

영한기계번역에서 자질정보를 이용한 한국어 종속부사절의 연결어미 생성

이영우¹⁾, 안동언, 정성중
전북대학교 컴퓨터공학과

apollo@calhpl.chonbuk.ac.kr, {duan,sjchung}@moak.chonbuk.ac.kr

A Conjunctive Generation of Korean Subordinating Adverb Clause using Feature Information In English-Korean Machine Translation

Young Woo Lee, Dong Un Ahn, Sung Jong Chung
Department of Computer Engineering, Chonbuk National University

요 약

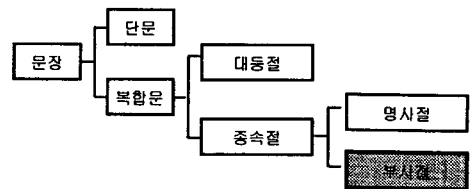
영한기계번역에서 영문의 복합문중 종속부사절이 한국어로 생성될 때, 종속절의 주절에 대한 의미에 따라 다양한 한국어 연결어미를 생성하게 된다. 주절의 의미를 보완하는 종속절은 그 연결어미에 의해 그 역할이 결정되는 것이다. 종속부사절을 이끄는 부사어는 연결어미로 표현되는데 기존의 사전을 기반으로 하는 기계번역시스템에서 사전에 있는 부사어의 표층어만을 이용하였기 때문에 그 생성결과가 만족스럽지 않았다. 영문의 부사어중의 일부는 한국어로 생성될 때 의미적 구분에 따라 여러 가지의 연결어미로 생성이 되어야 하는데 영어 해석에서 종속절의 의미 정보를 충분히 분석하지 못하는 경우가 많다. 본 연구에서는 종속부사절을 이끄는 영어의 부사어와 각 부사어가 생성될 때 필요한 한국어 연결어미를 정리하였다. 또한, 형태소와 구문 자질 정보를 이용하여 여러 연결어미를 갖는 부사어의 경우에 하나의 연결어미를 선택한다.

1. 개 요

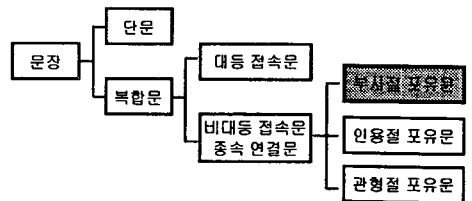
현재 영한 기계번역 시스템에서의 한국어 생성은 한국어 생성을 위한 정보를 수록한 생성사전과 각 어미와 조사의 결합을 나타내는 지식베이스에 의존되고 있다. 생성사전에 의해 한국어 생성에 필요한 각 단어의 한국어 대역어와 단어 정보가 생성시스템에 제공되며, 지식베이스에 의해 한국어에 맞는 조사와 어미가 생성되어 한국어 문장이 완성된다. 본 연구에서는 기존의 지식베이스의 범위를 넓혀, 영어문장이 갖는 구조에서 복합문중 종속절의 명사절과 부사절 가운데, 문장구조가 주절과 종속절로 이루어지는 부사절에 관해 영어와 한국어간의 종속적 연결어미에 대해 제시한다.

영한기계번역 시스템에서의 한국어 생성은 원시언어인 영어 문장의 자질 요소와 종속절을 이끄는 부사어로서 목적언어인 한국어 종속부사절의 연결어미로 대응될 수 있다. 즉, 본 논문에서는 영문에서 각 종속부사절을 이끄는 부사와 종속절의 자질 정보를 이용하여 영문 종속부사절이 한국어 종속부사절로 기계번역될 경우, 주절과 종속절을 연결하는 어미를 생성한다

한국어의 부사절과 영어에서의 부사절은 의미적인 정의는 같지만,



<그림 1> 영어 문장의 구조



<그림 2> 한국어 문장의 구조

실제 종속부사절의 연결어미의 분류상에서는 차이를 보인다. 한국어에서의 종속절 연결어미의 분류는 어미자체의 의미적인 차이로 분류를 하지만, 영문에서의 부사절은 종속절 전체의 의미적인 요소로 분류하고 있다. 영문에서 종속부사절을 이끄는 부사는 단일의 연결어미를 갖는 부사와 그 종속절의 의미, 역할 차이에 따라 여러 연결어미를 갖는 부사로 나눌 수 있다.

여러 연결어미를 갖는 부사어와 단일한 연결어미를 갖는 부사어를 정리해 이를 응용할 수 있도록 정리하고, 여러 연결어미를 갖는 부사에 대해 자질을 이용한 하나의 연결어미를 선택하여 생성하도록 하는 것이 본 연구의 목적이다.

2. 관련 연구

2.1 부사절

그림 1을 보면 영어문장의 구조는 크게 단문과 복합문으로 나뉘며 이 복합문은 다시 대등절과 종속절로 나뉘는 것을 알 수 있으며, 이 복합문은 두 개 이상의 낱말이나 또는 어구가 일정한 구문 안에서 등장하는 문장이다. 대등절은 일정한 낱말이나 어구가 이 사이의 관계를 나타내는 연결어(Connective)에 의해 대등하게 연결되는 것이고, 종속절은 종속접속사(Subordinating Conjunction)에 의해 주절과 종속절로 이루어지는 구문이다. 영문에서 이 종속절은 2가지로 나뉘는데, 명사절과 부사절 안김으로 나누어진다.

명사절이란 3가지 분류로, 주어의 역할을 하는 주어절, 목적어의 역할을 하는 목적어절, 서술어의 역할을 하는 서술어절로 나눌 수 있고, 종속절의 다른 분류인 부사절은 “주절에 대해 부사와 같은 역할을 하는 종속적인 문법적 비중을 갖게 되는 절”로 정의하고 있어[1], 종속부사절이 주절에 대해 그 문장역할을 가감하는 의미적인 요소로 작용하는 것이다.

영어에서의 부사절 분류는 다음과 같다. [1]

- 장소절, 시간절, 이유절, 목적절, 결과절, 양태절, 상황절, 정도절, 제한절, 비교절, 일치절, 대조절, 양보절, 조건절

한국어에서 단순문(Simple Sentence)이란 기본적으로 주어와 서술어의 관계가 한 번만 이루어진 것이고, 복합문(Multiple Sentence)은 두 개 이상의 절 또는 단순문이 결합하여 이루어진 것이다. 복합문은 대등접속문과 비대등접속문으로 나눌 수 있고, 대등문은 대등한 구조를 가진 두 성분 절이 대등 접속사(Coordinator)에 의해 맞섬 관계로 이어지는 구문이고, 비대등문이란 복합문 중에서 대등접속문이 아닌 것을 통틀어 가리킨다.

한국어에서의 부사절의 분류 다음과 같다. [2]

- 상황접속소, 계기한정접속소, 까닭접속소, 조건접속소, 첨가접속소, 점중접속소, 비교접속소, 반복접속소, 시간접속소, 양보접속소, 의도목적접속소, 결과관계접속소, 전환접속소

한국어에서 부사절이란 종속절이 갖는 의미적인 기준에서의 부사절의 분류가 아닌, 연결어미의 의미에 의존된 분류를 채택하고 있다. 그래서 그 분류상의 결과를 볼 때 영문과 국문에서의 부사절 분류가 다르지만, 부사절의 의미와 그 문장구조는 동일하다고 할 수 있다.

2.2 부사절 처리

기계번역 시스템에서 부사절 연결어미의 생성은 목적언어인 한국어의 생성결과에 영향을 미친다. 다음 a) 문장들은 상용 영한기계번역 시스템을 이용하여 부사어의 연결어미를 생성한 결과이고 이를 문맥에 맞도록 수정한 결과는 b)이다. 현재, 단일 연결어미를 갖는 부사어의 경우는 종속절에 맞는 어미로 생성되지만, 아래와 같이 여러 어미를 갖는 부사어의 경우는 대표적인 하나의 연결어미로 생성되고 있다.

(예1) As I point my sword, the sun arises.

- a) 내가 내 검을 지적함에 따라, 태양은 뜹니다. (상용시스템)
- b) 내가 지적하는 곳에서 태양은 뜹니다. (개선된 문장)

(예2) They shouted as we were laughing.

- a) 우리가 웃고 있음에 따라 그들은 큰소리를 냅니다. (상용시스템)
- b) 우리가 웃고 있을 때에 그들은 큰소리쳤습니다. (개선된 문장)

(예3) As he is an honest boy, we can trust him.

- a) 그가 정직한 소년임에 따라, 우리는 그를 신뢰할 수 있습니다. (상용시스템)
- b) 그가 정직한 소년이므로, 우리는 그를 신뢰할 수 있습니다. (개선된 문장)

종속부사절은 일정한 접속사와 이와 대등한 접속구에 의해 유도되지만, 전통적인 절의 형식을 취하지도 않고 동등한 역할을 하는 동등구문은 분사나 부정사 동명사 등을 동원하거나 그렇지 않으면 생략 또는 각종 어순의 변화 등에 의하여 적절히 부사절과 동등한 구문으로 나타나는 구문이다. 본 연구에서는 동등구문이 아닌 부사어 중심 구문에 기반하여 연결어미에 대해 연구한다.

2.3 선행 연구

대량의 코퍼스로부터 지식을 습득하고 이를 이용하는 방법(Corpus-Based approach)은 크게 2가지로 나뉜다. 직접적인 방법인 번역의 변환, 생성 분석을 위해 코퍼스에서 뽑아진 정보를 이용하는 방법과 간접적인 방법으로 원시언어와 목적언어의 통계정보와 형태소, 문법, 지식구축을 위해 코퍼스를 바탕으로 하는 것이다[6]. 본 연구에서는 태깅되지 않은 영한 대역 코퍼스(Bilingual Corpus)로부터 영문과 한국어의 상호 번역 관계와 그 정보를 얻어서 이용하였다.

3. 부사절 어미

3.1 부사어 선택

종속부사절에 사용되는 부사는 먼저 문법적 분류에 근거해 일차로 분류하고 이를 다시 교과교과서, 맨투맨 독해, 성문해심, 시사고교 정리, Ohenry 단편, World News, 소설 Gatsby, Pearl, 오만과 편견(아가사 크리스티), Synnery 고교독해집, 중학교 영어교과서의 영한 대역 코퍼스(Bilingual Corpus)를 통해 22,378개의 영한 대역 문장을 구축하고 이를 통해 코퍼스에서 사용되는 62개의 종속부사절을 이끄는 부사어를 수집했으며, 자질정보를 추출하였다. 또한, 코퍼스의 일부 문장을 한국어 종속부사절의 연결어미 생성의 실험에 사용했다.

[표 1] 부사어

according as	after	although	anywhere
as	as far as	as if	as long as
as many as	as much as	as regards	as respects
as soon as	as though	because	before
but	considering that	directly	every time
for fear that	for regards	for the reason that	however
if	immediately	in order that	in so far as
instantly	lest	just so	like
like as	nor	now that	once
seeing that	since	so	so long as
so much as	so that	than	that
the instant	the moment	though	till
to the end that	unless	until	whatever
when	whenever	where	whereas
whether	wherever	whichever	while
whoever	without		

3.2 부사어의 분류

[표 2] 단일 연결어미를 갖는 부사어 (일부)

부사어	영문절	사전적 영문 의미	국문절 분류	국어 어미
According as	일치절	...에 따라서	정중접속소	e 수륙
After	시간절	...뒤에	시간전후접속소	L 뒤에
Although	양보절	비록 ...임지라도	양보접속소	비록, 지라도
Anywhere	장소절	어디든지	양보접속소	L 곳은 어디라도
As far as	제한절	...하는 한은	제한접속소	L 한은
As if	양태절	...처럼	양태접속소	L 것처럼
As long as	제한절	...하는 한은	제한접속소	L 한은
As many as	정도절	...만큼	정도접속소	L 것만큼
As much as	정도절	...만큼	정도접속소	L 것만큼
As regards	정도절	...만큼	정도접속소	L 것만큼
As respects	정도절	...만큼	정도접속소	L 것만큼
As soon as	시간절	...하자마자	시간순차접속소	자마자
As though	양태절	마치 ...인 것처럼	양태접속소	L 것처럼
Because	이유절	...때문에	까닭접속소	기 때문에
But	조건절	...않고	조건접속소	L 다면
Considering that	이유절	...이므로	까닭접속소	기 때문에
Directly	시간절	...하자마자	시간순차접속소	자마자
Every time	시간절	...할 때마다	시간유발관계	e 때마다
For fear that	이유절	...하지 않도록	결과관계접속소	기 않도록
For regards	제한절	...에 관해서	계기한정접속소	L 것에 관해서

위 62개의 선정된 부사어를 통합하여 전체 부사어에 대한 테이블을 작성하였다. 부사어의 의미적인 분류를 위해 앞에 나온 국문 부사절의 종류를 다음과 같이 확장하여 단일 연결어미를 갖는 부사어 55개[표1]와 여러 어미를 갖는 부사어는 7개[표2]를 결정하였다.

테이블에서 영문절은 [1]의 분류에 의해 정리하였고, 영문의 부사절과 국문의 부사절의 분류가 다르므로, 국문의 분류는 [2]의 분류를 확장하여 대응되도록 하였다.

[표 3] 여러 연결어미를 갖는 부사어

부사어	영문절	사전 영문의미	국문절 분류	국어 어미
As	정도절	...만큼	정도접속소	L 것만큼
	양태절	...처럼	양태접속소	L 대로
	시간절	...하고 있을 때	시간동시접속소	e 때에
	이유절	...이기 때문에	까닭접속소	기 때문에
	일치절	...할수록	정중접속소	e 수록
Before	대조절	...이지만	양보접속소	지만
	시간절	...보다 이전에	시간전후접속소	기 전에
If	대조절	...보다 차라리	대조접속소	느니 차라리
	조건절	만약 ...다면	조건접속소	L 다면
	양보절	인지 아닌지	양보접속소	L 지 아닌지
	양보절	비록 ...임지라도	양보접속소	e 지라도
Since	대조절	...이지만	양보접속소	지만
	시간절	...할 때	시간유발관계	e 때마다
	이유절	...이므로	까닭접속소	기 때문에
That	결과절	...하기 위해	계기한정접속소	L 후에
	이유절	...때문에	까닭접속소	기 때문에
	상황절	...이다	판단접속소	다
	제한절	...바로	제한접속소	L 한은
Whenever	양보절	언제 ...하든지	양보접속소	언제 ...지라도
	시간절	...할 때마다	시간유발관계	e 때마다
While	시간절	...하는 동안	시간지속관계	L 동안
	양보절	...할지라도	양보접속소	지만

[표 4] 종속부사절의 빈도

부사어	국어 어미	전체	해당 어미	
As	L 것만큼	287	12	4%
	L 대로		135	47%
	e 때에		57	19%
	기 때문에		29	10%
	e 수록		42	17%
before	지만	217	12	4%
	기 전에		217	100%
	느니 차라리		0	0%
	L 다면		722	98.4%
	e 지라도		3	0.4%
If	L 지 아닌지	733	5	0.6%
	지만		3	0.4%
	e 때마다		0	0%
	L 후에		56	67%
	기 때문에		27	33%
Since	어시	141	129	91%
	기 위해서		4	3%
	기 때문에		6	5%
	다		2	1%
	L 한은		0	0%
Whenever	언제 ...지라도	9	0	0%
	e 때마다		9	100%
	L 동안		165	96%
While	L 동안	171	6	4%
	지만		6	4%

위 [표 4]는 여러 연결어미를 갖는 부사어에 대해 영한 대역 코퍼스에 의한 빈도를 나타낸 것이다. "as"와 "since"의 경우는 고른 분포를

보였고, "before", "if", "whenever", "while"은 여러 연결어미중 하나에 편중되는 현상을 보였다. "that"의 경우는 결과를 나타내는 연결어미 "어서"에 집중되었지만, "so that"과 "in order that" 같은 숙어적 표현이 전부를 차지했다.

현재 영한기계번역시스템을 통해서는 부사어에 의해 이끌어지는 종속부사절의 역할과 의미를 알 수 없고, 이러한 종속절의 역할을 규명하지 못하면 그 종속절이 주절과 연결되며 상호간 의미적인 관계를 설정하는 연결어미 역시 결정될 수 없다. 이를 위해, 현재 기계번역 시스템 상에서의 종속부사절의 역할은 그 종속절과 주절의 정보에 의해 분류하게 된다. 즉, 종속절의 역할과 의미에 대한 분류정보를 현재 기계번역 시스템의 영어 해석에서 충분한 정보를 얻을 수 없으므로, 본 연구에서는 코퍼스로부터 얻어진 직접적인 형태소 정보와 양상, 구문정보를 이용하기로 한다.

4. 실험

종속절을 이끄는 부사어가 여러 연결어미를 갖는 부사에 대해, 영한 대역 코퍼스의 문장에서 빈번히 나타나는 형태소, 양상, 구문적 특성을 찾아 그 정보로 사용하였다. [표 5]에서는 "as"와 "since"에 대한 자질정보를 보여주고 있다.

[표 5] 자질정보 (일부)

부사어	국어 어미	순위	분류	정보
As	ㄴ 것만큼	5	형태소정보	
			양상정보	
			구문정보	"As" ~ "as"
	ㄴ 대로	1	형태소정보	"As you know"
			양상정보	"Just as"
			구문정보	"As" + 종속절, "so" + 주절
	ㄴ 때에	2	형태소정보	"As" + 명사
			양상정보	주절/종속절간 시제 일치
			구문정보	
	기 때문에	4	형태소정보	
양상정보				
구문정보			"As" 종속절이 문두로 온다 형용사부사/명사 + "as"	
ㄴ 수록	3	형태소정보		
		양상정보		
		구문정보		
지만	5	형태소정보		
		양상정보		
		구문정보	형용사부사/명사 + "as" 동사 + "as" + 조동사 주절/종속절간 주어 일치	
Since	ㄴ 후에	1	형태소정보	"Ever since" "It is"의 주절
			양상정보	주절이 완료형이다. 종속절이 과거이다.
			구문정보	
	기 때문에	2		

[표 5]의 자질정보를 이용하여 실험을 하였으며 생성에서 해당하는 자질정보가 없을 경우에는 연결어미의 출현빈도에 따라 가장 많이 나오는 연결어미를 선택하였다. [표 5]에서 순위는 코퍼스에서 해당하는 연결어미의 출현빈도순이다.

"as"를 볼 때, 종속절이 양태적 의미를 갖는 연결을 위해 "ㄴ 대로"로 대표 연결어미를 결정해 실험했지만, 실제 문장을 조사해 볼 때, "같이"와 "대로"의 2가지가 서로 혼용되어 나타나는 형태를 보였으며, 영문과 국문의 번역시의 모호성으로 다른 분류의 연결어미이지만, 때로는 같은 의미로 사용될 때도 있었다.

[표 6] 연결어미 생성 실험 결과

부사어	연결어미	테스트 문장수	해당 문장	오류 문장	정확률
As	ㄴ 것만큼	60	26	0	100%
			12	0	100%
			9	8	12%
			5	3	40%
			7	3	57%
			1	0	100%
Since	ㄴ 후에	20	18	0	100%
			2	2	0%

[표 6]은 연결어미를 생성한 실험 결과이다. "as"에서 연결어미 "ㄴ 것만큼"으로 생성되는 경우에는 (예4)와 같이 "as" ~ "as"의 특성만으로 100% 정확률을 보인다.

(예4) A Child is as important as a man.

- a) 아이는 어른만큼 중요합니다. (상용시스템)
- b) 어린이는 어른만큼 중요합니다. (개선된 문장)

그렇지만 "as"가 연결어미 "ㄴ 때에"로 생성되기 위한 자질정보는 충분하지 않아서 매우 낮은 정확률을 보이고 있다. (예5)와 (예6)에서 보듯이 정확한 연결어미가 생성되지 않았다.

(예5) As the fish loses a tooth, another moves into its place.

- a) 물고기는 이빨을 잃음에 따라, 또 다른 것이 그것의 장소 안으로 이동합니다. (상용시스템)
- b) 물고기는 이빨 잃을 때에, 또 다른 것이 그것의 장소 안으로 이동합니다. (개선된 문장)

(예6) They use their thumbs to hold on to branches as they climb trees.

- a) 그들은 나무에 오름에 따라 그들은 나뭇가지를 잡기 위해 그들의 엄지를 사용합니다. (상용시스템)
- b) 그들은 나무에 오를 때에 그들은 나뭇가지를 잡기 위해 그들의 엄지를 사용합니다. (개선된 문장)

이 자질 특성값 외에도 본 연구에서는 의미적 분류를 적용하여 실험을 하였다. 의미적 분류는 영한 대역 코퍼스의 문장을 대상으로 각 구성단어에 의미자질[4]을 적용하여 실시하였는데, 주절 또는 종속절 각각의 문장내 단어들간의 의미적 공기관계는 얻을 수 있었지만, 연결어미 선택을 위한 주절과 종속절간의 의미적, 문맥적 역할을 밝혀 알맞은 연결어미를 생성할 수는 없었다.

5. 결론

본 연구는 자질 정보를 이용하여 종속부사절을 이끄는 부사어의 연결어미를 생성하였다. 영어 해석에서 종속절의 의미정보가 주어지지 않았을 때 형태소, 구문, 양상 정보를 이용하여 하나의 연결어미를 선택하도록 하였다.

뚜렷한 형태소, 구문 자질정보가 있는 경우에는 실험 결과가 매우 좋았으며, 상용 기계번역시스템에서 단순히 하나의 연결어미를 생성하는 결과보다는 향상된 연결어미의 생성 결과를 보이고 있다.

그러나, 형태소, 구문, 양상과 같은 표층적인 자질정보만을 가지고 생성할 때, 해당하는 자질정보가 없을 경우에는 정확한 연결어미를 결정할 수 없었다.

올바른 연결어미를 생성하기 위해서는 의미자질을 사용하여야 하지만 기존의 주절 또는 종속절의 한 문장내의 단어들 간에 의미적 공기관계만으로는 부사어 해결에 큰 도움이 되지 않았다. 의미자질의 적용을 위해서는 단어간의 공기정보가 아닌 주절과 종속절간의 관계를 나타내주는 의미자질을 이용하여야 한다.

6. 참고문헌

- [1] 조성식, “영문법 연구Ⅴ”, 신아사, p239 ~ 508, 1990
- [2] 서정수, “국어문법”, 한양대학교 출판원, p1189 ~ 1289, 1996
- [3] 윤명현, “국어의 접속어미 연구”, 한신문화사, 1989
- [4] Anna Weerzbicka, “Semantics”, OXFORD Press, p35 ~ 111, 1996
- [5] 임철수, 이현아, 최명석, 강병규, 이공주, 김길창, “어휘화된 규칙에 기반한 영한 기계번역 시스템”, 한국정보과학회 학술발표논문집 제 24권 2호, 1997
- [6] Proceedings, “MT SUMMIT IV”, Kobe Japan, p16, 1993