

EW/HA,
THE FUTURE
WE CREATE

범용엔진과 커스텀엔진의 한영 법률번역 품질 평가 사례연구

이지은 · 최효은

(이화여자대학교 통역번역대학원)



연구 배경과 연구 문제

연구 필요성

- 선행연구 범용 인공신경망번역엔진(NMT) 성능 평가 위주 & 국내외적으로 법률 MT 연구 희소
- 법률 MT에 대한 기대와 수요(Roberts 2022; 김한균 2021)

연구 목적

- NMT 3종의 한국어-영어 법률번역 품질 평가 :
범용 NMT(구글 번역과 네이버 파파고)과 법률특화 NMT(오토란)의 자동번역 성능 평가

MT 평가 방법 - 자동평가

- MT 결과가 레퍼런스인 인간번역(HT)과 얼마나 닮아 있는지를 평가하며, BLEU, METEOR, NIST, TER and chrF 등이 있음 (Chatzikoumi 2020; Rossi & Wiggins 2013)
- 빠르고 객관적이나 인간번역과의 유사성 정도를 측정하기 때문에 번역의 품질을 직접적으로 평가하기는 어려움

MT 평가 방법 - 수동평가

- 사람이 MT 품질을 직접 평가
- 대개 정확성(충실성)과 유창성을 기준으로 함
- 리커트 척도를 활용하는 방법을 비롯해서 점수 또는 순위 매기기, PE 기반 평가, HT 수정률 등의 방법을 활용

⇒ 자동평가와 수동평가의 병행이 가장 바람직하나 현실적인 문제로 이 두 평가가 병행되어 이루어진 연구는 그다지 많지 않음

MT 평가 방법 - 오류 분류

- Dynamic Quality Framework (DQF) – 적절성과 유창성을 기준으로 4점 척도(none, minor, major and critical)로 평가
- Multidimensional Quality Metrics (MQM) – 평가 기준이 정확성, 유창성, 용어, 스타일, 로케일 관습, 디자인이며 각 기준 별로 하위 분류를 가짐 (Chatzikoumi 2020)

⇒ 본 연구에서는 이와 같은 오류 분류를 단순화하여 적용

MT 평가 연구

- MT의 품질은 언어쌍, 문장의 길이와 구조, 어휘의 형태, 텍스트 종류에 따라 달라지는 경향(Enríquez Raído 2016; Guo 2016; Ragni & Vieira 2022; 박옥수 2017; 한현희 2020)
- 특히 한-영 기계번역은 고유명사나 다의어가 포함되어 있을 때 정확성이 현저히 저하되는 경향(서보영과 김순영 2018) & 길고 복잡한 문장 구조, 한국어의 주어 생략 등도 번역 오류를 일으키는 원인(박옥수 2017)

⇒ 원문이 길고 복잡한 구조를 가지며 복합명사와 한자 기반의 어휘가 다수 포함된 국문법령의 경우 MT 오류 예상

연구 방법

- 대상 - 범용 NMT 2종(구글 NMT1, 파파고NMT2)과 법률특화 NMT 1종(오토란 NMT3)
- 법령 - 물관리기술 발전 및 물산업 진흥에 관한 법률 14개 조와 외국인투자촉진법 13개 조의 총 180개 세그먼트를 평가
- 자동평가
 - BLEU, METEOR 활용
- 수동평가
 - 4명의 법률번역 전문가(석사 이상, 최소 5년 이상 경력)
 - 4점 척도(0~4)
 - 평가기준: 정확성(추가, 누락, 오역, 미번역), 유창성(통사, 문법), 용어(용어 정확성, 용어 일관성), 기타(스타일)

연구 결과 - 자동평가 결과

MT엔진/ 평가 종류	BLEU	METEOR
NMT 1	0.421	0.376
NMT 2	0.395	0.352
NMT3	0.585	0.445

연구 결과 - 자동평가 결과

MT엔진/ 평가 종류	BLEU	METEOR
NMT 1	0.421	0.376
NMT 2	0.395	0.352
NMT3	0.585	0.445

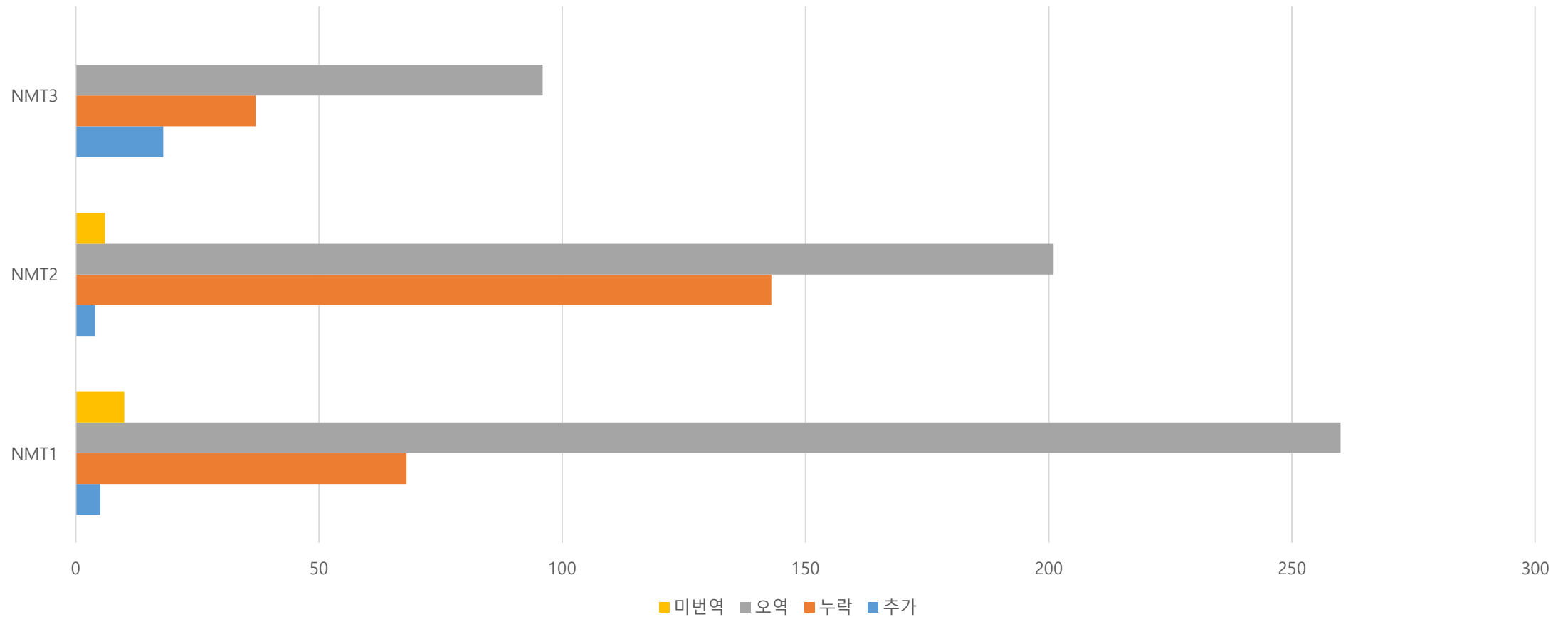
연구 결과 - 수동평가 결과

MT엔진/ 평가자	E1	E2	E3	E4	평균 점수
NMT 1	2.5	3.3	2.8	2.6	2.8
NMT 2	2.5	3.1	2.9	2.7	2.8
NMT3	2.8	3.6	3.4	3.2	3.5

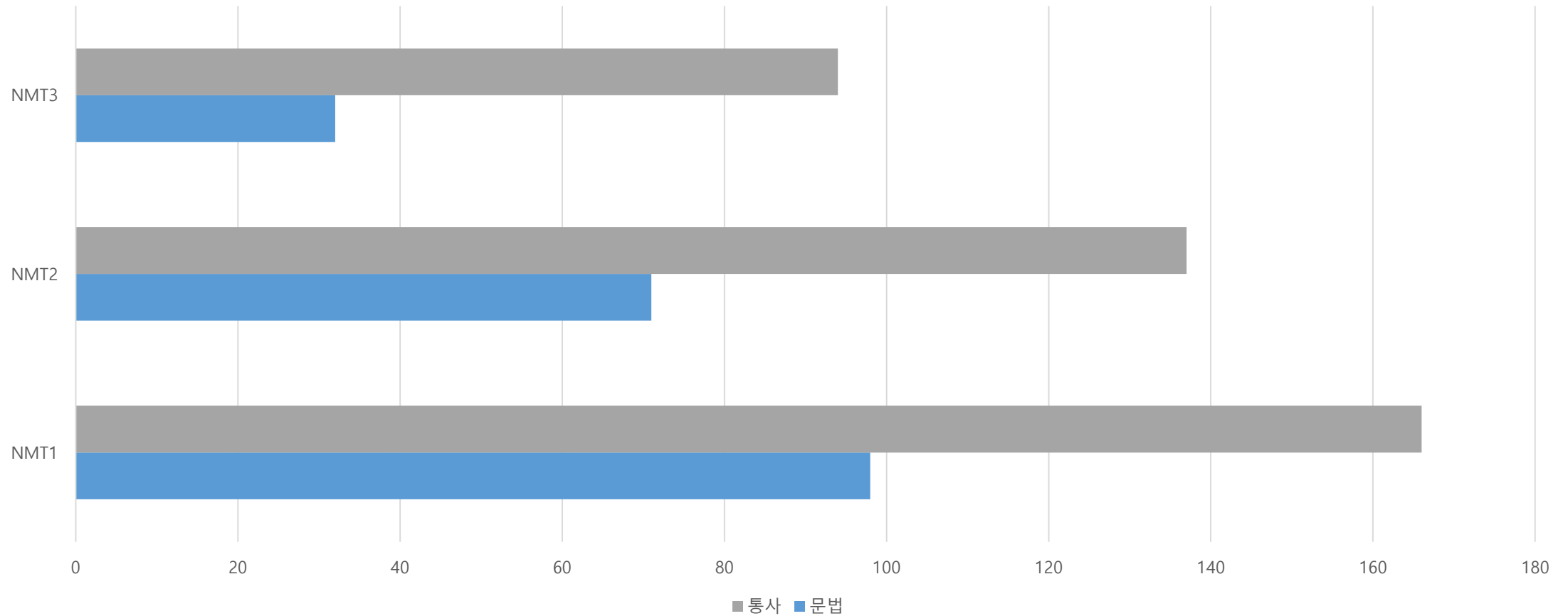
연구 결과 - 오류 개수

분류 (오류 개수, %)	NMT1		NMT2		NMT3	
정확성	343	35.8%	354	28.7%	151	22.4%
유창성	264	27.6%	208	22.8%	126	18.5%
용어	123	12.8%	151	16.5%	229	33.5%
기타	228	23.8%	201	22%	174	25.6%
총합	958	100%	914	100%	680	100%

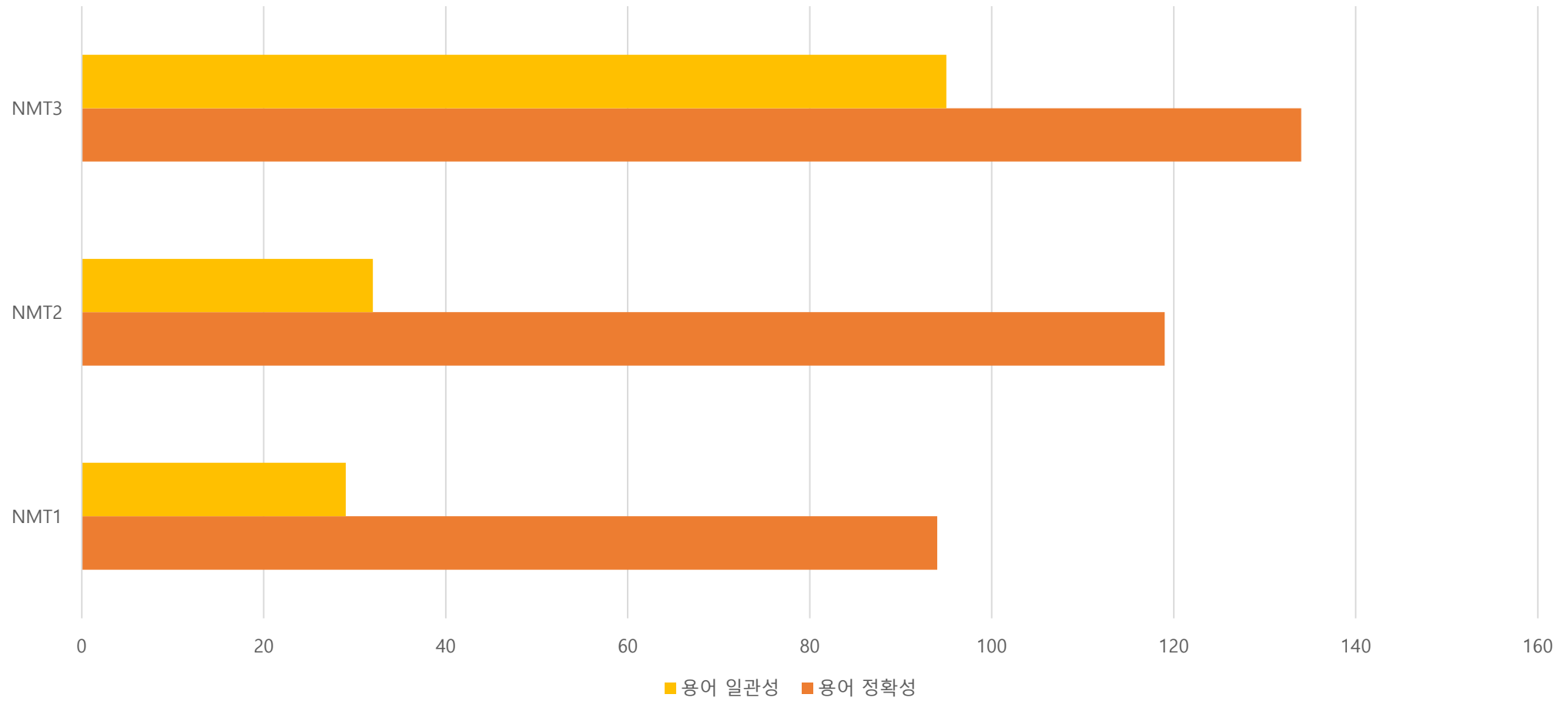
연구 결과 - 정확성 오류



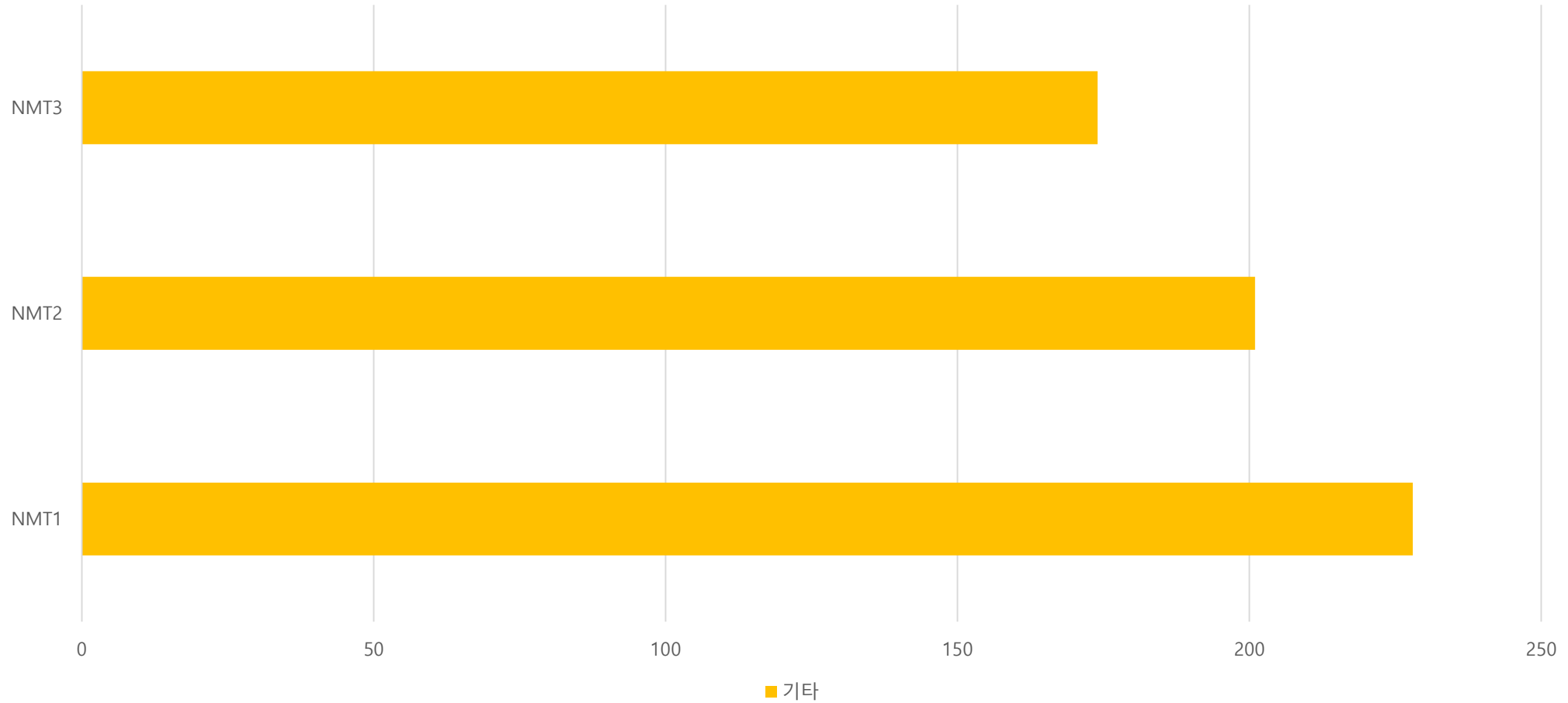
연구 결과 - 유창성 오류



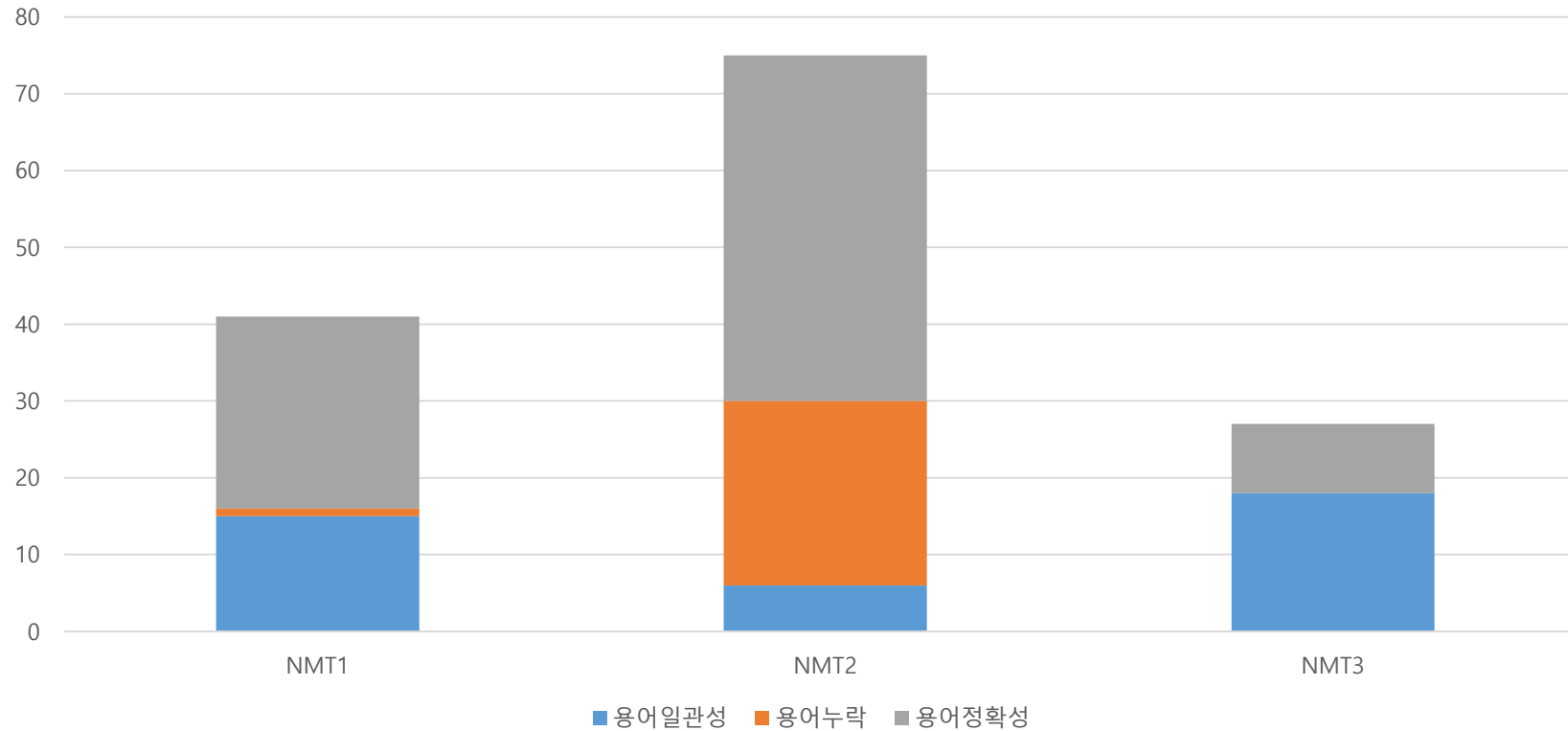
연구 결과 - 용어 오류



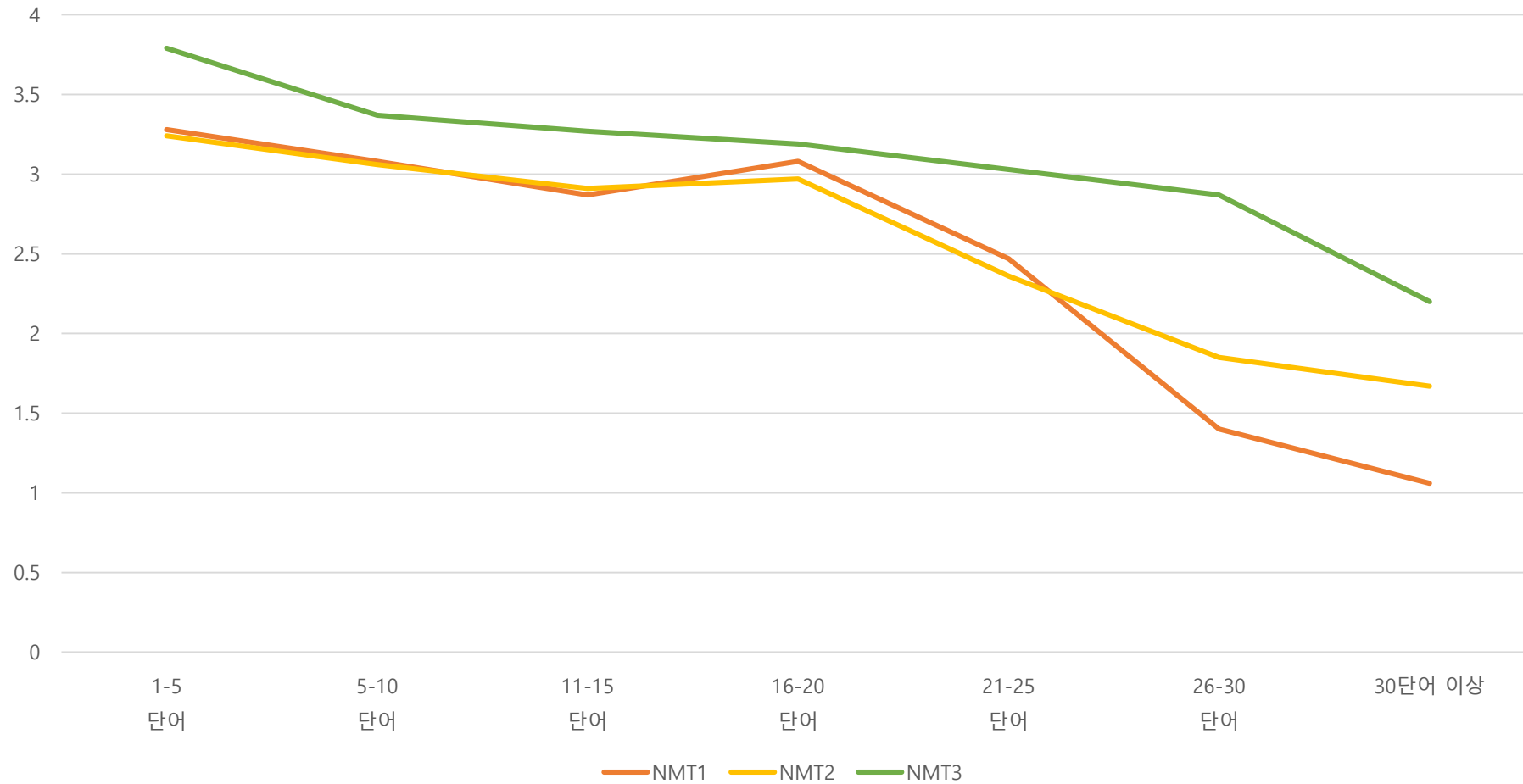
연구 결과 - 기타 오류



연구 결과 - 복합명사의 번역 오류 분석



연구 결과 - 세그먼트의 길이별 품질



결론

- BLEU와 METEOR 점수 기반 자동평가에서 커스텀엔진은 범용엔진에 비해 참조 번역과 더 유사하여 성능이 우수한 것으로 나타남
- 평가자 4인에 의한 수동평가에서 커스텀엔진의 평균 점수가 범용엔진 평가 점수 상회함
- 자동평가 결과와 수동평가 간 상관관계 어느 정도 확인됨
- 오류 분석에 의하면 커스텀엔진은 정확성과 유창성에서 두 범용엔진의 성능을 훨씬 능가한 반면 용어에서는 그렇지 못함
- 단, 25단어 이상의 장문과 복합명사를 중심으로 한 법률용어 번역에서는 커스텀 엔진의 우월한 성능 확인
- 3종의 NMT 모두 포스트에디팅 없이 사용하는 것은 불가하며 신뢰할 만한 수준의 법령 번역 생산을 위해 인간 개입은 필수적임

참고문헌

- Chatzikoumi, Eirini. (2020). How to evaluate machine translation: A review of automated and human metrics. *Natural Language Engineering*, 26, 137–161.
- Enríquez Raído, Vanessa. (2016). Translators as adaptive experts in a flat world: From Globalization 1.0 to Globalization 4.0? *International Journal of Communication*, 10, 970–988.
- Guo, Jia Wen. (2016). Is Google Translate adequate for facilitating instrument translation from English to Mandarin? *Cin-Computers Informatics Nursing*, 34(9), 377–383.
- Ragni, Valentina & Lucas Nunes Vieira. (2022). What has changed with neural machine translation? A critical review of human factors. *Perspectives*, 30(1), 137–158.
- Roberts, Ben. (2022). Machine vs. Human translation: When to use which for legal translation. Retrieved October 18, 2022, from <https://www.attorneyatwork.com/machine-translation-vs-human-translation-when-to-use-which-for-legal-translation/>
- Rossi, Laura & Dion Wiggins. (2013). Applicability and application of machine translation quality metrics in the patent field. *World Patent Information*, 35, 115–125.
- 김한균 (2021). 「형사절차상 인공지능기반 통번역 애플리케이션 활용과 적법절차. 『형사소송 이론과 실무』 13(3): 147-174.
- 박옥수. (2017). 「한영 기계 번역에서 ST의 유형적 특징에 따른 번역 오류 분석」. 『동아인문학』, 41, 155-183.
- 서보현, 김순영. (2018). 「기계번역 결과물의 오류유형 고찰」. 『번역학연구』, 19(1), 99-117.
- 한현희 (2020). 「한-노 기계 번역, 어디까지 왔나?: Google과 Papago 번역 성능 비교를 기반으로」. 『노어노문학』, 32(3), 63-93.

EWHA,
THE FUTURE
WE CREATE

감사합니다.

