

한일 양국어 연결어미의 비교연구

- 일한 번역기에서 ~ㄷ의 애매성 해소를 통하여 -

허남원†, 정유진, 문경희, 이종혁‡
계명문화대학 전산정보처리과†, 포항공과대학 전자계산학과‡

A Study on the Korean and Japanese Connective Endings

Namwon Heo†, Yujin Chung, Kyonghee Moon, Jong-Hyeok Lee‡
Dept of Computer Information Science, Keimyung College† Dept of Computer Science, POSTECH‡
{nwheo,prizer,khmoon,jhlee}@madonna.postech.ac.kr

요약

본 논문에서는 한일 양국어의 연결어미에 관한 문법적 연구결과를 이용하여 효율적인 연결어미 번역방법을 제안한다. 특히, 일본어 연결어미 가운데서 가장 빈번히 사용되고 있으나 한국어로 기계번역할 경우 심각한 중의성을 보여온 「-ㄷ」를 이에 상응하는 대표적 한국어 연결어미 「-아/-고」와 문법적 차원에서 비교·분석하여 자연스럽게 번역하는 방법을 제안한다. 기존의 일한기계번역 시스템에서는 일본어 연결어미 「-ㄷ」¹의 한국어 번역 정확률이 약 54% 정도에 불과하였으나 제안하는 번역방법을 적용하면 73% 이상으로 개선됨을 보인다.

1 서론

교착어의 특성을 가지는 한국어와 일본어는 연결어미가 매우 발달한 언어들로서 양국어에서 연결어미가 차지하는 비중은 막중하다고 할 수 있다. 양국어에서의 연결어미는 그 사용빈도가 높을 뿐만 아니라 어떻게 해석하느냐에 따라 문장 전체의 의미가 달라지게 된다. 따라서, 한국과 일본의 문법연구 학자들은 각각의 언어에서의 연결어미는 많이 다루어 왔지만 양국의 연결어미를 함께 연구한 사례는 거의 없었다. 그러나 한일 양국어간의

원활한 기계번역을 위해서는 이들 연결어미들간의 비교분석이 반드시 선행되어야 한다. 본 논문에서는 기존의 문법연구에서 시도된 접근방법을 이용하여 한국어와 일본어의 연결어미를 비교·분석함으로써 양국어간의 자연스러운 연결어미 번역방법을 제시하고자 한다.

먼저, 일본 중·고등학교 국어교과서에 나타나는 연결어미들을 분석해 본 결과 출현빈도가 높은 연결어미들 중 대부분은 오직 하나의 고유 의미만을 갖고 있기 때문에 한국어로 기계번역을 하는 경우에 있어서도 그다지 문제가 되지 않는다. 그러나, 일본어 연결어미 「-ㄷ」는 한국어로 번역할 경우 일반적으로 종속의 연결어미 「-아」로 번역되거나 또는 대등의 연결어미 「-고」의 두 가지로 번역된다. 그런데 연결어미 「-아」와 「-고」가 갖고 있는 의미기능 사이에는 공통요소가 회박하기 때문에 이들이 서로 바뀌어 번역될 경우 원래의 문장이 의도했던 의미와는 전혀 별개의 의미로 변질되어 버린다. 더구나, 이렇게 심각한 중의성을 발생시키는 「-ㄷ」가 [표-1]에서 보는 바와 같이 일본어 연결어미 중에서 70% 이상의 출현빈도를 나타내고 있으므로 효율적인 일한기계번역을 구현하기 위해서 이는 반드시 해결되어야 하는 문제이다.

본 논문에서는 일본어의 대표적 연결어미 「-ㄷ」를 이에 대응하는 한국어 연결어미 「-아/-

¹ 실제 일본어에서는 연결어미가 아니라 접속조사라고 부른다.

고」와 언어학적인 차원에서 비교 분석하여 그 결과를 일한기계번역에 적용함으로써 「-て」의 중의성 해결을 위한 새로운 연구방향을 제시하고자 한다.

[표-1] 일본 중고등 국어교과서에 나타난 연결어미들의 빈도수

접속조사	빈도수	대응 한국어 연결어미
て / で	7208	①아(서) ②-고
ても / でも	519	-더라도
と	463	①-하자마자 ②-면
が	405	-지만
ば	355	-면
たり / だり	221	A도 하고 B도 하고
ながら	214	-면서
たら(ば)	145	-면
ので	141	-때문에
から	125	-때문에
し	99	-고
ために	76	-하기 위해
なら	76	-면
けれど(も)	65	-지만
たって / だって	48	-더라도
のに	43	①-는데도 ②-르텐데
기타	76	
총합	10275	

2 「-て」의 번역방법

2.1 기존의 접근방법

지금까지 상용화된 몇몇 일한기계번역기에서는 일본어 연결어미 「-て」의 번역방법을 위하여 한국어 연결어미 「-아」나 「-고」 가운데 어느 한가지로만 번역하도록 고정시켜 놓거나 또는 「-て」 다음에 쉼표가 있으면 「-고」, 그렇지 않으면 「-아」로 번역하는 방법 등을 사용해 왔다. 이러한 방법들은 연결어미 「-아」와 「-고」가 한국어 연결어미들 중 가장 대표성이 높고 광범위하게 사용되는 특성으로 인해 해석정확도가 50% 가까이 이를 뿐만 아니라 기계번역 시스템에서의 구현도 간단하다는 장점이 있었다.

그러나, 해석정확률이 50%를 상회한다고 하더라도 잘못 번역된 연결어미들은 문장 자체를

비문으로 변화시켜 버림으로써 문장의 의미과약을 불가능하게 만들기 때문에 기계번역 시스템의 전체적인 해석 품질을 저하시키는 요인이 되어왔다는 점을 감안하면, 「-て」의 해석정확률을 현재의 수준 이상으로 향상시키는 작업은 일한기계번역 분야에서 필수적으로 고려되어야 할 연구과제라 하겠다.

2.2 본 논문의 접근방법

기계번역을 연구하는 사람들과는 달리, 한일 양국의 국어학자들 사이에서는 연결어미의 의미를 확인하는 연구가 언어학적 측면에서 활발하게 수행되어 왔다. 그 중에서도 대표적인 3가지 접근방법을 응용하여 일한기계번역에서 연결어미 「-て」의 중의성 해결을 시도해 보고자 한다.

1) 형태론적 접근

연결어미와 함께 나타나는 특정 단어나 문형의 출현여부를 파악한다. 즉, 연결어미 「-て」의 의미를 알아내거나 대응하는 한국어 연결어미를 결정하기 위하여 일본어 문장에서 「-て」가 어떤 의미일 때 특정 단어나 문형과 함께 출현하는가의 여부를 알아보는 방법이다[8]. 이 방법은 매우 간단하나 의미 구분을 위한 고유 단어나 문형파악이 어렵고 이 방법으로 번역될 수 있는 문장의 수가 많지 않다는 단점이 있다.

이 방법을 이용하기 위하여 연결어미 「-て」가 문장 속에서 갖는 의미에 따라 일본어 문장을 14그룹으로 구별하고 각 그룹별로 고유 패턴을 추출해 보았다. [표-2]에서 분리된 14개의 의미그룹 중에서 8개 그룹의 고유패턴은 확인되었으나 나머지 6개 그룹의 문형이 파악되지 않았다. 이 방법은 특수의미를 지닌 「-て」의 번역을 쉽게 할 수 있으며 기계번역기에서의 구현 또한 간단하다. 그러나, 고유문형이 파악되지 않은 나머지 6개 그룹의 출현빈도가 상대적으로 높기 때문에 이 방법만으로 「-て」의 정확한 번역이 이루어질 수는 없다.

[표-2] 의미별 고유문형 파악 상황

의미	예문	고유문형	한국어 연결어미
순서	映画を見て考えた.	발견못함	-고
동시진행	本を読んで暮らす.	발견못함	-고
원인이유	明るくて眩しい.	발견못함	-아
수단방법	バスに乗って行く.	발견못함	-고
결과	全部食べて千円	발견못함	-아
양보허용	起きていいよ.	양보허용어형이 연결됨 ²	-도
역설조건	義を見て行かない.	발견못함	-도
병렬공존	運動もして, 勉強もする.	N1도V1, N2도V2 N1도A1, N2도A2	-고
병렬대비	冬は過ぎて, 春は来た.	N1はV1, N2はV2 N1はA1, N2はA2	-고
연속적 나열	秋の空は青くて, 美しくて, 高い.	「て,」 2번 이상 출현	-고
자격제시	彼を大将にして, 敵兵を破った.	자격제시어형이 연결됨 ³	-아
대상제시	家に向かって走る.	대상제시어형이 연결됨 ⁴	-아
지속적 행위상태	春から夏にかけて 咲く.	지속상태어형이 연결됨 ⁵	-아
강조	寒くて寒くてたまらない.	て 2번 이상 출현	-고

2) 통사론적 접근

한국어 연결어미 「-아」 나 「-고」 가 포함된 문장의 구성요소 또는 품사의 차이점을 이용하는 방법이다. 즉, 일본어와 한국어는 통사적 구조가 비슷하다는 특성을 이용하여 서로 대응하는 연결어미를 파악하는 방법이다. 이 방법은 그 적용범위가 광범위하나 한일 양국어에서는 주어, 목적어 등의 문장 주요 요소가 자주 생략되기 때문에 적용이 애매한 경우가 나타나기도 한다.

문법학자들은 한일 양국어의 유사성에 기반하여 한국어와 일본어 연결어미를 문법적으로 고찰한 결과, 연결어미 「-て」 에 대응하는 한국어 연결어미는 「-아」 와 「-고」 라는 사실을

2 양보허용 어형 : -て いい, -て かまわない 등
 3 자격제시 어형 : -に して, -に なって 등
 4 대상제시 어형 : -に ついて, -に くらべて 등
 5 지속상태 어형 : ~に かけて 등

발견하였다. 그리고 연결어미를 통사적으로 고찰한 여러 연구들에서 문장상에서의 통사적 비교 기준들이 몇 가지 제시되었지만 실제 적용 가능한 것으로 판별된 것은 주어의 동일성/비동일성 여부와 서술어 품사의 동일성/비동일성 여부의 두 가지 정도였다. 이 두 통사적 비교 기준에 입각하여 한국어와 일본어의 문장들을 비교 관찰한 결과, 일본어 연결어미 「-て」 와 한국어 연결어미 「-고」 는 통사적인 면에 있어서 비교적 유연성을 갖는 반면, 한국어 연결어미 「-아」 는 통사적 제약이 철저함을 발견할 수 있었다. 그러므로 이 관찰 결과를 이용하여 「-て」 를 한국어로 번역할 경우 연결어미 「-아」 가 갖고있는 통사적 제약조건이 그 일본어 문장에서 발견되면 「-아」 로 번역하고 그 나머지들은 모두 「-고」 로 번역시킴으로써 「-て」 의 중의성을 일부 해소할 수 있다.

이 방법을 사용하면 일본어 문장상에 나타나는 대부분의 「-て」 가 번역 가능하나 두 서술어가 연속하는 경우에 사용된 「-て」 는 주어와 목적어가 존재하지 않기 때문에 통사적 분석이 어려워 본 방법의 적용이 어렵다. 그 중에서도, 예를 들어 동사+형용사의 조합처럼 연속한 서술어의 품사들이 서로 다른 경우에는 품사의 상이성으로 인한 종속관계가 발생하여 「-아」 로 번역이 가능하나 동일 품사의 서술어로 연결된 경우에는 대응하는 한국어 연결어미를 결정하기가 어렵다.

3) 의미론적 접근

「-て」 앞·뒤에 연결된 서술어들에 대해 의미 분류를 한 후 그 분류 결과에 의해 「-て」 에

[표-3] 연결어미 「-아」 의 제약조건

	일본어	한국어	제약조건
접속문	伯父が来て私は嬉しかった.	삼촌이 와서 나는 기뻐다.	비동일 주어 (伯父 ≠ 私) 비동일품사 술어 V(来る) ≠ A(嬉しい)
접속 용언	腹が痛くて起きられない.	배가 아파 일어날 수 없다.	비동일품사 술어 A(痛い) ≠ V(起きる)

대응하는 연결어미를 결정하는 방법이다.

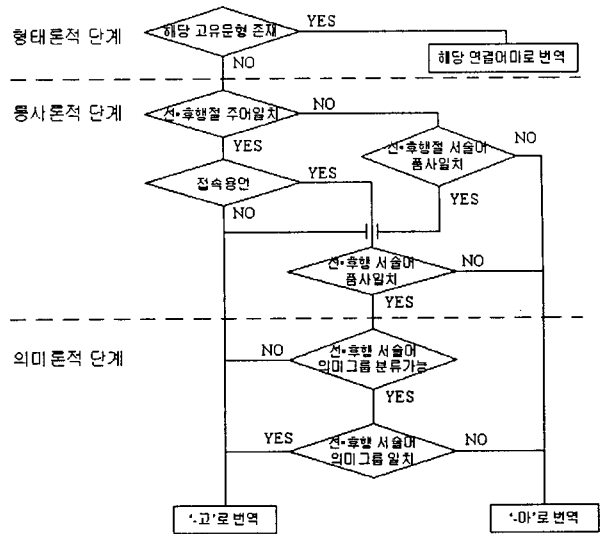
먼저, 이 방법을 적용하기 위해 서술어들을 상적자질에 따라 [표-4]와 같이 네 그룹으로 나누었다[8]. 동일 그룹에 속한 두 서술어를 연결한 경우의 「-て」는 서술어간에 대등한 의미를 갖게 하므로 한국어 연결어미 「-고」로, 비동일 그룹에 속하는 서술어가 연결된 경우에는 종속의 의미가 발생하므로 한국어 연결어미 「-아」로 번역한다. 이 방법은, 서술어들만을 접속한 「-て」의 번역은 가능하지만 두개의 문장을 접속하는 「-て」의 번역이 어려운 단점이 있다.

[표-4] 의미별 용언의 분류

	상적 자질			예
	상태	순간	완성	
상태용언	○	X	△	아프다, 고프다 길다, 착하다
순간용언	X	○	△	끝나다, 죽다 끊다, 이해하다
완성용언	X	X	○	비우다, 입다 외우다, 열다
비완성용언	X	X	X	웃다, 사랑하다 사용하다, 놀다

2.3 본 논문의 번역방법

지금까지 일본어의 대표적 연결어미인 「-て」를 이에 상응하는 한국어 연결어미로 번역할 경우 적용가능한 언어학적 접근방법들을 살펴보았다. 형태론적 접근으로는 특수어의 「-て」를 해석할 수 있었으나 범용성이 부족하였고, 통사론적 접근으로는 접속문의 해석은 가능했으나 접속용언의 처리가 어려웠으며, 의미론적 접근으로는 접속용언의 처리만이 가능하였다. 그러므로, 이들 3가지 언어학적 접근법을 적절히 적용하면 특수적인 경우는 물론, 일반적인 경우에 있어서도 접속문 및 접속용언의 번역이 가능하리라 생각된다. 제안하는 번역방법은 지금까지 기술된 내용을 기반으로 하여 세단계로 구성되며 그 흐름은 [그림-1]과 같다. 위의 세단계에서 처리하지 못한 문장은 기존의 방법에서와 같이 「-고」로 번역되도록 하였다.



[그림-1] 제안하는 「-て」의 번역방법

3. 실험 및 평가

제안한 「-て」의 번역방법에 대한 성능평가를 위하여 연결어미 「-て」가 일본어 문장에서 동사와 형용사에 접속하는 문장만을 대상으로 하여 실험하였다. 실험에 사용한 문장은 일본어사전[10,11,12]에서 추출한 332개이며⁶ 추출된 문장을 구성에 따라 접속문 형태와 접속용언 형태로 구분하였다.

먼저, 접속문 형태를 구성하는 연결어미 「-て」에 대한 번역실험 결과는 [표-5]와 같다.

[표-5]에서 나타난 것과 같이, 기존의 방법은 적용률이 100%이지만 정확도는 55.4%로 비교적 저조함을 알 수 있었다. 또한, 형태론적 접근 방법이나 통사론적 접근 방법을 적용하면 좀더 높은 정확률을 기대할 수 있으나 고유문형으로 매칭되는 문장이 많지 않기 때문에 전체 번역률이 매우 저조하였다. 의미론적 방법은 접속문 형태에서는 적용되지 않으므로 문장 실험을 시도하지 않았다. 그리고 본 논문에서 제안한 방법을 사용한 결과,

⁶ 추출한 예문은 편의상 문두, 문중, 문말을 생략했거나 수정하여 작문한 것임을 명기해 둔다.

정확률과 번역률이 기존의 방법에 비해 향상되었음을 볼 수 있었다.

다음으로 접속용언 형태를 구성하는 연결어미 「-ㄷ」의 번역실험 결과는 [표-6]과 같다.

접속용언을 이루는 「-ㄷ」는 접속문의 경우에 비해 그 출현빈도가 약간 낮았다. 형태론적 접근방법은 접속문 형태에서와 거의 비슷한 번역성과를 보였고 통사론적 접근방법의 경우는 제약조건에 부합되는 문장의 출현빈도가 접속문의 경우보다 현저히 작아서 적용률이 상당히 낮았다.

접속용언 형태를 이루는 「-ㄷ」의 번역에 있어서도, 본 논문에서 제안한 방법이 높은 번역 정확도와 번역률을 보였다. 결과적으로, 번역대상 문장의 구성 형태에 관계없이 특정 한국어 연결어미로만 번역하는 기존의 「-ㄷ」 번역방법을 사용하면 저조한 정확도와 번역률을 보였다. 또한, 문법학자들이 제안한 접근 방법들은 각각의 장점과 함께 치명적인 단점이 발견되어 실제 사용에 어려움이 있음을 짐작할 수 있었다. 즉, 형태론적인 접근방법과 통사론적 접근방법은 번역률이 현저히 낮았으며 의미론적인 접근방법은 접속문을 구성하는 「-ㄷ」의 번역에 적용되지 못하는 단점을 보였다.

그러나 본 논문에서 제안한 방법은 문장형태에 관계없이 「-ㄷ」의 번역 정확도와 번역률에 있어서 고른 우수성을 보였다. [표-7]은 접속문과 접속용언에서의 실험을 통합한 결과이다.

[표-5] 접속문에서의 「-ㄷ」 번역결과

		매칭된문장수	번역성공수	정확률 ⁷	적용률 ⁸	번역률 ⁹
기존의 방법		204	113	55.4%	100%	55.4%
개별적접근법	형태적	25	23	94.3%	12.3%	11.3%
	통사적	177	118	66.7%	86.8%	57.8%
	의미적	-	-	-	-	-
논문의 방법		204	152	74.5%	100%	74.5%

(※ 접속문 문장수 204개)

⁷ 정확률 = 번역성공수 / 매칭된 문장수
⁸ 적용률 = 매칭된 문장수 / 전체 문장수
⁹ 번역률 = 번역성공수 / 전체 문장수

[표-6] 접속용언에서의 「-ㄷ」 번역결과

		매칭된문장수	번역성공수	정확률	적용률	번역률
기존의 방법		128	67	52.3%	100%	52.3%
개별적접근법	형태적	15	15	100%	11.7%	11.7%
	통사적	27	20	74.1%	21.1%	15.6%
	의미적	128	75	58.6%	100%	58.6%
논문의 방법		128	92	71.9%	100%	71.9%

(※ 접속용언 문장수 128개)

[표-7] 전체 「-ㄷ」 번역결과

		매칭된문장수	번역성공수	정확률	적용률	번역률
기존의 방법		332	180	54.2%	100%	54.2%
개별적접근법	형태적	40	38	95.0%	12.1%	11.4%
	통사적	204	138	67.6%	61.4%	41.6%
	의미적	128	75	58.6%	38.6%	23.5%
논문의 방법		332	244	73.5%	100%	73.5%

(※ 전체 문장수 332개)

4. 결론

본 논문에서는 일한 양국어의 연결어미를 문법적 차원에서 비교 분석하여 번역기에 응용하였다. 특히, 양국어에서 빈번하게 나타나나 심각한 중의성을 야기시키는 일본어 연결어미 「-ㄷ」와 이에 대응하는 한국어 연결어미 「-아/-고」를 비교하고 중의성을 해결하기 위한 번역방법을 제시하였다. 제안한 번역방법을 실제 문장에 적용해 본 결과 번역정확률이 73% 이상으로 크게 개선됨을 보였다.

문법학적인 연구결과들을 기계번역에 이용하는 연구는 자연스런 번역결과물 생성에 보다 근접해 갈 수 있는 지름길이라고 생각하며 앞으로도 기계번역 시스템의 개발에 있어서 기존의 언어 및 문법에 관한 연구결과들이 효율적으로 사용될 수 있기를 기대한다.

참고문헌

- [1] 권경희, “현대 국어 접속어미의 통어론적 연습”, 동아대학 석사논문, 1990
- [2] 김옥희, “日本語の助詞「て」の機能と意味 関係 : 「て」に対応する韓国語との比較 対照を通して”, 한국외국어대학 석사논문, 1989
- [3] 남기심, “국어 연결어미의 쓰임 : 「-고」, 「-어서」, 「-니까」, 「-다가」의 의미· 통시적 특징”, 서광학술자료사, 1994
- [4] 서정수, “연결어미 「-고」와 「-어(서)」”, 언어와 언어학 8집, 1982
- [5] 서정수, “한국어와 일본어의 접속어미 비교연구”, 문법연구 5집, 문법연구회, 1984
- [6] 선홍남, “연결어미 「-아」, 「-고」와 동사의 의미자질에 대한 연구”, 한국외국어대학 석사학위논문, 1992
- [7] 이시형, “한국어 연결어미 「-어」, 「-고」에 관한 연구”, 서강대학 박사학위논문, 1990
- [8] 정문수, “상적 특성에 따른 한국어 풀이씨의 분류”, 문법연구 5집, 문법연구회, 1984
- [9] 정병남, “일한 양국어 연결어미 「-て」와 「-고」의 대조연구”, 계명대학 석사학위논문, 1985
- [10] 尚学図書(編), “国語大辞典”, 小学館, 昭 57
- [11] 日本語教育学会(編), “日本語教育辞典”, 大修館書店, 1982
- [12] 松村 明(編), “日本文法大辞典”, 明治書院, 昭 56