06	43.09×100.00
	Cumulative	63.48	54.85	48.58	43.09	

Table 5.4 Text 2 (AI) WIPO Translate & Google Translate (BLEU Score)

		1-gram	2-gram	3-gram	4-gram	Precision × brevity
WIPO	Individual	83.13	64.78	49.34	37.24	56.09×98.21
	Cumulative	81.64	72.07	63.14	55.08	
Google	Individual	79.43	59.52	45.34	33.12	51.62×100.00
	Cumulative	79.43	68.76	59.85	51.62	

Table 5.5 Text 3 (Vehicle) WIPO Translate & Google Translate (BLEU Score)

		1-gram	2-gram	3-gram	4-gram	Precision × brevity
WIPO	Individual	74.74	53.01	36.36	27.81	44.74×100.00
	Cumulative	74.74	62.94	52.42	44.74	
Google	Individual	81.18	60.34	45.35	33.94	52.40×100.00
	Cumulative	81.18	69.99	60.56	52.40	

Table 5.6 Text 4 (Chemistry) WIPO Translate & Google Translate (BLEU Score)

		1-gram	2-gram	3-gram	4-gram	Precision × brevity
WIPO	Individual	78.10	47.00	30.53	21.11	39.22×81.87
	Cumulative	63.94	49.60	39.47	32.11	
Google	Individual	75.83	58.26	43.64	35.24	51.05×95.12
	Cumulative	72.13	63.23	54.95	48.56	

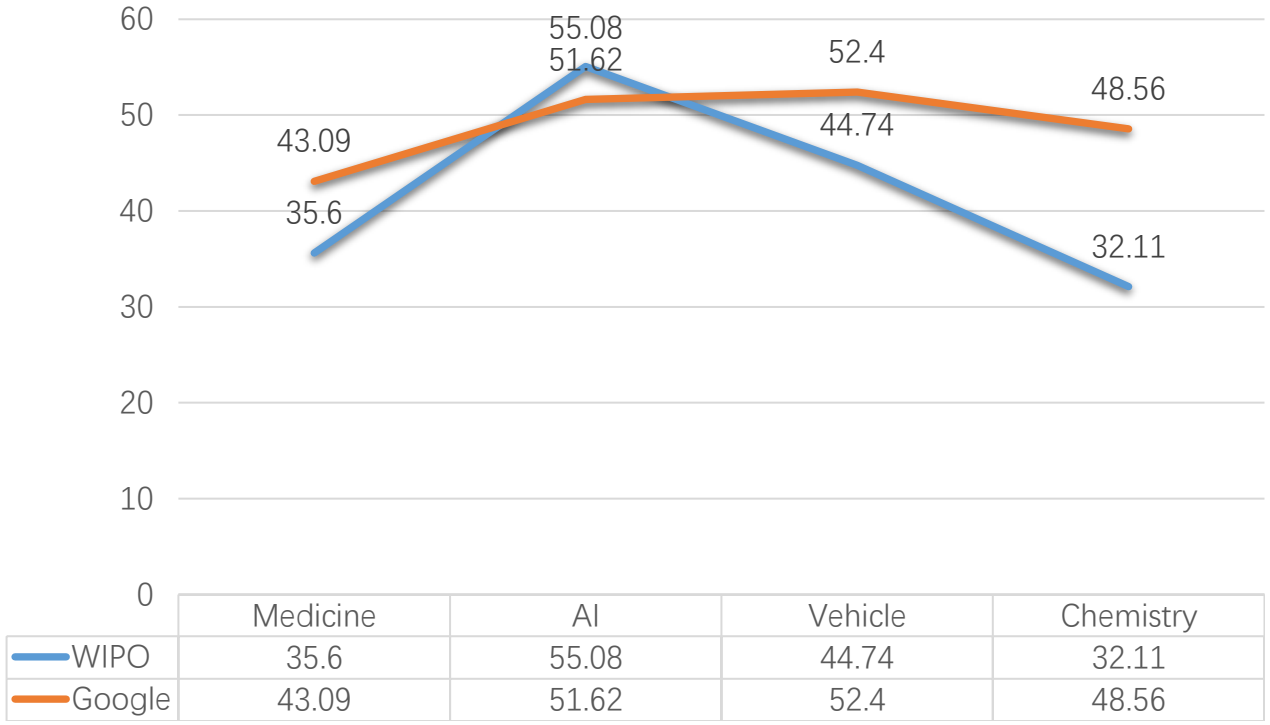


Fig.2 Comparison of BLEU Scores of WIPO & Google