

Translation of Korean Object Case Markers to Mongolian's Suffixes

Khulan Setgelkhuu[†] · Joon Choul Shin^{**} · Cheol Young Ock^{***}

ABSTRACT

Machine translation (MT) system, especially Korean-Mongolian MT system, has recently attracted much attention due to its necessary for the globalization generation. Korean and Mongolian have the same sentence structure SOV and the arbitrarily changing of their words order does not change the meaning of sentences due to postpositional particles. The particles that are attached behind words to indicate their grammatical relationship to the clause or make them more specific in meaning. Hence, the particles play an important role in the translation between Korean and Mongolian. However, one Korean particle can be translated into several Mongolian particles. This is a major issue of the Korean-Mongolian MT systems. In this paper, to address this issue, we propose a method to use the combination of UTagger and a Korean-Mongolian particles table. UTagger is a system that can analyze morphologies, tag POS, and disambiguate homographs for Korean texts. The Korean-Mongolian particles table was manually constructed for matching Korean particles with those of Mongolian. The experiment on the test set extracted from the National Institute of Korean Language's Korean-Mongolian Learner's Dictionary shows that our method achieved the accuracy of 88.38% and it improved the result of using only UTagger by 41.48%.

Keywords : Korean, Mongolian, Machine Translation System, Object Case Marker, UTagger

한국어 목적격조사의 몽골어 격 어미 번역

Khulan Setgelkhuu[†] · 신 준 철^{**} · 옥 철 영^{***}

요 약

최근 기계 번역에 관한 연구들이 활발하게 이루어지고 있고 한국어와 몽골어 간의 상호 기계 번역 시스템도 개발되고 있다. 한국어와 몽골어는 계통적으로 같은 어족에 속하며 '주어+목적어+서술어'라는 비교적 자유로운 어순을 가지는 언어이고 어미와 조사가 발달한 것이 그 특징이다. 따라서 기계 번역 시 양언어의 조사나 어미의 의미를 잘 번역하는 것이 중요하다. 그러나 한국어 목적격 조사를 몽골어로 번역할 때 한국어 목적격 조사가 몽골어의 여러 격 어미로 번역이 될 수 있는데, 기존의 연구들은 한 가지 격 어미로만 번역해 정확한 의미를 전달하지 못하는 문제점이 있다. 본 논문에서는 이러한 문제점을 개선하기 위하여 한국어 형태소 분석과 동시에 품사 및 동형어의어 태깅 시스템인 유테거(UTagger)를 기반으로 하여 한국어 목적격 조사의 몽골어 격 어미 결정 방법을 제안한다. 제안하는 방법에서는 한국어 목적격 조사에 대응하는 몽골어 격 어미들을 살펴보고 데이터 테이블을 설계하여 적절한 격 어미를 결정한다. 제안한 방법의 성능을 검증하기 위하여 한국어기초사전에서 데이터를 추출하고 유테거와 비교 실험하였다. 실험 결과 목적격 조사를 바로 대격 어미로 번역한 유테거의 정확률은 46.9%인데 반해 본 논문에서 제안한 방법은 88.38%로 제안한 방법이 41.48%p 더 우수한 결과를 보였다.

키워드 : 한국어, 몽골어, 기계 번역 시스템, 목적격 조사, 대격 어미, 유테거

1. 서 론

최근 기계 번역에 관한 연구들이 활발하게 이루어지고 있다[1, 2]. 특히 한국어-영어 기계 번역[3, 4], 한국어-일본어의

기계 번역 등이[5] 활발히 연구되고 있다. 그러나 한국어 몽골어 간의 기계 번역 연구는 그다지 많지 않다[6-8]. 한국어 격 조사들이 몽골어로 번역이 될 때 여러 가지 격 어미로 번역이 되는데 기존 한국어-몽골어 기계 번역 연구들[6, 9, 10]은 모두 한 가지 격 어미로 번역되는 문제점이 있다. 언어학 논문들[11-13]을 보면 한국어 목적격 조사가 여러 가지 격 어

* 이 논문은 2016년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2016S1A5B6913773).

[†] 준 회 원 : 울산대학교 IT융합학부 석사

^{**} 비 회 원 : 울산대학교 IT융합학부 연구교수

^{***} 종신회원 : 울산대학교 IT융합학부 교수

Manuscript Received : October 5, 2018

First Revision : December 3, 2018

Accepted : December 18, 2018

* Corresponding Author : Cheol Young Ock(okcy@ulsan.ac.kr)

1) 몽골어 문법에서는 '대격 어미'라고 지칭한다. 몽골어의 몽골어학에서는 전통적으로 '조사'라는 개념을 쓰지 않기 때문에 한국의 몽골어학에서는 이를 사용하여 '격 어미'라는 용어를 쓴다. 한국어의 격조사와 몽골어의 대격 어미는 문법적인 기능으로 봤을 때 차이가 없다(강남옥, 2010, p.243).

미로 번역될 수 있다는 것을 자세히 살펴볼 수 있다. 한국어 목적격 조사는 몽골어로 번역이 될 때 대부분은 몽골어의 대격 어미랑 대응되지만 경우에 따라 방향격 어미 ‘-руу,-руу,-луу,-луу’[에/(으)로], 여격/처격 어미 ‘-д/г’[에/에게], 도구격 어미 ‘-аар,-ээр,-оор,-өөр’[(으)로], 공동격 어미 ‘-тай,-тэй,-тэй,-тэй’ [와/과], 탈격 어미 ‘-аас,-ээс,-оос,-өөс’[에서], 관형격 어미 ‘-ын/-ийн/-н’[의] 로 번역이 될 수 있다. 그러므로 모든 한국어 목적격 조사를 단순히 몽골의 대격 어미로 번역하는 것은 번역상 오류가 발생할 수 있다.

또한 서술어가 목적어를 필수적으로 요구하는 타동사인지 목적어를 요구하지 않은 자동사인지에 따라서도 목적격 조사를 번역할 것인지 결정된다. 예를 들어, 한국어에서 타동사인 동사 ‘만나다’, ‘좋아하다’, ‘싫어하다’, ‘민다’, ‘슬퍼하다’ 등은 몽골어에서는 자동사이므로 이때의 목적격 조사는 번역하지 않아야 한다.

기존의 기계분석에서의 연구들이 목적격 조사를 한 가지로만 번역한다는 문제점이 있기 때문에 본 논문에서는 형태소 분석과 동시에 품사 및 동형의이어 태깅 시스템인 유태거(UTagger)를 기반으로 하여 더욱 정확하게 의미를 분석하여 한국어 목적격 조사에 대응되는 몽골어 격 어미 결정 방법을 제안하고자 한다.

[10]에 의하면 한국어 형태소를 분석하여 실질형태소에 대해 10개의 언어로 대역어를 제공하며, 이 중에서 몽골어를 선택하면 몽골어의 명사, 동사, 형용사 등의 몽골어 대역어를 제시하지만 조사는 번역을 하지 않는다.

본 논문에서는 조사 중 목적격조사를 중심으로 한국어-몽골어 번역 방법을 제안하고자 한다. 한국어-몽골어간 번역에 대한 언어학 논문[14]에서 몽골인 한국어 학습자들의 구어 자료에서 나타난 격조사 오류 중 목적격 조사의 대치 오류 빈도가 많았음을 주장하였다. 이에 본 논문에서도 한국어-몽골어간 번역에서 문제가 많은 목적격 조사에 대한 문제를 논의 대상으로 삼았다. 또한 목적격 조사의 정확한 기계 번역을 위한 프로그램을 제안하고자 한다. 이 프로그램을 기존의 유태거의 번역 작업과 비교 실험을 통해 본 논문에서 제안한 방법의 성능을 검증할 것이다.

본 논문의 구성은 다음과 같다. 2장에서는 한국어-몽골어 기계 번역 및 한국어-몽골어 격조사에 관한 연구들에 대해 살펴보고, 3장에서는 본 논문에서 제안하고자 하는 한국어 목적격 조사에 대응되는 몽골어 격 어미 결정 방법을 제시하도록 하겠다. 다음으로 4장에서는 본 논문에서 제안한 시스템의 성능을 측정하는 실험 및 결과를 살펴보고, 마지막으로 5장에서는 결론과 향후 연구를 제시하도록 하겠다.

2. 관련 연구

본 논문에서는 현재까지의 한국어 몽골어 간 기계 번역에 대한 연구와 한국어 몽골어 격조사 비교 연구들을 살펴보았

다. 또한 온라인 기계 번역 틀이 있는데, 그 예로는 구글의 문장 번역에서 한국어 목적격 조사를 몽골어로 번역한 결과를 살펴보았다.

먼저 한국어-몽골어 기계번역에 대한 논의를 살펴보자.

[6]에서는 기존 한국어 형태소 분석 방법을 토대로 몽골어와 한국어를 비교해서 몽골어 형태소 분석에 적용하였고, 한국어-몽골어 문장 기계 번역 시스템의 구성을 설계하고 흐름도를 제안하였다. 시스템은 입력 블록, 사전 블록, 번역 블록, 번역 메모리 블록으로 구성하였고 이 중에서 제일 중요한 블록은 번역 블록이며, 사전 블록을 참조해서 형태소 분석 모듈, 문형 분석 모듈, 구문 분석 모듈을 통해 문장을 분석하였다. 정확한 번역을 위해 체언, 용언의 변환 성분들을 포함한 성분 사전 모듈(IDM: ingredient Dictionary Module)을 정리하고 사전 블록에 추가했다. 하지만 [6]에서는 다른 논문과 같이 한국어 목적격 조사를 몽골어 대격 어미로만 대응된다고 하는 문제점이 있다.

[8]에서는 규칙 기반 방법 중 transfer-based 번역 방법을 선택해 한국어 동사를 몽골어 동사로 번역 하는 transfer rules을 개발하였다. Transfer-based 번역 모델은 주로 feature 구조를 intermediate representation로 사용하지만, 이 연구에서는 constituent 구조는 문법적 특성이 같은 언어에서 사용한다. 동사 번역에만 중점을 두고 있어 목적격조사의 번역 오류와 같은 점을 나타내지 않는다.

[9]에서는 한국어-몽골어 기계 번역 시스템을 위하여 한국어 조사 및 어미의 대응하는 몽골어 조사와 어미를 비교하여 1:1, 1:N, 비정상적, 대응하지 않다 등으로 분리하였다. A. Enkhsaruul(2007)도 한국어 목적격 조사를 1:1 부분에 분리하여 다른 격 어미로의 번역을 포함하지 않았다. 다른 격 조사를 여러 격 어미로 번역이 될 수 있다는 것을 밝혔지만 언제 어떤 격 어미로 번역이 된다는 해석 방법은 없었다.

다음으로 한국어-몽골어간 언어학 논의를 살펴보자.

[13]에서는 몽골어의 ‘Ыг/ийг г/г’ 목적격 어미와 한국어의 ‘을/를’ 목적격 조사를 대조 분석하여 유사점 및 차이점을 밝히고, <외국인을 위한 한국어 문법2>에 용례에 대응하는 순서에 따른 한국어 조사 ‘을/를’의 의미 분석을 하였다. 또한 한국어 목적격조사를 몽골어로 9가지 다르게 번역될 수 있다는 것을 Table 1과 같이 각 각 예를 들어 밝혀주었다. Table 1에서 강조의 뜻을 나타낸 경우와 행위의 기준이 되는 대상임을 표현할 때 쓰이는 경우는 언급하지 않았다.

[15]에서는 한국어의 격조사와 몽골어의 격어미를 대상으로 하여 문장성분의 격을 표시하는 방식을 검토하면서 통사·의미적인 특성을 비교하여 공통점과 차이점을 밝혔다. 한국어의 목적격조사가 이동 동사의 도달점 성분을 목적격조사로 표시할 수 있지만, 몽골어에서는 대격 어미가 사용되지 못한다는 것을 밝히고 한국어에서 있는 동쪽목적어가 몽골어에서는 없다는 것을 예를 들어 설명했다.

Table 1. Multiple Mongolian Object Case Markers Corresponding to the Korean Object Case Markers

Korean sentence	Mongolian sentence
아이가 음료수를 마신다	아이가 음료수를 마신다
그는 어제 대구를 출발한다	그는 어제 대구에서 출발한다
저는 일요일마다 교회를 다녀요	저는 일요일마다 교회에 다녀요
우리 과는 내일 강원도로 답사를 떠난다	우리 과는 내일 강원도로 답사로 떠난다.
자녀를 위해 교육보험 하나쯤은 들어 보세요.	자녀의 위해 교육보험 하나쯤은 들어 보세요.
늦었으니 택시를 타고 갑시다	늦었으니 택시에 타고 갑시다

[16]에서는 몽골인 한국어 학습자들의 한국어의 목적격조사 ‘을/를’을 누락한 오류, 대치 오류, 첨가 오류를 설문조사를 통해 밝히고 이유를 설명했다. 여기에서는 “를 위해서”, “마음을 먹다” 등 관용적 표현에서 많이 누락되는 것을 보이고 ‘사랑하다’, ‘좋아하다’, ‘싫어하다’ 등의 동사는 몽골어에서 여격/처격 어미 ‘-г/т’를 쓰고 한국어의 부사격 조사 예, 예게, ‘에서’의 기능을 담당한다고 한다. 이 연구에서 몽골인 학습자들은 목적격조사 ‘을/를’을 특정 동사 앞에서 부사격 조사로 대치하는 오류를 몇 가지만 보였지만 자세히 없는 것으로 보인다.

마지막으로 번역 툴을 유태거와 구글 번역을 중심으로 살펴보도록 하겠다.

[10]에 따르면 유태거는 세종 말뭉치를 학습 기반으로 사용하여 형태소 분석과 동시에 품사 및 동형이의어를 분별한다. 품사와 동형이의어 모두 96.44%의 태깅 정확률을 보였다. 이 시스템은 문장에서 어절의 형태소와 동형이의어를 좌우 문맥 정보에 의해 결정한다. 유태거의 출력은 격조사를 정확히 번역하지 못하지만 명사나 형용사, 동사에 대한 번역은 정확률이 높다. 유태거의 몽골어 번역 화면은 Fig. 1과 같다.

```

버스를 타다
버스(автобус) 를
EH(унах, суух) да
버스_020001/NNG+를/JK0 EF_020100/VV+다/EF

1 2 버스를 버스_020001/NNG + 를/JK0
2   타다   EF_020100/VV + 다/EF

버스를 FWD AF
[ 1] 버스_02[bus_합승_자동차]/NNG+를/JK0 --- 321.18380070
    
```

Fig. 1. Output of the UTagger

구글 번역은 구글이 무료로 제공하는 영어와 100개 이상의 다른 언어 간의 단어, 구문을 즉시 번역하는 기계 번역 서비스이다. 구글 번역기는 동형이의어 분별 기능이 약하고 아주 쉬운 일상생활 문장의 한국어 몽골어 격 조사 부분은 비교적 정확하게 번역하지만, 조사 번역에서 오류가 많은 편이다. 다음 Table 2는 구글에서 한국어 문장을 몽골어로 번역한 결과이다.

Table 2를 살펴보면 1번 2번 문장은 정확하게 잘 번역이 되었지만 3번 4번 문장에는 오류가 나타났다. 3번 문장의 경우 ‘여행을’ 어절이 빠져 있고, 4번 문장의 경우 한국어 동사 ‘사랑하다’는 몽골어에서 도구격 어미 ‘-аар, -ээр, -оор, -өөр’의 주요 기능과 대응되는 한국어의 부사격 조사 ‘(으)로’ 번역되어야 하는데 [15] ‘에서’로 잘못 번역되어 있다. 즉 ‘사랑하다’ 동사는 도구격 어미랑 대응되어야 하는데 구글에서는 탈격 어미인 ‘에서’로 번역한 것이다.

3. 한국어 목적격 조사의 몽골어 격 어미 결정 방법

다음으로 본 논문의 몽골어 격 어미 결정 방법을 제안하기 위해 사용한 도구와 데이터 테이블 설계 방법을 살펴본다. 먼저 본 논문에서 사용된 도구인 유태거와 한국어기초사전에 대해 소개한다.

3.1 유태거(UTagger)

유태거는 한국어 품사 태깅뿐만 아니라 동형이의어를 분별하는 태깅 시스템이다. 먼저 한국어 형태소를 분석하는데, 형태소 분석은 2 단계로 나뉜다. 첫 단계에서는 세종말뭉치에서 기분석 부분 어절 사전을 구축하고 이를 활용해서 하나의 어

Table 2. Translated Results of Object Case Markers on the Google

No.	Source sentences	Output by “Google translator”	Target sentences by human
1	이 물을 누가 마셨어?	Энэ усыг хэн уусан бэ? 이 물을 누가 마셨어?	Энэ усыг хэн уусан бэ? 이 물을 누가 마셨어?
2	난 버스를 탔다	Би автобусанд суув. 난 버스에 탔다.	Би автобусанд суув. 난 버스에 탔다.
3	난 여행을 떠났다	Би явсан. 난 떠났다.	Би аялалаар явсан. 난 여행으로 떠났다.
4	난 몽골 사람을 자랑한다	Монголчуудаас бахархдаг 몽골인들에서 자랑한다.	Би монгол хүнээр бахархдаг 난 몽골 사람으로 자랑한다.

절을 분석하여 형태소와 품사 그리고 동형이의어 번호를 가진 후보들을 생성한다. 두 번째 단계에서는 생성된 후보들 중에서 하나의 적절한 후보를 선택하는데, 자료 부족 문제를 해결하기 위해 단계별 전이모델을 사용하며, 각 전이모델마다 다른 가중치를 적용한다. 또한 부분적으로 전이빈도가 0인 경우를 위해 최소 전이점수를 계산하는 예외 처리 루틴을 포함한다.

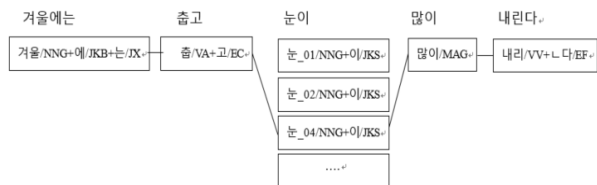


Fig. 2. Example of the Homograph Tagging

이 시스템은 문장에서 어절의 형태소와 동형이의어를 좌우 문맥 정보에 의해 결정한다. Fig. 2에서는 ‘눈’과 ‘내리다’의 형태소 태깅 결과는 동일하지만 ‘눈/NNG’는 {감각 기관, 얼음의 결정체, 식물, 이어 이론 구멍} 등의 동형이의어를 가지며, ‘내리/VV’도 {눈/비/이슬이 오다, 탈것에서 밖이나 땅으로 옮겨 서거나 나오다, 막이나 커튼 등이 위에서 아래로 옮겨 가다, 명령이나 지시 등을 알려 주다, 먹은 음식물 등이 소화되다}의 동형이의어를 가진다. 그러므로 ‘눈’과 ‘내리다’와 같이 동형이의어를 가진 경우 주변 문맥 정보에 의해 각 단어의 형태소가 정확하게 분석되어 단어의 문맥에 맞는 의미로 형태소를 분석 태깅하고 있다.

3.2 한국어기초사전

한국어기초사전은 한국어 교육의 기본이 되는 5만 어휘를 수록한 한국어 학습용 웹 사전이다. 한국어 공부를 시작한 학

습자를 위해 쉬운 뜻풀이, 일상생활에서 자주 쓰는 예문 및 문법과 표현을 제시하고, 어휘 별로 다중 매체 정보(삽화, 사진, 동영상, 발음 등), 관련어 정보를 연결하는 등 한국 언어 문화 교육 정보를 최대한 모은 사전이다.

‘한국어기초사전’은 러시아어, 몽골어, 베트남어, 스페인어, 아랍어, 영어, 인도네시아어, 일본어, 타이어, 프랑스어 등 10개 언어로 번역한 최초의 한국어 학습용 이중 언어화 사전이기도 하다. 학습자의 언어와 문화를 반영하면서 뜻풀이를 하고, 언어권별 사용 환경에 맞추어 만들어졌다.

한국어기초사전의 한국어-몽골어 학습 사전에는 검색한 단어의 여러 뜻풀이, 발음, 설명, 예문이 나오는데, 뜻과 설명은 몽골어로 번역이 되어 나오지만 예문은 번역되지 않았다. 다음 Fig. 3은 한국어-몽골어 학습 사전의 화면이다.

유태거는 표준국어대사전을 바탕으로 형태소 분석과 동형이의어, 다의어 정보를 분석하여 정보를 제공하고 있다. 즉, 유태거의 동형이의어 번호 체계는 국립국어원의 표준국어대사전을 따른다. 대역어 제공을 위해 표준국어대사전과 한국어기초사전을 어휘 별로 뜻풀이별로 매핑하였다. 이 매핑 테이블을 통해 몽골어 대역어를 제공하고 있다. 본 논문에서는 외국인 한국어 학습자를 위한 용도로 만들어진 한국어기초학습사전을 기반 사전으로 하여 데이터를 구축하였지만, 번역틀은 유태거에 탑재된 표준국어대사전과 한국어기초사전의 매핑 테이블을 사용하였다.

3.3 데이터 테이블 설계

한국어 몽골어 기계 번역 연구를 보면 한국어 목적격 조사를 몽골어 대격 어미로만 번역했지만, 한국어 몽골어 언어학 연구에서는 몽골어 대격 어미뿐만 아니라 다른 격 어미로 번역이 될 수 있다는 것을 보여주고 있다. 기존의 언어학 연구



Fig. 3. Learner’s Dictionary of the Korean-Mongolian

에서 번역되는 것을 살펴보면 목적격 조사가 동사에 의하여 달라지고 있는 것을 알 수 있다.

본 논문에서는 실험을 위해 한국어 어휘 중 고빈도순으로 목적격 조사를 논항으로 가지는 동사 350개를 대상으로 추출하였다. 350개의 동사를 몽골어 격 어미와 연결하는 작업을 통해 80%는 대격 어미로 매핑되었고 나머지 20%는 대격 어미 외의 격 어미로도 매핑되었다. 본 논문에서 중점적으로 연구한 대상은 대격 어미로 매핑되지 않은 20%의 동사이다.

또한, 대격 어미로 매핑되지 않은 경우를 살펴보면 한국어 동사에 의해서 결정되는 경우(55%)와 번역된 몽골어 동사에 의해 결정되는 경우(40%)가 대부분이고 명사에 의해 격 어미가 결정된 경우(5%)는 제일 적게 나타났다. Fig. 4는 대상 어휘에 대한 정리이다.

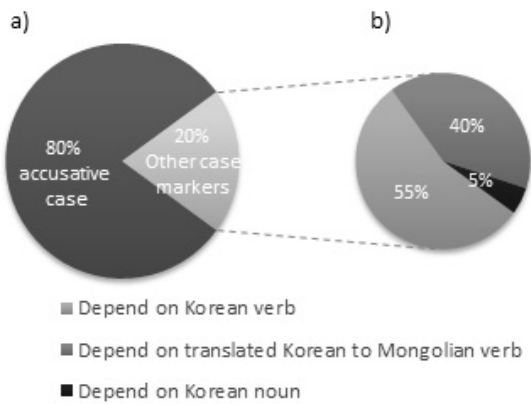


Fig. 4. (a) Translated Ratio of the Korean Object Case Markers to the Mongolian Object Case Markers, (b) Translated Ratio of the Korean Object Case Markers to the Other Mongolian Object Case Markers

한국어 목적격 조사가 바로 몽골어 대격 어미와 대응되는 80%에 해당하는 예 중 ‘만들다’가 있다. 한국어 동사 ‘만들다’는 한국어기초사전에서 14개의 의미를 가진 다의어이며 각 의미별로 다른 몽골어 동사로 번역되지만, 번역된 몽골어 동사들이 대격 어미와 어울려 쓰이기 때문에 실험에서 제외하였다.

가. 로봇을 만들면 인간 생활이 더 편리해질 것이다.

Роботыг бүтээвэл хүн төрөлхтөний амьдрал илүү хялбар болох болно.

나. 사람에게 대한 믿음이 있어야 좋은 사회를 만들 수 있다.

Хүнд итгэх итгэл бай ж бай ж сай хан ний гми й г бий болгож чадна.

다. 김 대리는 회의 시간에 쓸 서류를 만드느라 바빴다.

Ким орлогч хуралд хэрэглэгдэх материалыг боловсруулах гээд завгүй бай сан.

여기에서는 ‘만들다’ 동사가 몽골어로 번역이 다르게 되지

만 조사는 변하지 않는다. (가)에서는 ‘만들다’의 몽골어 번역이 ‘бүтээх’ 이고 (나)에서는 ‘бий болгох’, (다)에서는 ‘хий х’ 이다.

한국어 목적격조사가 몽골어의 다른 격 어미로 번역이 되는 20%의 동사들은 몽골어 격 어미를 결정할 때 다음 3 가지 조건으로 나누어 데이터 테이블을 설계하였다.

1) TYPE 1. 한국어 동사에 의해 몽골어 격 어미를 결정하는 경우

TYPE 1은 한국어 동사의 목적격 조사의 몽골어 격 어미가 항상 1:1로 대응되는 경우이다. 동사는 ‘믿다, 좋아하다, 싫어하다, 돕다, 자랑하다, 만나다, ...’가 있다. 이들 동사는 논항에 ‘을/를’을 가지는 타동사이다. 그러나 몽골어로 번역되면 그 대응되는 형태가 꼭 대격 어미는 아니지만, 항상 하나의 격 어미로만 번역이 된다. 그러므로 이러한 동사와 대격 어미를 테이블에 등록하여 번역시 바로 몽골어 격 어미를 제공하게 하였다. 동사 ‘믿다’의 몽골어 번역 예를 통해 격 어미 결정 예를 살펴보자.

가. 민준이는 종교는 없지만 신의 존재를 믿는다.

Минжү шүтлэггүй боловч бурхан бай дагт итгэдэг.

(번역: 민준이는 종교는 없지만 신의 존재에 믿는다.)

(가)의 동사 ‘믿다’는 몽골어로 한 가지 의미로만 번역이 되기에 바로 한국어 동사만 보고 목적격 조사를 결정할 수 있는 것이다. ‘믿다’에 대응되는 몽골어 동사 ‘итгэх’는 한국어 부사격 조사 ‘에게/한테’에 대응되는 ‘д/т’에 어울려 쓰인다[13]. 따라서 ‘믿다’의 앞에 온 목적격 조사는 체언에 상관 없이 부사격 조사 ‘д/т’로 번역이 된다.

2) TYPE 2. 번역된 몽골어 동사에 따라 몽골어 격 어미를 결정하는 경우

TYPE 2는 다의어 동사이고 다의어 의미에 따라 다른 몽골어 대역어를 가지면서 대역어에 따라서 격 어미도 달라지는 경우이다. ‘마시다, 구하다, 타다’ 등이 대표적인 예이다. ‘마시다’를 예로 살펴보자.

가. 나도 시원한 공기를 좀 마시고 싶었어.

Би ч бас цэвэр агаараар амьсгалахыг хүссэн.

(번역: 나도 시원한 공기로 좀 마시고 싶었어.)

나. 누가 여기 있는 물을 마셨어요?

Хэн энд бай сан усыг уусан бэ?

(번역: 누가 여기 있는 물을 마셨어요?)

한국어로는 ‘마시다’는 “물이나 술 따위의 액체를 목구멍으

로 넘기다.”와 “공기나 냄새 따위를 입이나 코로 들이쉬다.”의 다의어이며, 몽골어 대역어도 의미에 따라 “공기를 마시다”의 ‘마시다’는 ‘амьсгалах’, “물을 마시다”의 ‘마시다’는 ‘уух’로 번역된다. 다른 단어로 번역이 된다는 점이 격 어미 결정 요인이 되는 것이다. “공기를 마시다”의 번역된 몽골어 동사인 ‘амьсгалах’은 도구격 어미 ‘аар’를 쓴다. ‘аар’에 대응되는 한국어의 조사가 바로 부사격 조사 ‘(로)’이다. 반면에 “물을 마시다”의 번역된 몽골어 동사인 ‘уух’은 대격 어미가 생략된다.

3) TYPE 3. 한국어 명사에 의해 몽골어 격 어미를 결정하는 경우

TYPE 3은 한국어 다의어 동사가 하나의 몽골어 동사로 번역이 되지만, 목적격 조사가 결합된 명사에 의해 격 어미가 결정되는 경우이다. 해당 예는 다음과 같다.

- 가. 아이는 멀리서 나를 보고 달려와 품에 안겼다.
Хүүхэд холоос намайг г хараад гүй н ирж тэврүү лэв.
(번역: 아이는 멀리서 나를 보고 달려와 품에 안겼다.)
- 나. 나는 거울을 보면서 머리를 가다듬었다.
Би толинд харангаа үсээ зассан.
(번역: 나는 거울에 보면서 머리를 가다듬었다.)
- 다. 사전을 보고 찾았다.
Тольноос хараад олсон.
(번역: 사전에서 보고 찾았다.)

(가)~(다) 모두 동사 ‘хараад’로 번역된다. 한국어 문형은 ‘[~을] 보다’이다. 그러나 몽골어로 번역하면 (가)의 ‘보다’는 대격 어미로, (나)의 ‘보다’는 부사격 어미 ‘에’로 (다)의 ‘보다’는 부사격 어미 ‘에서’로 결정된다. 그러므로 앞에 오는 체언에 따라 격 어미가 달라진다.

본 논문을 위해 만든 데이터베이스에 있는 테이블은 Fig. 5와 같으며, 각 필드는 Fig. 6에서 설명하였다.

mongolian_case 필드에 있는 몽골어 격 어미 부분에 교체할 수 있는 모든 어미를 저장하고 있다. 한국어 목적격 조사는 ‘을/를’이고 자음이나 모음이냐에 따라 교체하는데, 몽골어 격 어미의 교체 조건은 한국어의 격 조사보다는 더 복잡하다.

한국어의 목적격조사는 ‘을/를’이 있으며 받침 여부에 따라 음운론적 이형태를 구성한다. 몽골어 대격어미에는 ‘-ЫГ /-ИЙ Г/ -Г / Ø’ 등이 있는데 모음의 길이(단모음/장모음/이중모음), 끝 자음의 형태, 선행체언의 성(남성/여성)에 따라 모음 조화 조건에 맞추어 이형태로 실현된다. ‘-ЫГ /-ИЙ Г/ -Г’은 몽골어 생략이 될 때 교체하고, ‘Ø’은 몽골어 대격 어미가 생략되지 않는다는 의미다. 대격 어미가 반드시 쓰이는 경우에

는 목적어가 고유 명사일 경우와 인칭 대명사나 지시 대명사가 한정하고 있는 명사일 경우, 그리고 목적어가 인칭 대명사나 형용사, 수사일 경우이며 목적어가 불특정한 것을 나타내는 경우에는 대격 어미가 생략된다[16].

verb_id	Korean_verb	type	mongolian_case	noun	mongolian_verb
69	민다	1	-л/г		ингэх (민다)
70	자랑하다	1	-аар /-ээр /-оор /-өөр		бахархах(봉내다)
71	갈기다	2	-руу/ -рүү/-луу/ -лүү/-аал/-өөд		алгалах (때리다)
72	갈기다	2	-ыг /-ийг /-г / Ø		гяглах (글씨를 마구 쓰다)
73	갈기다	2	-аар /-ээр /-оор /-өөр		буулах (충 쓰다)
74	위하다	1	-ын /-ийн/-н		төлөө (위하다)
75	때리다	3	-ыг /-ийг /-г / Ø	뺨, 머리	цохих(때리다)
76	때리다	3	-руу/ -рүү/-луу/ -лүү/-аал/-өөд	사람, 동생	цохих(때리다)
...					

Fig. 5. Structure of the Table

verb_id: Id field

korean_verb: Korean transitive verb

type: Determination of Mongolian's suffixes type. It can be included following types: 1, 2, and 3

- 1: Depend on Korean verb
- 2: Depend on translated Mongolian verb
- 3: Depend on Korean noun.

mongolian_case: Accessible Mongolian case markers

noun: Korean noun

mongolian_verb: Mongolian verb which has different meaning from Korean

Fig. 6. Role of the Fields

- 가. 나는 책을 읽었다.
Би ном(Ø) уншсан.
- 나. 나는 이 책을 읽었다.
Би энэ номыг уншсан.

3.4 조사 결정 과정

한국어 목적격조사의 몽골어 격 어미 결정 과정의 전체 시스템 구조는 Fig. 7과 같다.

본 논문의 연구 방법은 다음과 같다. 우선 유태거를 기반으로 분석된 문장을 어절 단위로 확인한 후 어절 중 목적격 조사를 찾는다. 목적격 조사를 찾은 후 후행 어절의 형태소를 분석하여 동사 여부를 확인한다. 후행 어절이 동사가 아니라면 다음 어절로 넘어가 다시 동사여부를 확인하는 과정을 거치도록 하였다.

이러한 일련의 과정을 통해 동사를 찾은 뒤 데이터 테이블을 통해 몽골어 격어미를 결정한다. 몽골어 격 어미 결정과정은 3개의 부분으로 구분하고 만약 3개의 특징이 드러나지 않는다면 몽골어 대격 어미(-ЫГ /-ИЙ Г/ -Г / Ø)로 결정하도록 프로그래밍하였다.

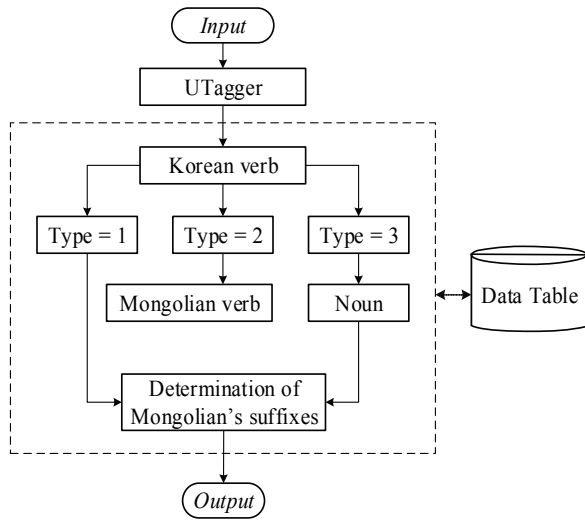


Fig. 7. Process of the Object Case Marker Decision

테이블 결정 과정은 Fig. 7에 제시된 TYPE 1, 2, 3이다. TYPE 1은 한국어 동사 정보에서 바로 몽골어 격어미를 결정할 수 있는 것이고, TYPE 2는 몽골어로 번역을 한 뒤에 격 어미를 변별하여 결정하는 것이며, TYPE 3은 동사의 번역어로는 격 어미 결정이 어려운 경우로 동사와 공기관계에 있는 명사를 찾아 그 명사의 의미에 따라 몽골어 격 어미를 결정하는 것이다.

출력 부분에는 입력된 한국어 문장의 품사 및 동형이의어의 의미를 분석한 다음 몽골어로 번역된 결과와 격 조사까지 출력된다. Fig. 8은 그 예이다.

“버스를 타다” 문장의 형태소 분석 및 품사 동형이의어, 몽골어 번역, 목적격 조사까지 번역해 출력한다.

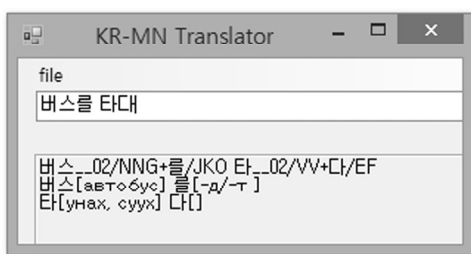


Fig. 8. Final Output

4. 실험 및 결과

4.1 실험 결과

본 논문에서 제안한 한국어 목적격 조사의 몽골어 격 어미 결정 방법의 성능을 평가하기 위하여 한국어기초사전에서 목적격 조사를 포함한 2,040개 문장을 추출하였다. 한국어기초사전에는 뜻풀이는 몽골어로 번역이 되어 있지만 그와 관련된 예문들은 번역이 되어 있지 않다. 한국어기초사전의 뜻

풀이는 대체로 {무엇을 말한다, 어떤 것을 표현한다} 등의 표현으로, 번역된 몽골어 뜻풀이 문장에서 목적격 조사의 종류는 한두 가지로만 나타나기 때문에 실험 데이터로 추출하기는 부족했다. 따라서 번역되지 않은 예문에서 목적격 조사를 가진 2,040개 문장을 추출하였다. 목적격 조사를 사람이 번역한 다음 기존에 유태거와 본 논문의 제안한 방법으로 각각 비교실험 하였다. 실험 데이터의 전체 목적격 조사 중에서 번역된 격조사에 대한 평가 기준은 정확도를 기준으로 하였고 다음과 같이 계산하였다.

$$\text{정확률} = \frac{\text{정확하게 분석된 목적격 조사 수}}{\text{전체 목적격 조사 수}}$$

한국어에서는 ‘잠자다/잠을 자다’, ‘공부를 하다/공부(를) 하다/공부하다’와 같은 동족목적어를 가지는 예가 상당히 많다. 그러나 몽골어는 이러한 동족목적어를 가지는 동사가 없다. 그러므로 추출한 2,040개의 문장 안에는 동족목적어를 가진 예는 포함되어 있지 않다. 실험 방법은 문장마다 있는 목적격 조사를 발견하고 미리 번역해 놓은 데이터와 시스템을 통해 번역한 목적격 조사 부분을 비교하였다. 실험 결과는 Table 3과 같다.

Table 3. Result of the Examination

Approach	Output accuracy
UTagger	46.9%
UTagger+Josatranslator	88.3%

4.2 오류 유형 분석

잘못 번역된 오류의 유형을 살펴보면, 유태거 의미 분석이 잘못 된 경우, 데이터베이스에 등록되지 못한 명사의 경우 번역 오류가 나타났다. type 별로 정확률을 계산하여 다음 Table 4에 제시하였다.

TYPE 1에는 54개의 동사가 있고 그 동사를 포함한 712개의 문장을 실험해 3개의 오류가 발생했다. TYPE 2는 40개의 동사가 등록되어 있으며 881개의 문장을 실험하였으며 유태거의 의미분별 오류로 인해 발생했다. TYPE 3은 데이터 테이블에 등록되지 못한 명사에 의해 오류가 발생했다. 오류 갯수에서 괄호 속 숫자는 유태거가 형태소분석을 잘못하여 발생한 경우이다(4.2.4에서 설명).

Table 4. Data of the Examination

type	verb	sentence	error	accuracy
TYPE 1	54	712	3(2)	99.5%
TYPE 2	40	881	213(5)	75.8%
TYPE 3	11	447	21(16)	95.3%
total	105	2,040	237(23)	88.38%

1) 의존관계 오류(TYPE 1)

TYPE 1에 해당된 동사 “돕다”를 포함한 “부모님 가게를 일삼아서 돕다.” 문장에 의존관계 오류가 있었다. “가게를 돕다”로 격조사를 “[-д/-г]”로 결정해야 했지만, “가게를 일삼다”로 격조사를 “[-ьг/-ий г/-г]” 결정해 오류가 났다. 다음은 결정 결과의 예이다.

가. 부모님/NNG 가게/NNG+를/JKO 일삼/VV+아서/EC 돕/VV+고/EC 있_01/VX+어/EF+./SF

부모님[эцэг эх]
 가게[дэлгүүр] 를[-ьг/-ий г/-г]
 일삼[ажил болгох] 아서
 돕[туслах, дэмжих] 고
 있[бай х] 어 .

이러한 오류는 격 어미 결정 과정을 좀더 정교하게 보완하여야 해결될 것이다.

2) 유태거 오류(TYPE 2)

‘갈기다’은 Table 5와 같이 여러 뜻이 있다.

Table 5. Various Meaning of the Verb ‘갈기다’

Korean	Mongolian	Meaning
갈기다	алгадах	to slap
	татлах	to scribble
	буудах	to shoot

유태거에서 분석된 동사 ‘갈기다’의 번역이 ‘буудах’로 나와야 목적격조사가 도구격 어미 ‘aav’로 결정이 되는데, 번역이 ‘алгадах’로 되어 몽골어 동사 ‘алгадах’와 대응되는 방향격 어미 ‘pyu’로 결정되는 오류가 있었다. Table 6이 그 예이다.

Table 6. Type of the Error

대포를 갈기다		
Target sentence	Их буугаар буудах	to shoot a gun (대포를 사용해서 갈기다)
result of translation	Их бууруу алгадах	to slap a gun (대포로 향해서 때리다)

3) 명사의 데이터 테이블 등록 오류(TYPE 3)

데이터 테이블에 등록되지 않은 명사들로 인한 오류가 발생하였다.

‘때리다’의 경우에는 몽골어 번역이 똑같이 ‘цохих’로 되지만 명사에 따라 목적격조사가 달라진다.

가. 친구를 때렸다.
 Най зьг цохисон.

나. 빵을 때렸다.
 Хацарлуу цохисон.

(가)처럼 친구, 동생, 엄마 등 사람을 가리키는 명사 뒤에는 ‘цохих’ 동사가 들어간 경우에는 대격 어미 ‘ьг’를 취하고, (나)처럼 사람의 신체 부위를 가리키는 명사 뒤에는 ‘цохих’ 동사가 들어간 경우에는 방향격 어미 ‘pyu’[에,(으)로]를 가지게 된다. 데이터 테이블에 사람이거나 신체 부위를 가리키는 명사를 가급적 등록했지만 등록되지 않은 명사가 있어 오류가 발생하였다. 이러한 문제는 어휘의미망을 활용하여 특정노드 하위 명사들을 같이 처리하도록 하는 방법이 있을 수 있으나, 이는 다음 연구과제에서 다룰 예정이다.

4) 유태거 형태소분석 오류

유태거가 {각인하다(2회), 어려워하다(5회), 보다(16회)}의 3가지 어절을 잘못 형태소 분석하였다.

첫째 ‘각인하다’는 “성경의 구절을 각인하여 교회 ...” 문장에서 ‘각인하여’를 ‘각인하/VV’로 분석하지 않고 ‘각인_04/NNG+하/XSV’로 분석하여 테이블에서 검색하지 못했다.

둘째, “어려워하다”는 유태거가 ‘어려워하/VV’와 ‘어렵/VA+어/EC+하_01/VV’의 두가지로 분석하였으며, ‘하_01/VV’의 대역어를 선택하여 오류가 발생하였다.

아버지/NNG+를/JKO 어렵/VA+어/EC+하_01/VV+다/EF+./SF

아버지[aav] 를[-ьг/-ий г/-г]

어렵[хэцүү, хүнд, бэрх] 어[] 하[гүй лдэх, хий х, гүй ц этгэх] 다[] .

셋째, ‘보다’는 다음과 같이 ‘보_01/VV+다/EF’로 분석하지 않고 ‘보다_02/MAG’ 부사로 분석하였다.

상황_02/NNG+을/JKO 보다_02/MAG+./SF

상황[бай дал, нөхцөл, төлөв, явц бай дал] 을[-ьг/-ий г/-г]

보다[илүү] .[]

5. 결론

최근 기계 번역에 대한 연구들이 활발하게 이루어지고 있다. 인터넷에서 무료로 번역해 주는 구글, 파파고 등의 시스템도 있다. 그러나 한국어-몽골어 기계번역 시스템은 많이 없는 편이다. 현재 연구되어 있는 한국어-몽골어 기계번역 연구들은 한국어 목적격 조사와 몽골어 대격 어미로 바로 대응하여 지정하는 문제점이 있었다. 본 논문은 더 정확한 번역을 위해서 기존에 있는 유태거 시스템을 기반으로 하여 한국어 목적격 조사의 몽골어 격 어미 결정 방법을 제안하였다. 제안한 방법에서 기존의 언어학 연구들을 살펴보고 목적격 조사가 여러 격 어미로 번역될 때 가진 특징을 고려하여 데

이터 테이블을 설계하였고, 유태거에서 문장의 형태소와 의미 분석, 몽골어 번역 등을 통하여 출력된 문장의 격조사 부분을 설계한 테이블을 통해 결정하였다.

본 논문에서 제안한 방법의 우수성을 검증하기 위해서 한국어기초사전에서 추출한 문장들을 통해 유태거와 비교 실험하였고 그 결과 유태거는 46.9%, 본 논문에서 제안한 방법은 88.38%로 정확률에 큰 차이를 보였다. 실험 결과를 보면 한국어기초사전에서 추출한 문장들이 여러 격조사로 번역이 되는 문장들을 정확히 선택하여 추출했다는 것을 알 수 있었다. 12%의 오류를 분석한 결과 유태거의 번역 오류, 명사를 데이터 테이블에 등록하지 않은 오류 등이 있었다. 더욱 정확한 번역을 위해 데이터 테이블 내 데이터를 확대하면 더 좋은 결과를 얻을 수 있을 것이다. 또한 본 제안 방법은 다른 한국어 격조사에도 쓰일 수 있으며 향후 연구 과제는 모든 한국어 격조사를 정확하게 번역하도록 연구를 확대하는 것이다.

References

- [1] L. Bates, "Korean Sentence Complexity Reduction for Machine Translation," Master's thesis, University of Seoul, 2017.
- [2] W. K. Kim, "The Case information Processing and Korean-English Machine Translation," *Korean Semantics*, Vol.35, pp.23-43, Aug. 2011.
- [3] C. S. Lim, H. A. Lee, M. S. Choi, B. G. Chang, K. J. Lee, and G. C. Kim, "English to Korean Machine Translation System based on Lexicalized Grammer," in *Proceedings of Korea Information Science Society*, Vol.24, No.2, pp.161-164, Oct. 1997.
- [4] J. S. Kim, W. S. Kang, K. S. Choi, and G. C. Kim, "Design and Implementation of English Syntactic Analysis System for English-Korean Machine Translation," in *Proceedings of Korea Information Science Society*, Vol.16, No.2, pp.563-566, Oct. 1989.
- [5] K. H. Moon, J. H. Lee, J. I. Kim, and G. J. Yang, "Resolution of Word Sense Ambiguities using Collocation Pattern in Japanese-to-Korean MT system," in *Proceedings of Korea Information Science Society(B)*, Vol. 25, No. 8, pp. 1270-1280, Aug. 1998.
- [6] B. Tsend Ayush, "Design of Korean and Mongolian sentence machine translation system," Master's thesis, Kangwon National University, 2012.
- [7] Y. M. Kim, "Mongolian-Korean language Machine Translation System," *Korea Multimedia Society International Conference*, pp.50-52, Jul. 2011.
- [8] T. Gantur, "Rule-based Machine Translation from Korean Verb to Mongolian Verb," Master's thesis, Soongsil University, 2010.
- [9] A. Enkhsaruul, C. G. Song, and Y.-S. Kim, "A comparison between Korean and Mongolian eomi and josa for Korean to Mongolian machine translation system," in *Proceedings of the Korean Association of Information Science and Technology Conference*, pp.228-232, Oct. 2007.
- [10] J. Ch. Shin and C. Ock, "A Korean Morphological Analyzer using a Pre-analyzed Partial Word-phrase Dictionary," *Journal of KIISE:Software and Application*, Vol.39, No.5, pp.415-424, May 2012.
- [11] M. Altantsetseg, "A Comparative Study on the Case Markers between Korean and Mongolian languages: On Objective case and Adverbial case markers," Master's thesis, Kangwon National University, 2016.
- [12] E. Khulan, "A comparison between Korean objective proposition and Mongolian," *Journal of Korean Culture*, Vol.12, pp.105-134, 2009.
- [13] U. Bolor Erdene, "Error Analysis and Teaching Method of Case Markers 'eul/leul' for Mongolian Learners of Korean," *New Language Education*, Vol.102, pp.273-304, 2015.
- [14] L. Buyandelger, "A Study on Lexical and Grammatical Errors of Mongolian Learners of Korean," Master's thesis, Keimyung University, Jun. 2015.
- [15] B. Munkhtsetseg, "A Comparative study of Korean Case Marker and Mongolian Case Ending," Master's thesis, Chungbuk National University, 2013.
- [16] S. J. Nam, "A Study about the interlanguage appeared on the Mongolian Korean Language Learners : Focused on the Case Particle," Master's thesis, Keimyung University, 2010.



Khulan Setgelkhuu

<https://orcid.org/0000-0002-6876-1864>

e-mail : setgelkhuukhulan@gmail.com

2013년 몽골과학기술대학교 전자공학과
(학사)

2019년 울산대학교 IT융합학부(석사)

관심분야 : Machine Translation



신 준 철

<https://orcid.org/0000-0002-6793-3687>

e-mail : ducksjc@nate.com

2007년 울산대학교 컴퓨터공학(학사)

2009년 울산대학교 컴퓨터공학(석사)

2014년 울산대학교 컴퓨터공학(박사)

2014년~현 재 울산대학교 IT융합학부
연구교수

관심분야 : Korean Language Processing, Word Sense
Disambiguation, Machine Learning



옥철영

<https://orcid.org/0000-0003-0020-1040>

e-mail : okcy@ulsan.ac.kr

1982년 서울대학교 컴퓨터공학과(학사)

1984년 서울대학교 컴퓨터공학과(석사)

1993년 서울대학교 컴퓨터공학과(박사)

1994년 러시아 TOMSK공과대학 교환교수

1996년 영국 GLSGOW대학교 객원교수

2007년 몽골국립대학교 IT학부 명예박사학위

2007년~2008년 한국정보과학회 언어공학연구회 운영위원장

2008년 국립국어원 객원연구원

2016년 한글발전유공자 근정포장

1984년~현 재 울산대학교 IT융합학부 교수

관심분야: Korean Language Processing, Word Sense

Disambiguation, Machine Learning