

한일 기계번역을 위한 보문의 수식 Scope 해석

이수현

시스템 공학 연구소, 자연어 정보처리 연구부

The Modification Scope Analysis of the Embedded Sentences in Korean and Japanese Machine Translation

SooHyun Lee

Natural Language Information Processing, Systems Engineering Research Institute

요 약

한일 양언어의 복합문은 여러가지의 통어 현상을 가지며, 주어, 목적어 등의 생략 현상으로 문장의 표층상에 나타나지 않는 것이 있기 때문에 수식구조의 처리가 복잡해지고, 구문해석에 있어서 애매성의 요인이 된다. 따라서, 본 논문에서는 DPN 에의하여 한국어와 일본어의 수식 scope 를 해석하는 방법에 대하여 설명한다. 먼저, 한일 양언어의 공통점과 차이점을 찾아내어, 한국어와 일본어의 보문을 표현형식으로 나타내고, 동사의 격정보로부터 DPN 을 구성하여 DPN 상에서 보문의 수식 Scope 를 해석하는 방법에 대해서 설명한다.

1. 서 론

한국어와 일본어는 통어적 유사성이 있고, 한자어 명사는 같은 개념으로 사용되고 있는 등 기본적으로는 유사성이 많으므로 타 언어간의 기계번역에 비하여 유리한 점이 많다. 그러나, 실제로 번역을 할 때에는 여러가지의 문제점이 발생한다. 즉, 양언어의 동사는 n:n 으로 대응하는 경우가 많고, 필수격 조사도 n:n 으로 대응하기 때문에 이에 대한 다역어 처리가 필요하다. 또한, 부정 표현의 경우 한국어와 일본어의 어순이 다르기 때문에 어순의 변환이 필요하다. 한국어와 일본어의 용언의 활용 형태에는 차이가 있으며, 특히, 사역, 수동, 가능의 경우, 한국어는 2 개 이상의 동사로 이루어진 복합 서술부로 구성되는 것에 반하여, 일본어는 동사가 활용하여 동사 어간에 사역, 수동의 조동사가 첨가되는 형태이다. 한국어와 일본어는 품사 분류 체계가 다르므로, 형용동사와 같이 한국어에 없는 품사는 의미적으로 대응하는 단어로 대체하는 등의 처리가 필요하다.

한편, 한일 양언어는 격표식이 발달하고, 용언이 문장을 지배하여, 용언을 중심으로 문장의 수식 구조가 형성되고, 수식어가 피수식어 앞에 놓인다. 특히, 한일 양언어의 복합문은 여러가지의 통어 현상을 가지며, 주어, 목적어 등의 생략 현상으로 문장의 표층상에 나타나지 않는 것이 있기 때문에 수식구조의 처리가 복잡해지고, 구문해석에 있어서 애매성의 요인이 된다. 일본어의 경우에는 문절간의 제약이 약하기 때문에 문법규칙이 방대해 지고, 처리효율이 저하된다 [11,12].

본 논문에서는 한일 기계번역에서 구문해석을 원활히 하기 위한 보문의 수식 scope 해석 방법에 대하여 논한다.

먼저, 한일 양언어의 공통점과 차이점을 찾아내, 보문의 표현 형식을 나타낸다. 동사가 가지는 격정보로부터 DPN

(Dynamic Pattern Net.)을 구성하여 [10], 구성된 DPN 상에서 보문의 수식 Scope 를 해석하는 방법에 대해서 설명한다.

2. 언어학적 고찰

2 장에서는 복합문의 종류와 구성에 대하여 살펴 보고, 한일 양언어의 용언의 기본 구조를 비교한다.

2.1. 복합문의 구성

복합문이란 여러 단순문이나 절들이 단어나 절의 뜻을 더해 주기 위해 결합하여 이루는 문장을 가리킨다.

복합문은 두 문장이 결합할때, 두 문장이 대등한 자격으로 접속되었는가, 종속적으로 접속되었는가에 따라 내포문과 접속문으로 나누어 진다. 즉, 접속문은 둘 혹은 그 이상의 독립 문장이 각각 수위적 성분관계를 가지면서, 접속어미에 의하여 연결되는 문장을 일컫는다[6,7]. 내포문은 어떤 문장이 다른 문장 안의 한 성분으로 내포된 문장을 일컬으며, 내포된 문장을 생략한 주문 만을 가지고는 완전한 독립 문장을

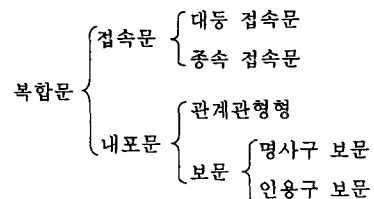


그림 1. 복합문의 종류

- 접속문의 예
나는 학교에 가다가 선생님을 보았다.
- 관계 관형형의 예
나는 학교에 가는 선생님을 보았다.
- 보문의 예
나는 선생님이 걷고 있는 것을 보았다.

구성할 수 없다. 또한, 내포문은 주절과 종속절로 구성되어, 종속절의 관형형 어미를 매개로 절 전체가 뒤의 명사를 수식하는 형태이고, 주절과 종속절의 관계와 기능에 따라 다시 관계관형형과 보문으로 구별된다 [5].

보문과 관계 관형형의 다른 점은, 종속절 직후의 HEAD 명사와 종속절 내의 주어가 일치하는 공지시성의 성립 여부에 있다. 즉, 공지시성이 성립하는 경우가 관계 관형형이고, 성립하지 않는 경우가 보문이다. 위의 예에서와 같이 관계관형형의 종속절 “학교에 가는”의 주어는 “선생님”으로서 주문의 목적어와 공지시성이 성립한다. 그러나, 보문의 예에서 보듯이 종속절을 생략하면 주절의 의미가 통하지 않음을 알 수 있다. 보문은 다시 명사구 보문과 인용구 보문으로 나누어 진다[5].

한국어와 일본어의 수식 구조는 수식어인 명사가 피수식어인 동사의 앞에서 동사를 수식하는 형태이고, 비교차성과 일문일격의 특징을 갖고 있다. 그런데, 보문의 예에서처럼 한 문장이 매립될 때에는 수식의 애매성이 발생한다. 즉, 명사구 “나는”이 동사 “걷고”와 “보았다”의 어느쪽을 수식하는지 애매하다. 본 논문에서는 이의 해결 방법에 대하여 논한다.

2.2. 한국어와 일본어의 용언의 기본구조

한국어 용언과 일본어 용언의 구조에는 근본적인 차이가 있다. 한국어 용언의 기본구조는 다음과 같이 “어간+어미” 혹은 어간과 어미 사이에 선어말 어미가 삽입된 형식이며, 일본어 용언은 “어간+조동사”의 형태로써, 조동사가 활용을 한다.

- 한국어 용언 :
- 종지형 : Vpk = STM + Fk
 - = STM + Aux_STM + Fk
 - 관형형 : Vpk = STM + F9
- 일본어 용언 :
- 종지형/연체형 :
 - VPj = STM + Fj
 - = STM + Aux_Vinf
- VP: 용언, STM: 어간
Fk, Fj: 어미, F9: 관형형 어미
Aux_STM: 한국어 보조어간
Aux_Vinf: 일본어 조동사

한국어 용언이 수식 구조의 관형형으로 활용되는 경우에는 어간에 관형형 어미가 첨가되어, 문장의 수식 구조가 용언의 어미로 인해 결정된다. 일본어 용언은 종지형과 연체형이 같은 형태의 어미를 취하기 때문에 어미 다음의 정보에

의해서 수식구조가 결정된다.

한일 양언어의 관형형과 연체형 어미 및 시제의 관계를 표 1에 보인다.

표 1. 한국어와 일본어의 어미

		한국어		일본어	
		동사/형용사		동사	형용사
		F9		Fj	
관형형	연체형	현재	는	-U	-い
		미래	-르, 을	-U	-い
		과거	-L	-た	-がった
		회상	-던	-た	-がった
종지형	현재	-다 / ㄴ다	-U	-い	
	과거	-었다 / 었다	-た	-がった	

U: う, る, す, ...

3. 보문의 표현 형식

한일 양 언어의 보문은 명사구 보문과 인용구 보문으로 나뉘어 지고 여러 가지의 용어 현상이 있으며[1,3], 표현 형식과 문법 구조에는 차이가 있다. 보문에는 공지시성이 존재하지 않기 때문에 식별자가 사용되지 않는다. 즉, 보문의 형식 명사는 의미를 갖고 있지 않기 때문에 매립문의 성분이 되지 않는다.

3.1. 명사구 보문

한국어의 명사구 보문에는 2가지의 기본형식이 있으며[4,8], 대응하는 일본어의 명사구 보문은 1가지의 기본형식이 있다[2]. 통상적으로 널리 사용되어 지고 있는 한일 양 언어의 보문의 예를 다음에 나타낸다.

- 예1) 나는 선생님이 걷고 있는 것을 보았다.
私は先生が歩いているのを見た.
- 예2) 나는 그녀가 시험에 합격한 것을 알았다.
私は彼女が試験に合格したことを知った
- 예3) 신문은 폴란드 정부가 자유 노조를 인정한 사실을 보도했다.
新聞はポーランド政府が自由労組を認めた事実を報道した
- 예4) 철수는 그가 시험에 합격했음을 알았다.
チョルスは彼が試験に合格したことを知った

한국어의 명사구 보문은 문장이 매립될 때, 예 1)~예 3)과 같이 종속절의 용언의 어미가 관형형 어미로 되면서, 주문의 형식명사 혹은 보문명사를 간접적으로 수식하는 형태와 예 4)와 같이 용언이 명사화 되어 주문을 직접 수식하는 형태가 있다. 이에 반하여, 일본어의 경우는 어미가 연체형 혹은 종지형으로 되면서 형식명사를 수식하는 형태이다.

한일 양언어의 명사구 보문의 표현형식을 (1),(2),(3)에 나타낸다.

한국어의 명사구 보문형식(1)은 종속절 D(S)의 용언 어간에 관형형 어미 F9 (ㄴ,는,르,을,던)이 첨가되어 보문의 의미를 전달시키는 형식명사 (것) 및 보문명사 (CMP)를 수식하는 형식이며, 예 1)~3)의 경우이다. 형식(2)는 D(S)의 용언 어간에 보문자(BMJ)가 붙어서 명사절이 되고, 거기에 조사가 더해져서 직접 주문의 목적어로 되는 형식이며, 예 4)의 경우이다. 형식(1)과 형식(2)는 서로 형태적 교체가 가능하다.

한국어의 표현형식 :

- Scomp -> D(S) F9 + CMP (1)
 Scomp : 보문
 D(S) : 문 S의 변형
 S : 문법적으로 정삼문이고, 종결 어미까지를 포함하고 있는 완전한 문장
 F9 : 관형형 어미
 CMP : 형식 명사 (것) 및 보문명사

- Scomp -> D(S) BMJ (2)
 BMJ : 보문자 (음, 口, 기, 것)

일본어의 표현형식 :

- Scomp -> S CMP (3)
 CMP : 형식명사 (こと, の) 및 보문명사

형식(2)의 보문자의 기능은 일본어에는 없으나 의미적으로 가까운 형식 명사 “こと, の” 에 대응된다. 보문 명사의 예를 표 2 에 나타낸다.

일본어의 명사구 보문형식(3)은 문 S가 연체형 및 종지형으로 되어 형식명사 “こと, の”나 보문명사를 수식하는 형태이고, 한국어의 형식(1)과 형태적으로 등가이다. 따라서, 한국어의 명사구 보문형식(1),(2)과 일본어의 보문형식(3)을 2:1로 대응 시키지 않으면 안된다. 일본어에서 한국어를 생성할 때에는 형식(1)과(2)는 같은 의미가 되므로 한국어로부터 일본어를 생성할 때에는 형식(1)과(2) 어떤 쪽도 형식(3)을 생성하도록 하고, 그반대의 경우에는 형식(3)으로부터 형태적으로 보다 가까운 형식(1)로 번역한다. 이때, 한국어의 관형형어미와 보문자가 일본어의 연체형어미 혹은 종지형어미에 대응한다.

표 2. 보문 명사의 예

한국어	일본어
사실	事實
소리	聲
사건	事件
이유	理由
모양	模樣
진실	眞實
목적	目的
용기	勇氣
소식	消息
경험	經驗
진의	眞意
약점	弱點

3.2. 인용구 보문

인용구 보문은 하나의 문장 안에서 타동사의 목적어가 문절이 되는 경우, 관형형/연체형이 인용 표식과 결합해서 심층구조상에서 목적어가 되는 형태로 다음과 같이 표현된다.

인용구 보문의 생성형식 :

X - Scomp - V -> X - Nobj - V

X는 보문의 앞문절, V는 뒤의 동사, Nobj는 보문 Scomp가 명사적 작용을 하여서, 심층상에서 서술어 V의 목적격으로 쓰여진다.

다음의 예 5)-예 7)은 일상 사용하고 있는 한일 양 언어의 인용 보문의 예로서, 인용 표식 “고,도록/계)와 “と,ように)가 각각 대응된다.

명사구 보문과 다른점은, 인용구 보문은 종속절의 용언이 부사형 어미와 결합하며, 인용 표식이 사용된다는 것이다.

예 5) 나는 여동생에게 일찍 돌아 가도록 권했다.

私は妹に早く歸るように勧めた

예 6) 철수는 영희가 바보라고 생각했다.

チョルスはコンヒがバカだと思った

예 7) 철수는 영희가 내일 온다고 알고 있다.

チョルスはコンヒが來ると思った

한일 양언어의 인용보문의 표현형식을 (4),(5),(6)에 나타낸다.

한국어의 표현형식 :

- Scomp -> S CMP (4)
 CMP : 인용표식 ((라)고)

- Scomp -> D(S) CMP (5)
 CMP : 인용표식 (도록/계)

일본어의 표현형식 :

- Scomp -> S CMP (6)
 CMP : 인용표식 (と, よう(に))

인용보문 Scomp는 한일 양언어가 거의 같은 구조로서 문 S의 끝에 인용 표식 CMP가 붙는다. 형식(4)의 CMP는 형식(6)의 CMP “と”와, 형식(5)의 CMP는 형식(6)의 CMP “よう(に)”와 각각 1:1로 대응 된다.

4. 보문의 수식 Scope 해석

4.1. DPN의 구성

한국어와 일본어는 용언이 문장을 지배하며, 조사가 격관계의 정보를 나타내는 표식이 된다[1]. 즉, 이들 격조사를 가진 명사구가 용언에 대해서 동작주, 장소 등의 의미 관계를 나타내기 때문에 한국어와 일본어의 처리에는 격문법이 이용된다. 따라서, 본 시스템에서는 용언이 가지는 격정보를 pattern화하여 구문해석시에 이를 이용한다.

DPN의 변환 과정은 그림 2와 같이 4가지 부분으로 구성된다. 즉, 계층구조의 격 Pattern 사전, 후보 문형의 생성, 문형 복합체의 생성, 심층구조의 DPN 변환의 4가지로 구성되며, 다음과 같은 순서로 변환된다.

- ① 사전의 격 Pattern [명사 (의미소성) + 격조사]는 필수격 (Vi)와 임의격 (Qj)로 나누어서 임의격이 유연하게 취하여 질 수 있도록 하여, 계층 구조의 Vi, Qj 사전을 구성한다.
- ② 용언에 수속되어 있는 격지배 정보 (Vi, Qj)를 사전으로부터 불러내어 계층구조의 격 Pattern을 Pass 하여 가상문형으로 조합하고, 이것을 표층격과의 Matching에 의해 후보 문형을 선택한다.
- ③ 후보 문형의 부분 Pattern을 재구성하여 문형 복합체를 생성한다.
- ④ 문형복합체를 심층구조의 DPN으로 변환한다.

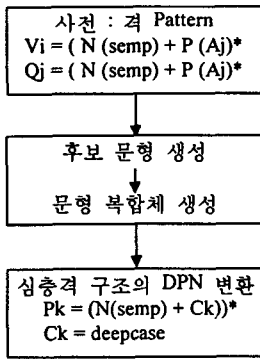


그림 2. 심층격 구조의 DPN 변환 과정

4.2. 수식 Scope의 결정 순서

보문의 수식 Scope는 1문1격의 원칙에 기반하여 다음의 순서에 의해 결정된다.

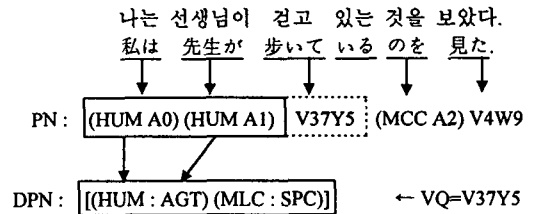
- ① 좌측부터 순차적으로 용언을 찾는다.
- ② 용언의 어미가 관형형인가를 확인한다.
- ③ 관형형이면 다음에 형식명사나, 보문명사가 있는가를 확인한다.
- ④ 입력문장의 형태소해석결과인 명사구의 의미 소성으로 PN을 구성하고, 관형형 용언의 DPN을 구성한다.
- ⑤ DPN 상에서 관형형 용언 앞의 성분을 테스트한다.
- ⑥ PN 상에서 관형형 앞의 복수개의 의미소성의 격과 DPN 상에서 하나의 격으로 일치할 때를 충돌한다고 한다.
- ⑦ 충돌하면, 앞의 명사구를 주문의 성분으로 한다.
- ⑧ 수식 scope는 충돌한 명사구에서부터 관형형까지로 한다.
- ⑨ 의미 소성간의 충돌이 없는 경우는, 관형절의 용언앞에 있는 모든 명사구를 관형절의 성분으로 한다. 이때, 수식 scope는 의미 소성의 충돌이 없는 명사구에서부터 관형형까지로 한다.

4.3. DPN에 의한 수식 Scope 해석

예 1)과 예 2)의 형식명사 “것/의, 것”은 어휘적 의미는 갖고 있지 않다. 또, 예 3)의 보문명사 “사실/事實” 등도 매립문의 성분으로는 작용하지 않지만, 보문의 의미를 전달한다. 따라서, 예 1은 관형형어미 “는, 한”과 형식 명사 “것/의, 것”에 의해서 보문임을 판별하고, DPN 상에서 보문의 의미 성분과 주문의 격구조의 식별에 의해서 수식 scope를 결정한다. 다음에 예 1의 수식 scope의 해석에 대하여 설명한다.

- ① 그림 3의 입력 문장의 형태소 해석 결과인 명사의 의미 소성과 동사의 pattern을 배열한다.
- ② 보조용언 “있는”의 관형형어미와 형식명사 “것”으로부터 입력문장이 보문임을 인지한다.
- ③ 주용언 “걷고”앞에 있는 명사의 의미소성으로 PN을 구성한다.
- ④ 동사 “걷고”의 pattern “V37Y5”로부터 DPN을 구성하여 PN과 일치시킨다.
- ⑤ 명사 “나:私”와 “선생님:先生”이 DPN 상의 (HUM:AGT)에서 충돌되어 “나는:私は”은 규칙에 의해서 밀어 내어 지고, 주문의 주격이 된다.

- ⑥ 따라서, 수식 scope는 명사 “선생님:先生”에서부터 동사 “있는:いる”까지로 되어 다음과 같이 수식 scope가 결정된다.
- ⑦ Scope 내의 문장과 Scope 밖의 문장을 별개의 문장으로 보고, 각각에 대하여 구문해석을 한다. 구문해석에 의해 작성된 tree를 그림 4에 보인다.



수식 Scope 해석결과 :

나는 { 선생님이 걷고 있는 } 것을 보았다.
私は { 先生が歩いて 있는 } のを見た.

그림 3. DPN에 의한 수식 scope 해석

5. 결론

본 논문에서는 DPN에 의하여 한국어와 일본어의 수식 scope를 해석하는 방법에 대하여 설명했다. 그를 위하여 먼저, 한일 양언어의 공통점과 차이점을 찾아내어, 한국어와 일본어의 보문을 표현형식으로 나타내고, 동사의 격정보로부터 DPN을 구성하여 DPN 상에서 보문의 수식 Scope를 해석하는 방법에 대해서 설명했다.

본 방법은 구문해석에 있어서 종래의 방법보다 필요한 문법 규칙이 적어지므로 처리 효율이 향상되었으며, DPN은 한국어와 일본어의 양쪽에 동시에 적용시킬 수 있으므로, 한일 양방향 시스템에 적용하기에 적합하다. 본 시스템의 형태소 해석은 문헌[11]의 형태소 해석기를 이용하였으며, Lisp으로 작성되어 IBM PC에서 구현되었다.

6. 참고 문헌

- [1] 紫谷方良, “일본어의 분석”, pp.65-165, 大修館書店 (1988).
- [2] 久野すすむ, “일본문법연구”, pp.137-142, 大修館書店 (1988).
- [3] 이홍배, “한국어의 관계절화”, 어학연구, 11, 2, pp.289-300 (1975).
- [4] 남기심, “완형 보문 구조 고”, 어학연구, Vol.9, No.2, pp.162-182 (1973.12).
- [5] 남기심, “국어 완형 보문법 연구”, 국어학회 (1989).
- [6] 최재희, “국어의 접속문 구성연구”, 탑출판사 (1992).
- [7] 권재일, “국어의 복합문 구성 연구”, 집문당 (1985).
- [8] 심재기, “관형화의 의미구조”, 어학연구, Vol.15, No.2, pp.109-121 (1979.12).
- [9] S.H.Lee and S.Ozawa, “Modification Scope Analysis of the Embedded Sentences in Korean-Japanese Machine Translation”, proc. of JTC-CSCC'92, pp.130-133 (1992).
- [10] S.H.Lee, K.R.Han, and J.K.Lee, “A Bidirectional MT System between Korean and Japanese based on a Pattern Net”, proc. of

SITA'87, pp.405-410, (1987).

[11] 一本木眞史 외, “黑板モデルを導入した階層的な知識分割に基づく日本語意味解析システム”, 電子情報通信學會論文集(D), J70-D, 11, pp.2294-2299 (1987.11).

[12] 荒木健治 외, “附屬語パターンを用いたルールによる表層レベルでの係り受け關係の解析”, 情報處理學會自然言語處理研究會報, 88-NL-68-6 (1988.9).

심층적 구조

N1P : (N1(나 (HUM)A0))-(HUM:TOP)
 N2P : (N2(선생님(HUM)A1))-(HUM:AGT)
 VQ(1) : (V(걸(V37Y5))/교(F9))
 Aux_V : (V(있/는(F1))
 N3P : (N3(것(MCC)A2))-(HUM:OBJ)
 VQ(2) : (V(보(V4W9))/았다(F1))

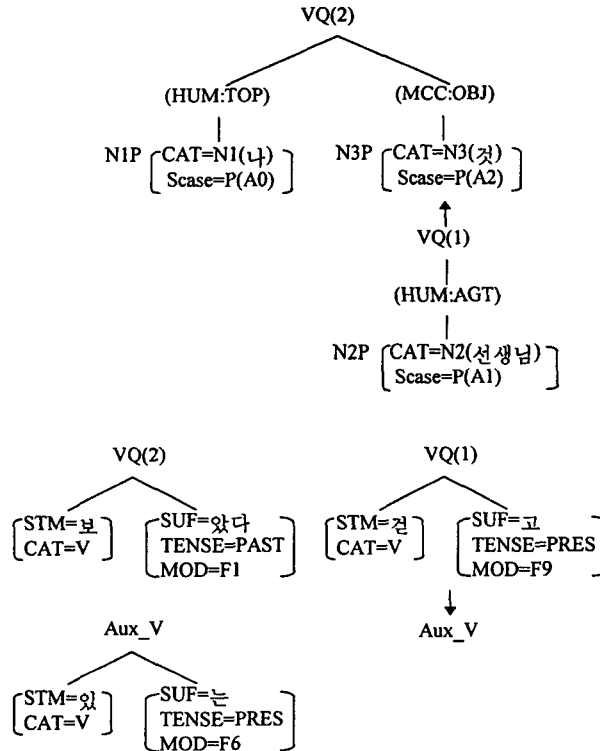


그림 4. 보문의 구문해석 결과