

일한 기계번역에서 조동사 “れる, られる” 의 번역처리

김 정인¹, 문 경희², 이 종혁³, 이 근배⁴

포항공과대학교 정보통신연구소¹, 포항공과대학교 전자계산학과²

Translation of Auxiliary Verbs “-reru, -rareru” in Japanese-Korean Machine Translation

Jung-In Kim¹, Kyong-Hi Moon², Jong-Hyeok Lee³, Geunbae Lee⁴

POSTECH Information Research Laboratories¹, Dept. of Computer Science & Engineering, POSTECH²

요약

일본어에서 조동사 “れる, られる” 는 ‘괴동’, ‘가능’, ‘자발’, ‘존경’ 등의 의미로 두루 쓰이고 있다. 일한 번역에서 이들은 여러 가지 대역어로 나타내며 일정한 규칙이나 패턴을 취하지 않으므로, 기계 번역시 조동사 “れる, られる” 는 그 처리가 쉽지 않다. 더구나, 조동사 “れる, られる” 는 일본어에서 높은 빈도로 등장하여 무시하기 어렵고 의미별 분포가 고루 퍼져 있어, 대표적 대역어인 “-아어여 지다, -되다” 등으로만 대응시킬 경우의 번역 에러는 의외로 치명적이다. 따라서, “れる, られる” 의 번역을 고려한 특수 처리를 행할 필요가 있다. 먼저, 본 논문에서는 조동사 “れる, られる” 가 포함된 아사히 신문 기사의 5,800 여 문장을 대상으로 각각의 의미에 대한 분포 및 한국어 대역어의 빈도를 조사하였다. 대역어는 크게 8 종류의 형태로 나누었으며 각 동사별로 “れる, られる” 와 결합된 경우의 의미 출현 빈도를 참고하여 대응 가능한 대역어 형태들을 미리 결정하였다. 그리고, 대역어가 여러 개 존재하는 경우는 패턴 매칭을 통하여 적절한 대역어를 선택할 수 있도록 하였다. 그 결과, 약 87% 의 “れる, られる”가 적절한 대역어로 번역되어, 본 논문에서 제시한 의미 출현 빈도에 기반한 각 동사별 대역어 형태 결정 방법이 “れる, られる”의 다의성 해소에 유효하다고 판단된다.

1. 서론

일한 기계 번역은 다른 언어들 사이의 번역보다는 비교적 양질의 번역어를 얻을 수 있는 것으로 알려져 있다. 이것은 두 언어 사이의 유사성에 기인한 것으로, 대부분의 일한 기계 번역에 관한 연구들은 양 언어의 유사성을 최대한 이용하는 방법을 제시하고 있다[1-4]. 이로 인해 높은 번역률의 실용화된 상용 시스템들이 등장하기 시작했으며, 현재 알려진 번역 시스템들 중에서는 일한 번역 시스템이 가장 좋은 성능을 보이고 있다.

그러나, 유사성을 지나치게 강조한 나머지, 양 언어의 차이

점에 대한 세밀한 분석이 부족하였으며, 따라서 양 언어의 차이에서 나타나는 문제점들의 많은 부분이 무시되어져 왔다. 일한 기계 번역 분야에서 다루기 힘든 대표적인 차이점으로는 각종 단어의 다의성과 슬부의 활용 및 양상표현 등을 들 수 있으며[5], 이들에 대하여 부분적인 연구가 계속 진행되고 있다[5-7].

본 논문에서는 다의성 문제의 일환으로 일본어에서 높은 빈도로 등장하는 조동사 “れる, られる” 가 한국어로 번역될 경우에 생기는 문제점들을 확실히 규정하고, 또한, 적절한 대역어가 생성되게 하는 방법을 제시해 보기로 한다.

2.“れる,られる”의 대역어별 특징

일본어 조동사 “れる,られる”는 ‘피동’, ‘가능’, ‘자발’, ‘존경’의 4 가지 의미를 가지고 있다. 특히, ‘피동’의 경우는 문장의 유형에 따라 한국어에서 “지다, 받다/당하다, 되다” 등의 여러 가지 대역어의 형태로 나타난다.

“れる,られる”의 한국어 대역어들을 종류별로 나열한 후, 대역어별로 어느 정도의 분포로 나타나는지를 조사하여 표 1을 만들었다. 표 1의 수치는 아사히 신문의 6개월분(97.1-97.6) 기사에서 정치, 경제, 국제, 스포츠 4개면을 대상으로 조사한 것이다. 전체 문장의 개수는 26,944 개였으며, 그 중 5,797 문장이 “れる,られる”를 포함하고 있어, 문장 단위에서 약 22%의 높은 출현 빈도를 보였다. 또한, 5,797 문장에서 6,216 개의 “れる,られる”가 등장하여, 문장 당 1.07 개의 분포를 보였다.

표 1. “れる,られる”의 한국어 대역어 종류 및 빈도수.

의미	기호	한국어 대역어	빈도수	%
피동	P1	지다(피동)	1176	18.9%
	P2	받다,당하다	291	4.6%
	P3	이/히/리/기	817	13.1%
	P4	되다	2871	46.2%
가능	P5	리/을 수 있	206	3.5%
자발	P6	지다(자발)	21	0.3%
존경	P7	(으)시	5	0.1%
기타	P8	어간 바뀜 등	829	13.3%
합계			6216	100.0%

“습격 받다/습격 당하다”, “보인다/보여진다” 등과 같은 어느 쪽으로도 대역이 가능한 한국어 표현들은 문장에 따라 부드러운 표현이 되도록 주관적으로 분류하였다. 그리고, 대역어가 규칙을 따르지 않고 완전히 다르게 나타나는 경우, 전접하는 단어의 격조사가 바뀌어 나타나는 경우, ‘피동’, ‘가능’, ‘자발’, ‘존경’의 의미가 소멸되는 경우, 또는 이들 현상이 섞여서 나타나는 경우 등을 ‘기타’로 분류했다.

신문 기사를 대상으로 조사한 결과, ‘존경’에 대한 빈도는 상당

히 낮게 나타났으며, 이는 일반 문헌이나 신문, 잡지 등에서는 “れる,られる”가 ‘존경’의 의미로 거의 사용되고 있지 않음을 보여주고 있다. 대역어별로 대표적인 유형의 문장들을 살펴보면 다음과 같다.

<아이어여 지다(피동)>

貿易外收支などを合わせた經常収支の赤字も過去最悪になるとみられている。

무역 외 수지 등을 합친 경상 수지의 적지도 과거 최악이 될 것으로 보여지고 있다.

大統領選と同時に行われる共和国議會選挙は、定数63に對し、約1000人が立候補した。

대통령 선거와 동시에 행하여지는 공화국 의회 선거는, 정수 63에 대하여, 약 1000 명이 입후보하였다.

한국어 피동의 대표적인 대역어이며, 특징이 없는 피동문은 대부분 여기에 해당된다.

<받다,당하다>

再建のチャンスを与えられる企業は少ない。

재건의 찬스를 부여 받는 기업은 적다.

人質が政府側の速やかな對應を求めて手紙を書いたともみられるが、ゲリラ側に強要された可能性もある。

인질이 정부측의 빠른 대응을 구하여 편지를 적었다고도 보여지지만, 게릴라 측에 강요 당한 가능성도 있다.

일부의 동사들은 피동이 될 경우, “받다, 당하다”로 번역이 이루어져야 한다. 이익을 보는 피동으로 “받다”를, 그리고 피해를 입는 의미로 “당하다”를 사용하는 경향이 있으나, 구분하기 어려운 경우도 있다.

<이,히,리,기>

總理府の準備室には關係省廳から三十人が集められた。

총리부의 준비실에는 관계 관청으로부터 삼십인이 모였다.

名譽回復の一環として、日本航空協會の會長に22日選ばれるばかりだった。

명예 회복의 일환으로서, 일본 항공 협회의 회장에 22일 막 뽑혔다. 피동 접미사를 사용하여 피동문을 생성한다. 동사의 음운에 따라 “이,히,리,기”가 존재하며, 자연스러운 피동문을 생성하기 위해서는 꼭 필요한 처리이다.

<되다>

早ければ、1998年4月から施行される。

빠르면, 1998년 4월부터 시행된다.

爲券は、期間 60 日の政府短期証券なので日銀引き受けが認められている。

교환권은, 기간 60 일의 정부 단기 증권이기 때문에 일본 은행 인수가 인정되고 있다.

する동사의 피동형인 “される”의 대표적인 대역어이며, “れる, られる” 대역어 전체에서 약 46%의 가장 높은 출현 빈도를 보였다.

<리을 수 있>

都會にいと、あらゆる文化、藝術の恩恵を受けられるが、地方はそうはいかない。

도회에 있으면, 모든 문화, 예술의 은혜를 받을 수 있지만, 지방은 그렇지 않다.

野黨第 2 黨の自由民主黨は“わが黨の排除は許されない”と訴えている。 야당 제 2 당의 자유 민주당은 “우리 당의 배제는 용서할 수 없다” 라고 고소하고 있다.

‘가능’은 ‘피동’ 등의 다른 의미로 해석될 경우, 문외의 의미가 잘못 전달될 위험이 커서, 신중한 처리를 할 필요가 있다. 그러나, 실제 코퍼스에서는 약 3.5%의 출현 빈도를 보여, ‘가능’의 의미를 위하여 과도한 처리를 행할 경우, 오히려 ‘피동’이 ‘가능’으로 바뀌어 버리는 오류에 의해 대역어가 적절히 선택되지 못할 위험이 있다.

<지다(자발)>

緊急に必要と思われる 10 万トンの穀物など、全部で 20 万トン余りの食糧を送る計画だ。

긴급하게 필요하다고 생각하여지는 10 만톤의 곡물등, 전부 20 만톤여유의 식량을 보낼 계획이다.

永友主將は“今日は勝とうという意識が感じられなかった”とふ然とした表情だった。

나가또모 주장은 “오늘은 이기려고 하는 의식이 느껴지지 않았다” 라고 무연한 표정이었다.

‘자발’의 대역어 “지다”는 일반적인 피동의 대역어 “지다”와 형태가 같으므로 굳이 구분할 필요는 없다. 그러나, 감각동사류에 주로 나타나는 특징으로 인해 구분해서 처리할 수도 있다.

<(의시)>

天皇、皇后兩陛下は 10 日午後、アルゼンチンの首都ブエノスアイレス市内のパレルモ日本庭園で開かれた日系人らの歓迎式典に出席された。

천황, 황후 양 폐하는 10 일 오후, 아르헨티나의 수도 부에노스 아이레스 시내의 바레르모 일본 정원에서 열린 일계인들의 환영 식전에 출석하였다.

官房長官は 9 日の記者会見で、首相、蔵相経験者の参加について“財政に一番大きな関心を拂ってこられた方だし、世論形成にも大きな影響がある。”と説明した。

관방장관은 9 일의 기자회견에서, 수상, 장상 경험자의 참가에 대하여 “재정에 가장 큰 관심을 나타내어 오신 분이요, 세론 형성에도 큰 영향이 있다.” 라고 설명하였다.

‘존경’은 “れる, られる”의 4 가지 의미중의 한가지이지만, 일반적인 신문이나 잡지, 과학 기술 문헌등에는 상당히 낮은 비율로 나타났다. 본 실험에서도 5 건이 나타났지만, 이는 0.1%에 불과했으며, 따라서 전체의 번역률을 올리는 데에 ‘존경’의 의미를 취하기 위한 노력은 별 의미가 없다고 보여진다.

<기타>

東京とはかなり差があるが 2 位の規模とされる。

동경과는 상당히 차가 있지만, 2 위 규모로 여겨진다.

小澤氏は、オレンジ事件について“国民から信頼を問われる事態になり、誠に遺憾だ”と...

오자와씨는, 오렌지 사건에 대하여 “국민으로부터 신뢰를 의심받는 사건이 되어, 참으로 유감이다” 라고..

その間、國王の子供や孫十數人も處殺の犠牲になったといわれる。

그 동안, 국왕의 아들과 손자 십수인도 처형의 희생이 되었다고 한다.

國會審議では“検査、監督權限の分離は徹底されているか” “金融システムに不安が生じる際の體制は十分か”などの観点から質疑が行われそう하다.

국회 심의에서는 “검사, 감독 권한의 분리는 철저하게 되고 있는가”, “금융 시스템에 불안이 생기는 동안의 체제는 충분인가” 등의 관점으로 부터 질의가 행하여질 것 같다.

“れる, られる”와 결합한 동사가 어간이 바뀌는 경우로 ‘기

타'중 약 65% 를 차지하였다. 그 중에서는 “とされる”, “問われる”, “いわれる”등이 높은 비율을 차지하였다.

家電業界特有の價格下落競争に巻き込まれる恐れもある。

기전 업계 특수의 가격 하락 경쟁에 말려 들어갈 위험도 있다.

“れる, られる”의 의미가 무시되는 경우이다. ‘기타’중에서 약 25%를 차지하였다.

國民の厳しい視線にさらされることになりそうだ。

국민의 엄한 시선을 맞는 일이 될 것 같다.

省廳側には“のんびり構えていては優秀な學生に逃がられてしまう”という危機感が強まっていた。

관청측에는 “유유히 준비하고 있어서는 우수한 학생이 도망쳐 버린”이라는 위기감이 강해지고 있다.

“れる, られる”에는 ‘迷惑受身’ 라는 한국어에는 없는 표현이 존재하는데, 이는 양 언어의 피동 표현의 큰 차이이지만, ‘기타’의 2% 정도로 출현 빈도가 그다지 높지 않았다. 전집 조사와 해당 동사의 대역어를 적절히 조절하면 같은 의미의 한국어 문을 생성할 수 있으나, 기계에 의한 처리는 쉽지 않다.

전체적으로 ‘기타’는 코퍼스에서의 출현 비율이 13.3% 로 비교적 높게 나타났으며, 이들의 대역어 선택을 위한 적절한 처리가 필요하다고 하겠다.

3. “れる, られる”의 다의성 해소 방법

본 논문에서 조사한 대역어들의 출현 빈도를 살펴보면, 일반적으로 일본어 조동사 “れる, られる” 는 ‘피동’으로 약 80% 이상이 사용된 것을 알 수 있다. 따라서 보통의 번역 시스템들은 ‘피동’으로만 해석하여 기본적인 번역률에 만족하는 듯 하다. 그러나 표1에서 보았듯이 ‘피동’ 중에서도 대역어가 여러 형태로 다르게 나타나고, ‘피동’이외의 경우가 20% 정도의 빈도로 나타나, 번역의 질을 높이기 위해서는 대역어별로 최대한 구분이 가능하도록, 번역 처리를 행할 필요가 있다.

3.1 선행 연구

“れる, られる”의 다의성에 의한 번역 실패는 생성되는 대역문이 의미 불명이 되는 경우까지 나타나므로 그에 대한 연구는

많이 행하여졌다고 볼 수 있지만, 의외로 “れる, られる”의 다의성 해소 방법을 소개한 문헌은 거의 없는 편이다. 또한, 이 문제로 인해, 일반 번역 시스템들의 번역률도 다소 떨어지고 있는 것이 현실이다. 최근, 권[8]은 “れる, られる”의 다의성을 해소하기 위하여 3 가지 접근 단계를 제안하였다.

1 단계) 격형식 정보를 이용하여 ‘피동’, ‘가능’, ‘자발’을 한다.

<피동의 격형식 정보>

$N_1[が/は]+N_2[[に/から/によって]+V+“れる, られる”$

$V+“れる, られる”+てしまう$

<가능의 격형식 정보>

$N_1[が/は]+V(의지동사)+“れる, られる”+た/てしまう]$

$N_1[と/ならば]+V(의지동사)+“れる, られる”+た/てしまう]$

- 여기서 5단 동사는 대부분 가능형 동사를 가지고 있으므로 확률상 5단 동사를 제외시켰다. 이중 취소선은 접속이 불가능한 것을 나타낸다.

<자발의 격형식 정보>

$N_1[と]+N_2[が]+V(사고동사)+“れる, られる”$

2 단계) 존경을 나타내는 의미소성을 이용한다.

$N_1[が/は]+(N_2[[に]+N_3[を]])+V+“れる, られる”$

일반적으로 “れる, られる” 가 존경의 의미를 가지기 위해서는 N_1, N_2, N_3 의 의미소성이 존경의 의미를 가져야 하며, “선생, 대통령, 천황폐하,...” 등과 같은 명사들이 이에 해당한다. 그리고 존경의 접두사, 접미사가 붙은 명사들도 존경의 의미를 가지는 명사로 분류했다. (お/御) + N, N + (様/展)

3 단계) 입력문과 예문을 비교, 계산한다.

격형식 정보와 존경의 의미소성을 가진 단어들에 의해 “れる, られる” 의 의미가 정확히 구별이 되지 않을 경우를 위하여, “れる, られる” 가 포함된 예문을 준비해 두고, 예문과 입력문을 비교하여 의미를 구별하도록 하고 있다.

이 3 단계의 접근 방법은 각각 다음과 같은 문제점을 가지고 있다. 첫 단계에는 ‘피동’, ‘가능’, ‘자발’을 구별하는데 동사의 유형이 필요하다는 것이다. 즉, 의지 동사, 사고 동사, 가능 동사 등을 따로 분류해야 하는 등, 격 형식보다는 동사 자체의 의미를 우선 시키고 있다. 또한, 동사들의 분류 방법에도 문제가 남는다. 예를 들면 “考える”는 사고 동사이지만, “考えられる”가 될 경우,

대부분 ‘자발’ 보다는 ‘가능’의 의미인 ‘생각할 수 있다’ 로 번역 된다. 그리고 ‘피동’ 중에서의 여러 대역어를 구별시키는데 이 방법은 상당히 미흡하다. **두 번째 단계**는 ‘존경’의 의미소성을 지닌 명사가 나타나면 해당 동사를 ‘존경’의 의미로 바꾸려는 시도인데 이것도 정확하게 ‘존경’의 의미를 추출하기 어렵다. ‘존경’의 의미로 사용되는 경우가 극히 적은 데에도 ‘존경’의 의미소성이 나타나면 무조건 ‘존경’으로 의미를 고정시키는 것은 좋지 않다.

先生にしかられた。

선생님에게 아단치셨다.(X)

선생님에게 아단맞았다.(O)

위의 예에서 “아단치다” 의 피동은 어간이 바뀌는 “아단맞다” 로 처리되어야 하나, ‘존경’의 의미로 사용되어서 오히려 대역어가 어색해진 경우이다. **세 번째 단계**로서, 입력문과 예문을 비교하여, 높은 유사도를 가지는 의미로 해석하는 것은 정확한 의미를 추출해 내는 데에 어느 정도의 효과는 예상되지만, 이를 위하여 많은 예문을 준비해야 하며, 또한 번역에 걸리는 시간을 늦추는 단점을 가져, 실용 시스템에 적용하기가 힘들다.

위의 3 단계 방법은 약 100 문장을 대상으로 실험한 결과 85% 정도의 성공률을 보였다고 보고하고 있지만, 이는 만족할 만한 수치가 아니며, 또한 실험의 규모가 너무 작아서 결과의 신뢰성도 떨어진다고 하겠다. 따라서, “れる, られる” 의 의미별 분류가 적절히 행해지고, 또한 대역어를 가장 잘 선택할 수 있는 실용적인 새로운 방법이 필요하다.

3.2 “れる, られる” 의 다의성 해소를 위한 제안

“れる, られる” 는 특별한 규칙이 없이 여러 의미로 사용이 되므로 의지동사, 감정동사, 사고동사 등으로 해당 동사를 분류하여 처리하기에는 예외 단어가 다수 존재하여 효과를 거두기 어렵다. 따라서, 우리는 코퍼스로부터 각각의 동사별로 “れる, られる” 와 결합된 경우의 한국어 대역어 출현 빈도를 참고하여, 각 동사별로 많이 사용되는 의미를 우선시키는 방법을 제안한다. 특히 ‘피동’의 경우는 대역어의 형태가 다르게 나타나는데, 이들의 구분은 단어별로 따로 정의할 필요가 있으며, 동사에 따라 특별히 자주 나타나는 의미를 추출하여 대표적인 의미로 취하고, 거기에 맞는 대역어를 미리 준비해 두어 다의성을 해소해 나간다. 즉, 일본어 동사를 표1의 8 가지 형태로 분류하고, 출현 빈도를 참고

하여 거기에 맞게 미리 대역어를 준비해 두는 것이다.

3.2.1 다의성 해소의 순서

본 방법은 코퍼스로부터 미리 모든 일본어 동사에 “れる, られる” 가 결합할 경우의 한국어 대역어 빈도를 조사하여 사전에 기술하여야 하지만, 작업의 너무 방대하여 실제의 시스템에 도입 시키기가 어렵다. 따라서 대역어 선택에 잦은 오류가 일어나는 동사들을 중심으로 “れる, られる” 결합형의 다의성 해소를 행하기로 한다. 하나의 대역어 형태로 수렴할 경우는 1대1 대응에 의한 대역어 생성이 가능하다. 두개 이상의 대역어가 존재하는 경우는 “れる, られる” 결합형 동사를 대상으로 다음과 같은 처리를 행하여, 적절한 대역어가 선택되도록 할 수 있다.

해당 동사로부터 “れる, られる”의 대역어 선정을 위한 패턴을 추출하여, 준비된 패턴과의 패턴 매칭에 의해 적절한 대역어로 대응시킨다. **그림 1** 은 다의성 해소의 순서를 나타낸 것이다.

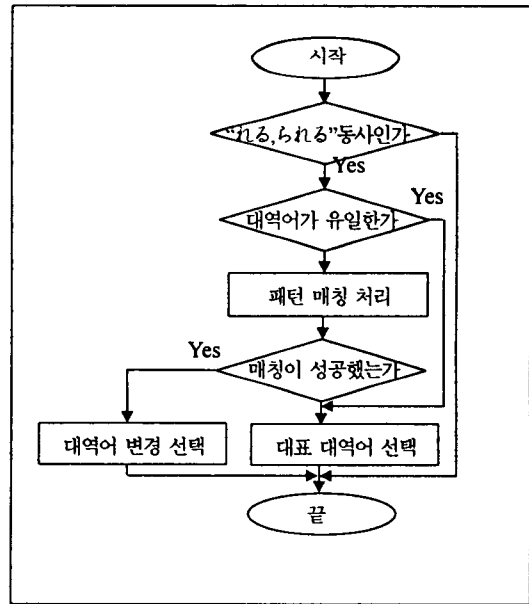


그림 1. “れる, られる”의 처리 흐름도

“れる, られる” 는 4 가지 의미로 사용되지만, ‘피동’의 한 대역어와 ‘자발’이 한국어로 해석될 경우, 같은 “지다” 로 대응이 가능하며, ‘존경’은 그 출현 빈도가 0.1% 이하로 미미하므로, ‘가능’의 의미

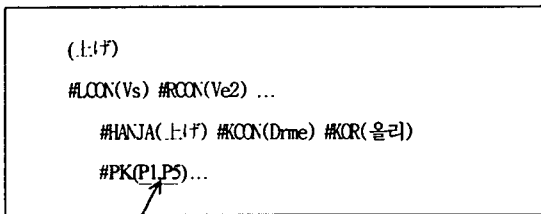
를 구별해 내면 나머지는 ‘피동’으로 처리하여도 큰 무리가 없다.

같은 의미 속에서의 대역어의 차이는 다시 상세 분류에 의해 구별해 준다. 즉, P21 은 “받다”, P22 는 “당하다”, P31 은 “이”, P32 는 “히” 등으로 나누었다. 특히, ‘기타’는 P81 이 어간이 바뀌는 경우, P82 가 “れる, られる”의 문법적인 의미가 사라지는 경우, P83 은 전접 조사의 격이 바뀌는 경우, P84 는 나머지 모든 경우로 분류하였다.

여러 개의 대역어는 주로 “되다”와 “받다/당하다”(피동기리), “지다”와 “르 수 있”(피동과 가능), “(으)사”와 “르 수 있”(존경과 가능) 등에서 나타나며, 이들을 구별하기 위한 패턴을 준비하여 이 문제를 해소한다.

3.2.2 “れる, られる”처리를 위한 동사 사전의 구조

동사 사전에는 “れる, られる”의 다의성 해소를 위하여 다음과 같이 대응 가능한 모든 대역어 형태를 준비한다.



대표적인 대역어를 ‘피동’(P1)으로 기술하고, ‘가능’(P5)의 경우도 동시에 정의해 둔다.

복수의 대역어 기호 중에서 두 번째 기호부터 패턴 매칭을 행하여, 준비된 패턴과 일치할 경우는 그 기호에 맞는 대역어로 바꾸어 한국어를 생성시키고, 그렇지 않을 경우는 대표적인 대역어(첫번째)로 한국어를 생성한다.

3.2.3 다의성 해소를 위한 패턴의 분류

모든 일본어 동사에 “れる, られる” 형태를 대비하여 8 가지로 나누어진 대역어 형태 중에서 하나를 등록해 둔다. 어느 대역어 형태를 기본으로 할 것인가는 코퍼스에 등장한 출현 빈도에 의존하며, 두 가지 이상의 대역어가 고르게 나타날 경우는 빈도수가 높은 한 대역어 형태를 기본으로 고정하고 다른 종류의 대역어 형태는 문장의 패턴을 참고하여 선택이 가능하도록 처리해 주도록 한다. 본 실험에서는 90% 이상의 출현 빈도로 한가지 대

역어 형태에 수렴하는 경우는 유일 대역어 형태로 지정하고, 그 이하 빈도의 경우는 대역어 형태를 선택할 수 있도록 한다. 대역어 형태의 선택은 문장으로부터 패턴을 추출하여 미리 준비된 패턴과의 비교를 통하여 알맞은 대역어 형태를 결정할 수 있도록 한다. “れる, られる”의 다의성 해소를 위해 준비할 패턴은 3 종류이며 다음과 같다.

<~할 수 있다> 류 : P5 패턴

[<が/は/を/も> * <동사> * れる, られる * <부정 조동사(ない/ず/ん/ぬ)ように>]

[<が/は/を/も> * <동사> * れる, られる * !(た/てしまう)]

(기호설명 : * : 인접, + : 접속, ! : 접속불가, / : 혹은, | : 혹은)

P5는 ‘가능’의 의미를 구별해 내기 위한 패턴이다. “れる, られる”의 뒤쪽에 부정 조동사나 “ように”와 결합할 경우는 ‘가능’의 의미로 해석한다. 단, “た/てしまう”와 결합할 경우는 ‘가능’의 의미를 취하지 않는 것을 의미한다. “答える, 避ける, 抑える, 得る, 生き残る, 耐える, いる” 등은 90%를 넘는 가능형 단일 대역어로 대응 가능한 단어들이며, “考える, 出る, 上げる” 등이 “れる, られる” 조동사가 같이 쓰인 경우, 여러 개의 대역어로 대응될 수 있는 것들이다. 이와 같은 경우, 미리 준비한 패턴과의 매칭을 통하여 적절한 의미를 취할 수 있다. 다음은 패턴 매칭에 의해 대응된 예이다.

競争上, 値段は上げられない

경쟁상, 가격은 올릴 수 없다.

PK#(P1,P5)-> 부정 조동사가 결합되었으므로 P5 패턴에 해당하며 ‘가능’의 대역어로 처리된다.

<되다> 류 : P4 패턴

[! を * <동사> * れる, られる]

“되다” 대역어는 조사 “을”과 같이 사용될 수 없다. 따라서 전접하는 조사가 “を”인 경우는 “を”를 “이/가”의 형태로 바꾸던지 “れる, られる”를 “받다/당하다”로 바꾸는 두 가지의 선택이 가능하다. 즉, P4 패턴으로 “되다”와 “받다, 당하다”의 구별을 어느 정도 해결할 수 있다.

育児, 介護など家庭責任のある労働者には深夜労働を免除される制度を設ける.

육아, 간호 등 가정 책임이 있는 노동자에게는 심야 노동을 면

제 받는 제도를 세운다.

<(으)사> 류 : P7 패턴

[<명사(존경의미)> * 가 + <동사> * れる, られる]

[<동사> * れる, られる(연체형) * <명사(존경의미)>]

‘존경’은 ‘존경’으로 사용되었다는 확신이 필요하다. 동사의 주어
가 정확히 파악되고 주어가 존경의 의미소성을 가지고 있을 경우
만 ‘존경’으로 분류한다. 본 실험에서는 존경의 의미소성이 적용
된 명사 사전을 준비하지 못한 이유로 ‘존경’ 처리 패턴은 적용
되지 않았다.

4. 실험 및 검토

본 논문에서 제안한 “れる, られる”의 다의성 해소 방법이
어느 정도 효과를 보이는지를 확인하기 위하여, 실제로 표1의
코퍼스를 대상으로 번역 작업을 행하였다. 번역은 현재 약 13만
단어로 구성된 COBALT-J/K 를 이용하였으며, 실험 결과의 평가
는 일본어를 전공한 3명에 의해 이루어졌다. 자연스러운 대역어
에는 ‘O’표를, 부자연스러운 대역어에는 ‘X’표를 주어, 각 대역어
패턴별로 헤아렸다.

4.1 실험 결과

6,216 개의 “れる, られる”에서 자연스럽게 번역된 것은 5,417
개였으며, 이는 약 87%가 적절한 대역어를 찾는데 성공한 것을
의미한다. 각 대역어별 선택 결과는 그림2 과 같다.

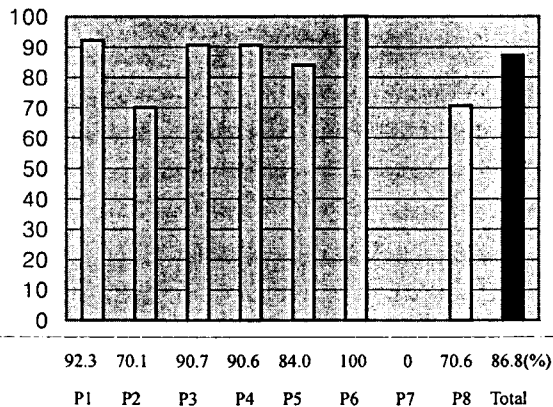


그림2. “れる, られる”의 대역어별 번역률

P1,P3,P4,P6 (지다(피동), 이/히/리/기, 되다, 지다(자발))가 대
역어 선택에서 비교적 높은 성공률을 보였으며, P2,P5, P7,P8 (받다/
당하다, 르 수 있,(으)시) 등은 상대적으로 낮은 성공률을 보였다.

4.2 검토

본 논문에서 제시한 “れる, られる”의 다의성 해소 방법은 적
절한 대역어 선정에 어느 정도 효과가 있었다. 그러나, 본 방법에는
두 가지의 결정하기 어려운 부분이 남아 있다. 첫번째는 일본
어 동사의 “れる, られる”형태에 대하여 유일한 대역어를 가진다
고 정의할 수 있는 동사들의 결정 문제이다. 본 실험에서는 90%
를 사용했지만, 이는 검증을 거치지 않은 수치이며, 이에 대한 충
분한 검증을 필요로 한다. 두 번째는 ‘가능’, ‘존경’ 등을 구분해
내기 위하여 패턴을 정의하는 경우, 어느 정도로 세밀하게 패턴
을 정의하느냐에 대한 문제이다. 즉, 재현율과 정확도의 배치 문
제로서, ‘가능’이나 ‘존경’은 코퍼스에서 출현 횟수가 상대적으로
낮았으므로 정확도를 최대한 높이는 패턴의 구성이 적절하리라
생각된다. 그리고, 본 방법은 이미 만들어진 일한 번역 시스템
COBALT-J/K 에 “れる, られる”의 번역시 일어나는 오번역을 줄이
기 위하여 제안한 것이다. 즉, 번역 시스템의 튜닝 과정에서 생
기는 “れる, られる”의 다의성은 일반적인 규칙을 적용시키기가 어
려워서 코퍼스로부터 통계적인 정보를 얻어 가장 가능성이 큰 대
역어를 동사 사전에 직접 등록하는 방법을 제시했다. 대용량 코
퍼스로부터 사전의 모든 동사들을 대상으로 이들 정보를 얻는 것
은 간단한 일이 아니다. 수작업에 의한 코퍼스 센스 태깅 작업이
선행되어야 하며, 코퍼스의 용량이 커질수록 신뢰성은 증가하는
반면에 작업량이 늘어나서 실용 시스템에 적용시키기는 더욱 더
어렵다. 그러나, 문제가 있는 동사들을 대상으로 코퍼스의 통계적
인 정보를 점차적으로 적용하는 것은 충분히 가능하며, “れる, ら
れる”와 결합한 동사의 다의성 해소에 좋은 개선책으로 생각할
수 있다. 실험에서 사용한 코퍼스는 5,800 여 문장으로 결코 큰
용량은 아니지만, “れる, られる”의 다의성에 코퍼스를 사용할 경
우 현재의 번역률보다 좋은 결과를 기대할 수 있다는 가능성을
보였다.

5. 결론

일본어 조동사 “れる, られる”는 한국어로 번역될 경우, 그

대역어가 여러 가지 형태로 존재한다. 본 논문에서는 대역어를 8 가지 형태로 나누고, “れる, られる”를 취하는 일본어의 각 동사를 8 가지 형태 중에서 한가지로 고정시켜 디폴트로 번역시키는 방법을 제안하였다. 물론, 번역 대상 문장에서 복수의 형태로 번역이 이루어져야 할 경우는 패턴에 의해 형태를 바꿀 수 있도록 하였다.

실험의 결과, 약 87%의 대역어 선택 성공이 이루어져, 본 논문에서 제안한 “れる, られる”의 번역 방법이 효과가 있음을 알았다. 본 방법은 단일 대역어에 대한 선정 기준과 여러 개의 대역어에 대해 하나의 대역어를 결정하는 패턴을 어떻게 구성하느냐에 따라, 대역어 선택의 성공률이 조정되는데, 너무 간단한 패턴을 이용하게 되면, 일반적인 대역어로 번역되어야 할 “れる, られる”의 경우도 패턴에 걸려, 다른 대역어로 바뀌어 버리는 오류를 범할 수 있다. 나머지 13%의 오번역도 현재의 직접 번역 방식을 채용하는 일한 번역에서는 해결하기 힘든 과제로 남아 있으나, 우선 “반다당하다, ㄹ 수 있, (으)사” 등의 70% 대 수준의 대역어 선택 방법을 점차 개선하고, 보다 큰 코퍼스를 이용하여 효율적인 “れる, られる”의 번역을 위하여 계속 노력해 갈 예정이다.

참고 문헌

- [1]H.D.Lee, M.Nakajima, T.Agui: A Japanese-Korean Predicate Translation System Using the Semantic Correspondence between Their Auxiliary Predicative Expressions, Trans. of IPSJ,31(6), 801-809(1990),(in Japanese)
- [2]J.I.Kim, S.Okoma : A Processing of Multi-Translatable Verbs in Japanese-Korean Machine Translation, the 45th annual conference of IPSJ, (3), 97-98(1991),(in Japanese)
- [3]T.S.Kim, U.Shoji: Processing of Negative Sentences in Japanese-Korean Machine Translation, Trans. of IPSJ, 34(5), 892-904 (1993) ,(in Japanese)
- [4]D.Hwang, M.Nagao, S.Sato : Construction of a thesaurus for Korean from a thesaurus for Japanese, SIG-NLP of IPSJ, 94-11,79-84(1993) ,(in Japanese)
- [5]김 정인, 문 경희, 이 종혁, 이 근배 : 일-한 기계번역에 있어서 한국어 술부의 생성과 평가, '96 한글 및 한국어 정보처리 학술대회,329-337(1997)
- [6]J.I.Kim, S.Okoma : A Method of Generating Korean Language by Using Extended Translation Tables on Japanese-Korean Machine Translation, Trans. of IPSJ, 37(9),1697-1707 (1996) ,(in Japanese)
- [7]C.J.Park, J.H.Lee, G.Lee, K.Kakehi: Collocation-based Transfer Method in Japanese-Korean Machine Translation, Trans. of IPSJ, 38(4),707-718(1997) ,(in Japanese)
- [8]T.K.Kwon, S.Okoma : A Study on Solution of Ambiguity of Auxiliary Verbs in Japanese-Korean Machine Translation Systems, Master Thesis of Keio University, Dept. of Administration Engineering (1997) ,(in Japanese)