

일한 기계 번역의 성능 향상을 위한

조동사 れる/られる의 애매성 해결

한명수*, 문경희**, 이종혁**

포항공대 정보통신 대학원*, 포항공대 컴퓨터공학과**

{mize,khmoon,jhlee}@kle.posech.ac.kr

Disambiguation of Auxiliary Verbs “*reru/rareru*” in Japanese-Korean MT System

Myoung-Su Han*, Kyong-Hi Moon**, Jong-Hyeok Lee**

Dept. of Graduate School for Information Technology, POSTECH*

Dept. of Computer Science & Engineering, POSTECH**

요약

본 논문은 고품질의 일한 기계 번역 시스템의 개발을 위해서, 코퍼스에서 높은 빈도로 나타나는 일본어 조동사 れる/られる의 애매성 해소 방법을 제안한다. 조동사 れる/られる는 동사와 접속하여, 가능, 자발, 피동, 존경의 네 가지 의미로 사용되며, 문맥에 따라서 그 의미가 결정된다. 본 논문에서는 특히, 한국어에는 존재하지 않는 간접 피동과 여격 및 대격을 동시에 가지는 수여동사의 이중피동에 중점을 두어, 동사의 결합가 정보와 구문분석을 통한 패턴 매칭 방법을 이용하여 애매성을 해소하였다. 코퍼스로부터 용법별로 추출된 300 문장을 실험 대상으로 기존의 방법과 논문의 방법을 비교, 평가한 결과 33%의 번역률 향상을 보였다.

1. 서론

한국어와 일본어는 같은 한자 문화권에 속하고, 비슷한 어순과 경어 표현 등 어휘적, 문법적으로 많은 유사성을 가지고 있다. 따라서, 이러한 특성을 최대한 활용한 일대일 대응의 직접번역방식이 일한 및 한일 기계번역 시스템의 주된 적용 방법이었다. 그러나, 고품질의 번역결과를 얻기 위해서는, 애매성을 가진 어휘에 대한 정확한 의미 결정이 반드시 필요하다. 본 논문에서 다루고 있는 れる/られる는 문장 단위에서 약 22%의 높은 빈도로 사용되며[1], 동사와 결합해서 네 가지 의미를 가진다. 이의 정확한 의미 결정을 위해서, 동사의 의미와 결합가 정보, 각 용법 별 패턴 정보들을 사용하여, 부분적인 구문분석으로 찾아낸 해당 동사의 격 정보와 대응시켜서 애매성을 해소하였다.

2. 기존 연구의 문제점

기존연구[1]에서는 <표 1>에서와 같이, 용법별로 れる/られる의 한국어 대역어를 우선 선정해서, 사전등록 시에 사전등록자의 직관에 따라 가장 빈도수가 높을 것으로

예상되는 대역어를 해당 동사 사전에 등록하였다. 그리고, 준비된 가능용법의 패턴과 대응되지 않는 경우에는 전부 직접피동으로 처리하는 방법을 사용하였다. 이 방법으로 아사히 신문기사 5,797 문장을 대상으로 실험한 결과, 87%의 번역률을 나타냈다.[1] 그러나 실험대상 코퍼스의 분야가 신문기사로 한정되어 실험의 80% 이상이 단순 피동 용법으로 나타나고 있기 때문에 れる/られる가 가진 네 가지 용법의 애매성 해소 평가로는 부족하다. 따라서 기존연구에서 처리되지 못한 간접피동과 이중피동 및 존경용법에 대한 보다 효과적인 처리방법이 요구된다.

<표 1> 기존연구에서 사용된 한국어 대역어 및 빈도수

의미	한국어대역어	빈도수
피동	지다	18.90%
	받다/당하다	4.60%
	이/해/리/기	13.10%
	되다	46.20%
가능	ㄹ/을 수 있다	3.50%
자발	지다	0.30%
존경	(으)시다	0.10%
기타	어간바꿈등	13.30%
합계		100.00%

다음은 기존 연구에서 존경과 간접 피동 및 이중 피동의 미처리로 인해 나타나는 오역 문장의 예이다

國體に天皇陛下が戦後初めて沖縄の土を踏まれます。
국가 체전에 천황 폐하가 전쟁 후 처음으로 오키나와의 땅을 밟습니다. (-> 밟으십니다)

正規表現はこの方程式の不動点として與えられる
정규 표현은 이 방정식의 부동점으로서 받는다. (-> 주어진다)

3. れる/られる의 용법 및 처리방법

이 장에서는 れる/られる의 네 가지 용법에 대해 각각 애매성 해소를 위한 정보들을 살펴보고, 이를 이용한 처리 알고리즘을 제안한다.

3.1 れる/られる의 각 용법별 애매성 해소 정보

1) 가능

私は早く起きられません.
나는 일찍 일어날 수 없습니다.

일본어에는 れる/られる 이외에, え단의 어미를 사용한 5 단동사의 가능동사 및 ことができる 등의 가능 표현이 존재하므로, 애매성 해소는 1 단동사를 중심으로 처리한다. 코퍼스를 통해서 획득된 공기 패턴 정보를 이용해서 자주 사용되는 가능의 패턴을 마련하여, 매칭되는 경우에 한해서 가능 처리를 한다. 코퍼스로부터 구한 가능의 패턴 정보는 아래와 같다.

[명사*が/は/を/も+ 동사*れる/られる* 부정표현/ように]

2) 자발

空気がきれいでおいしく感じられた.
공기는 깨끗하고 맛있게 느껴졌다.

감정을 나타내는 몇몇 동사의 활용형에서만 나타나므로, 감정 표현을 나타내는 동사의 DB 를 구축하여 애매성을 해소한다.

3) 존경

病氣ご療養中の天皇陛下の代りに皇太子殿下が出席された.
병 요양 중의 천황 폐하 대신 황태자 전하가 참석하셨다.

-なさる, お-になる, 召し上がる 등의 다른 일본어 존경 표현들에 비해 れる/られる의 존경용법이 가지는 존경의 정도는 아주 낮으므로[4,5], 존경표현을 사용하지 않는 대중매체에서 天皇, 殿下, 皇太子 등 반드시 존경으로 표현해야 하는 경우에 사용된다. 따라서 이러한 명사들의 DB 를 구축하여 れる/られる에 걸리는 주어가 이에 해당되는 경우, 존경으로 처리한다. 이를 패턴으로 나타내면 다음과 같다.

[존경대상명사*が는/ + 동사*れる/られる*]

4) 피동

<표 1>에서도 알 수 있듯이 れる/られる의 용법 중 가장 높은 빈도로 사용되는 것이 피동이다. 본 논문에서는 피동의 용법을 직접 피동, 간접 피동, 이중 피동의 세 가지로 분류하여 기존의 방법에서 다루지 못한 문제점들을 해결하였다. 분류에 필요한 결합가 정보에 따라 동사를 설명한 [7]을 표로 정리하여 모든 동사에 대해 가능한 피동문을 선택하여 해당 피동용법에 맞게 처리한다.

a. 직접 피동

調査された結果が発表された.
조사된 결과가 발표되었다.

능동문과 피동문의 격 이동이 1 대 1로 이루어지고, 피동 역어의 대응이 자연스러운 피동으로 기존 연구[1]에서 다루어진 피동 문장의 82%가 여기에 속한다.

b. 간접 피동

昨日は二階で寝がれて眠れなかった.
어젯밤은 이층에서 떠들어서 잘 수 없었다.

일본어에만 존재하는 독특한 표현으로, “迷惑의 피동”이라고 하며[2], 사건과는 관계없는 문장의 주체가 사건으로 인해 심리적, 간접적인 영향을 받았음을 나타내는 피동으로, “~해서 곤란한/나처한” 화자의 심리를 나타낸다. 따라서, 대역어를 직접 피동으로 생성시키면 무자연스러운 번역문이 되므로, れる/られる의 피동의미를 제외시킨 능동으로 처리해서 문맥 속에서 의미가 나타나게 한다. 능동문에서 대격을 취하지 않는 1 항동사와, 2 항 동사 중에서도 피동문에서 주어가 될 수 없는 장소/시간 등의 명사를 대격으로 취하는 동사들 및 공동격을 취해서 양자간의 상호작용을 나타내는 동사와 3 항 동사 중에서 그 동사 자체가 “받는” 동작을 나타내는 もう 등의 수여 동사는 직접피동이 아닌 간접 피동으로 처리한다.

c. 이중 피동

能力が彼に與えられる.
능력이 그에게 주어진다.
彼は神様から能力を與えられた.
그는 신으로부터 능력을 받았다.

여격 목적어와 대격 목적어를 동시에 가지는 수여 동사 중에서 여격 목적어가 주어로 파생되는 경우와, 대격 목적어가 주어로 파생되는 경우가 서로 다른 피동 역어로 나타나는 동사를 말한다. 따라서, 피동 역어를 두 개 준비하여, 피동문의 주어가 사람이거나, 앞에 사물의 목적격을 가지는 경우에는 “~받다/당하다/~게 되다”로, 그 이외의 경우에는 “-이/히/리/기”, “-되다”, “-어지다”로 처리한다.

[사람명사*は/가+ 이중 피동 동사*れる/られる*]

[사물 목적격*を+ 이중 피동 동사*れる/られる*]

<표 2> 테라무라 동사 분류표

격	격 종류	격 관계	예	피동형 종류
1항	가	주격	歩く,泣く,走る	간접
2항	가, を/に/と	동작부가	誘拐する,尊敬する	직접
		대면동작	反對する,會う	간접
		상호작용	戦う	간접
	가, に	출발점	出る,降りる	간접
		통과점	通る,飛ぶ	간접
도착지		降りる,着く	간접	
왕복동사		行く,来る	간접	
	변화동사	變える,發展する	간접	
3항	가, を, に	동작부가+이동	乗せる,残す	간접/직접
		동작부가+변화	變える,決める	간접/직접
		수여1	興える,求める	이중
		수여2	受ける,もらう	간접
		수여3	話す,命じる	이중

3.2 れる/られる의 애매성 해소 알고리즘

위에 기술한 각 용법별 정보들을 이용한 れる/られる 애매성 해소 알고리즘은 다음과 같다.

- * 가능/자발/존경의 처리
- 1. れる/られる 문장이 인식되면 구문분석으로 격과 격의 명사정보를 추출
- 2. 인식된 れる/られる가 1단 동사의 활용형인지 판단
 - 2-1 1단동사 활용형이면
 - 2-1-1 분석된 구문과 가능패턴을 비교, 매칭되면 가능으로 처리
 - 2-1-2 활용동사가 자발동사이면 자발로 처리
 - 2-1-3 분석된 구문과 존경패턴을 비교, 매칭되면 존경으로 처리
 - 2-1-4 나머지는 피동 처리
 - 2-2 1단 동사 활용형이 아니면
 - 2-2-1 활용동사가 자발동사이면 자발로 처리
 - 2-2-2 분석된 구문과 존경패턴을 비교, 매칭되면 존경으로 처리
 - 2-2-3 나머지는 피동 처리

- * 피동 처리
- 3. 동사가 취하는 결합가에 따라서
 - 3-1 결합가가 1 이면 간접피동 처리
 - 3-2 결합가가 2 이고
 - 3-2-1 동사가 변화동사/이동동사/상호동사이면 간접피동 처리
 - 3-2-2 나머지 동사는 직접피동 처리
 - 3-3 결합가가 3 이고
 - 3-3-1 수여동사 중 이중피동이 가능한 동사는 준비된 이중피동 패턴에 매칭시켜 대응되면 이중 피동

처리

3-3-2 수여동사 중 직접피동이 불가능한 동사는 간접

피동 처리

3-3-3 나머지는 직접피동 처리

4. 실험 및 평가

<표 3> 기존 및 제안하는 방법의 번역 성공률 비교

	제안방법	기존방법
가능용례	76%	70%
자발용례	96%	96%
존경용례	84%	0%
간접피동	88%	0%
이중피동	90%	70%
직접피동	96%	96%
평균	88.3%	55.3%

본 논문에서 제안한 방법의 성능 평가를 위하여 EDR 전자 사전에서 용법 당 각 50 문장씩 전체 300 문장을 실험 대상으로 하여, 기존방법과 비교 평가하였다. 평가결과 제안된 방법은 평균 88.3%의 번역 성공률을 보임으로써, 기존방법에 비해 번역률이 33% 향상되었다. 특히, 기존방법에서는 전혀 처리되지 못했던 존경용법과 간접피동용법, 이중피동용법에서 높은 성공률을 나타냈다. 실험에서 실패한 문장들은 보조사의 격 결정 문제 및 복합문에서의 구문분석오류로 인한 실패가 대부분을 차지했다.

5. 결론 및 향후 연구

고품질의 일한 번역시스템 구축을 위해 높은 빈도로 사용되는 일본어 조동사 れる/られる의 애매성 해소방법을 제안하고, 총 300 문장을 대상으로 기존방법과 비교 및 평가하였다. 평가한 결과, 제안한 방법이 기존방법보다 33% 향상된 번역 성공률을 나타냄으로써 れる/られる 애매성 문제에 효과적인 해결 방법이 될 수 있음을 보였다.

참고문헌

- [1] 김정인, 문경희, 이종혁, 이근배, "일한 기계 번역에서의 조동사 "れるられる"의 번역 처리", 96 한글 및 한국어 정보처리 학술대회, pp261-268, 1997
- [2] 안정환, "한일 피동문 대조 연구", 중앙대학교 박사학위논문, 1999
- [3] 차득호, "韓日兩國語の敬語の比較研究", 건국대 교육대학원 석사학위논문. 1982
- [4] <http://www2.plala.or.jp/ja/keig/keimn.htm> 敬意表現
- [5] <http://www2s.biglobe.ne.jp/~k15/> 言葉のmana-講座
- [6] 小泉保 外 4, "日本語基本動詞用法事典",大修館書店, 1989
- [7] 寺村秀夫, "日本語のシンタクスと意味".黒潮出版.1982