

피동 접사를 이용한 동사패턴의 확장

김창현^o 양성일 최승권
한국전자통신연구원
{chkim, siyang, choisk}@etri.re.kr

Extension of Verb Patterns Using Passive Affixes

Changhyun KIM^o Sungil YANG Sung-Kwon CHOI
NLP Research Team, Human Information Processing Dept., Electronics and
Telecommunications Research Institute

요 약

동사패턴은 원시 언어 분석을 위해 동사와 동사의 격성분 및 의미제약을 기술하고, 목적 언어 생성을 위해 동사의 대역어 및 격성분들의 생성 위치정보를 기술한다. 이러한 동사패턴의 구축은 시간적, 경제적 부담이 큰 작업이며, 동사패턴 구축의 자동화 혹은 반자동화에 대한 요구는 크다. 본 논문에서는 서술성 명사와 결합하여 동사를 생성하는 접사들인 ‘-하-, -되-, -받-, -당하-, -드리-’에 대해, 이들 간의 상호 변환 규칙을 이용하여 수동으로 구축된 동사패턴으로부터 새로운 동사패턴을 자동으로 생성한다. 변환 규칙에서는 명사 어휘별 접사 분포 정보와 함께, 접사와 결합된 파생동사의 구문정보가 요구된다. 그러나, 기존의 사전에는 서술성 명사들의 ‘-하다, -되다’ 분포 및 구문정보만이 기술되어 있고, ‘-받다, -당하다, -드리다’에 대해서는 기술되어 있지 않다. 본 논문에서는 서술성 명사들의 접사 분포 정보 및 구문정보를 파악하고, 이들 간의 상호 변환 규칙을 도출하여 새로운 동사패턴을 생성하는 2단계 작업을 수행한다.

1. 서론

동사패턴은 원시 언어 분석을 위해 동사와 동사의 격성분 및 의미제약을 기술하고, 목적 언어 생성을 위해 동사의 대역어 및 격성분들의 생성 위치정보를 기술한다([4]). 따라서, 현재 개발중인 한중 자동번역 시스템에서 동사패턴은 한국어 문장의 구조 파악과 중국어 대역어 선정에 핵심적인 역할을 수행한다([5]). 이러한 동사패턴의 구축은 시간적, 경제적 부담이 큰 작업이며, 따라서 동사패턴 구축의 자동화 혹은 반자동화에 대한 요구는 크다.

동사패턴의 (반)자동 구축에는 의미적 유사성을 이용한 방법([1],[2])과 의미적 동일성을 이용한 방법([3])이 있다. 의미적 유사성을 이용한 방법은, 의미적으로 유사한 동사들은 필수 논항에 있어서도 유사성을 가진다는 가정에 기반하고 있다. 동사의 의미적 유사성을 판단하기 위해서는 시소러스나 대역어 정보를 이용한다([3]). 명사에 대한 시소러스 구축이 잘 정비된 것과는 달리, 동사에 대한 시소러스는 정의가 용이하지 않으며, 대역어 사전에서의 대역어의 동일성을 이용하여 동사의 의미적 유사성을 판단하는 경우가 대부분이다.

의미적 동일성을 이용한 방법은, 기본 동사에 접사가 결합되어 파생되는 동사들을 대상으로 한 방법으로, 사동, 피동 변환 규칙을 이용한 사피동사 패턴의 생

성이 여기에 해당된다([5],[6]). 이들 규칙은 동사 파생 접사인 ‘-이-, -히-, -리-, -기-, -우-, -구-, -추-’와 보조 용언인 ‘-게 하다, -어지다’를 대상으로 하고 있다. 보조용언의 경우는 거의 모든 용언과의 결합이 가능하고 기본 동사의 형태 및 구문 특성이 유지되는 반면, 동사 파생 접사의 경우, 생산력이 낮아 일부 용언들과만 결합한다. [7]에서 조사한 피동 표현의 사용 빈도수를 살펴보면, ‘-되다, -받다, -당하다’의 쓰임이 전체 피동 발생 빈도수의 63%를 차지한다. 하지만, 이들에 대한 전산학적 연구는 아직 없다.

본 논문은 피동사 파생 접사인 ‘-하-, -드리-, -되-, -받-, -당하-’를 대상으로, 이들간의 상호 변환 규칙을 파악하여 동사패턴을 자동으로 생성한다. 명사 어휘별로 접사들에 대한 분포가 다르고, 파생된 동사들의 구문특성이 달라진다. 기존의 사전에는 서술성 명사들의 ‘-하-, -되-’에 대한 분포와 구문정보만이 기술되어 있다. 따라서, 명사 어휘별 ‘-받-, -당하-, -드리-’ 분포 및 구문정보를 말뭉치를 이용하여 파악하고,

2장에서는 본 연구에서 자동 생성하고자 하는 동사패턴을 살펴본다. 그리고, 파생접사를 이용한 기존의 피동화 연구 및 문제점을 살펴본다. 3장에서는 새로운 동사패턴 변환 규칙을 제시한다. 4장에서는 변환 규칙의 적용을 위해 필요한 명사 어휘별 접사 분포 정보 및 구문정보 획득에 대해 논한다. 5장에서는 실제 변

한 결과를 보이고 그 결과에 대해 논한다. 6장에서는 결론과 함께 향후 과제에 대해 논한다.

2. 관련연구

한중 자동번역에서 기술되고 있는 동사패턴은 “동사패턴번호 한국어동사패턴 > 중국어동사패턴”의 형태를 취한다. 동사패턴번호는 동사가 가지는 여러 동사패턴들에 대해 1번부터 순서대로 붙여지며, 의미별로 조직화되어있지는 않다. 다음의 예를 보자

타다16 A=건축물!가 불!에 타!다 > A 被!o:v [건물이 불에 났다]

타다32 A=사람!가 B=교통!에 타!다 > A 乘!v B [나는 배에 났다]

타다39 A=사람!가 B=재화!를 타!다 > A 得到!v B [나는 용돈을 댔다]

타다42 A=사람!가 자전거!를 타!다 > A 骑!v 自行车 [아이가 자전거를 탄다]

타다79 A=사람!가 B=약품!를 C=식품!에 타!다 > A 在 C 中放!v B [나는 약을 음식에 댔다]

한국어 패턴부에서 기능어와 실질어는 ‘!’ 기호로 구분되며, ‘=’기호의 왼쪽은 변수명을, 오른쪽은 명사 의미 코드¹⁾를 나타낸다. ‘=’ 기호가 없는 경우에는 어휘 자체를 의미한다. ‘타다16’의 경우, 주격명사는 ‘건축물’이라는 의미코드를 가지며, ‘불’은 ‘=’기호가 없으므로 어휘를 나타낸다. ‘타다42’에서도 ‘자전거’는 어휘 자체를 나타낸다. 중국어 패턴부에서, 동사는 ‘v’ 기호로 구분하고, 변수는 한국어 패턴부의 동일 변수에 대한 중국어 대역어를 의미한다. ‘패턴32’에서 주격명사 ‘A’에 해당하는 중국어 대역어는 동사 ‘乘’ 앞에 위치하고 ‘B’의 중국어 대역어는 동사 ‘乘’ 뒤에 위치함을 표시하고 있다. 각 동사패턴에는 실제 예제가 함께 기술되어 이해의 편의성을 도모하고 있다. 한중 동사패턴은 중국어 생성을 기준으로 하여 기술된다. 일례로, 패턴 ‘타다32’가 ‘타다42’를 포함하는 패턴이나²⁾, 중국어 대역부가 서로 다르므로, 별개의 동사패턴으로 기술된다. 만일 한국어 분석만을 위해서 동사패턴을 구축한다면, ‘타다42’와 같은 패턴은 제거될 것이다.

이와 같은 동사패턴을 이용한 기계번역에서는, 정확한 한국어의 분석이 곧 정확한 중국어 동사의 대역어 선택 및 올바른 생성정보의 제공으로 이어진다. 문장 ‘선생님이 버스를 타셨다’가 ‘타다32’에 의해 분석되면, ‘타다’에 대한 중국어 대역어 및 대역어 명사들의 생성 위치정보가 파악된다. 따라서 중국어 생성부에서는 개별 명사들에 대한 대역어 선택이 주요작업이 된다.

이러한 동사패턴은 한중 기계번역의 핵심 지식이며, 번역의 품질은 대량의 동사패턴 구축 여부에 달려있다. 그러나, 동사패턴의 구축은 많은 시간과 전문화된 인력을 필요로 하는 어려운 작업이며, 자동화에

대한 요구가 큰 부분이다. 피동표현의 대표적 형태인 ‘-이-, -히-, -리-, -기-’에 대해서는 많은 연구가 이루어졌으며, 표 1과 같은 능동문의 피동화 과정이 이루어진다([5],[6]).

표1. 능동형의 피동형 변환 규칙

능동사 주어표지	능동 -> 피동 격변환
유정성	주격 -> 부사격(-에 의해, -에게/-한테) 목적격 -> 주격
	엄마가 아기를 업는다 -> 아기가 엄마에게 업힌다
무정성	주격 -> 부사격(-로, -로부터, -에서, -에)
	먹구름이 온 마을을 뒤덮다 -> 먹구름에 온 마을이 뒤덮이다

능동문이 피동문으로 변경될 경우, 능동사는 타동사이고 피동사는 일반적으로 자동사가 된다. 그러나, 소수의 경우이긴 하지만 피동화된 동사가 여전히 타동사적 성격을 유지하기도 한다. ‘나라를 빼앗다’의 경우, 피동문은 ‘나라가 빼앗다’가 아닌 ‘나라를 빼앗기다’가 된다. 피동사의 타동성은 ‘-이-, -히-, -리-, -기-’의 경우 극히 적게 발생하지만, ‘-받-, -당하-’의 경우 빈번히 발생한다. 3장에서는 피동사의 타동성을 고려한 새로운 변환 규칙이 제시된다. 파생 동사들 간의 상호 변환은 어휘별 접사 분포에 따라 제약되지만, 어휘를 무시할 경우, 표 2와 같은 결과를 얻을 수 있다.

표2. 동사 파생 접사간 변환 가능 표

	하	되	당하	받	드리
하	X	O	O	O	O
되	O	X	O	O	O
당하	O	O	X	O	X
받	O	O	O	X	O
드리	O	O	X	O	X

거의 모든 경우에 대해 상호 변환이 가능하지만, ‘-당하-’와 ‘-드리-’ 간에는 변환이 불가능하다. 이는 의미적 상극에 기인한 것으로 보인다. ‘-당하-’의 경우, 주어가 원치 않는 행위가 진행되는데 반해, ‘-드리-’의 경우는 존경이나 존칭의 의도를 포함된다.

3. 동사패턴 변환 규칙

서술성 명사와 결합하여 동사를 파생시키는 접사인 ‘-하-, -되-, -받-, -당하-, -드리-’는 다음과 같이 분류할 수 있다.

능동화 접사	-하-, -드리-
피동화 접사	-되-, -당하-, -받-

동사패턴에는 이들 모두에 대한 패턴이 존재하므로,

¹⁾ 한중자동번역을 위해 동사패턴에 사용되는 위한 명사 의미코드는 약 200개이다.

²⁾ ‘자전거’의 명사코드는 ‘교통’이다.

변환규칙은 능동의 능동화, 능동의 피동화, 피동의 능동화, 피동의 피동화 등 모두 4가지 유형이 필요하다.

3.1 능동패턴의 능동화

능동화 접사 '-하-'와 '-드리-' 간의 동사패턴 변환에서는 구조적 변환은 발생하지 않고 단지 존칭의 의미만이 추가되는 것이므로, 접사만 교체하면 된다. 예를 들어, '아이를 부탁하다'의 경우 접사만 교체하면 '아이를 부탁드리다'가 된다.

3.2 능동패턴의 피동화

능동의 피동화는 피동사가 목적어를 가지는가의 여부에 따라 2가지로 나뉜다. '-되-'의 경우 피동사는 항상 자동사가 되며, '-당하-, -받-' 경우에는 어휘별로 자동사나 타동사가 가능하다.

능동화 접사 '-하-, -드리-'로부터 피동화 접사 '되-'로의 변환규칙은 다음과 같다.

능동 -하- -드리	피동 -되-	능동->피동
자	자	접사 변경
타	자	주격->부사격(에의해) 목적격->주격

능동 유형이 자동사인 경우에는 접사만이 교체되고 (예:되색) 타동사인 경우에는 기존의 연구와 같다.

'-받-, -당하-'의 경우, 능동 목적격이 동사의 유형에 따라 피동 주격으로 격 이동이 발생하기도 하고, 그대로 피동 목적격을 유지하기도 한다. 따라서, '-받-, -당하-' 피동으로의 올바른 변환을 위해서는 동사의 유형을 구분해야 한다. 이를 구분하는 기준은, 피동화된 동사의 유형이 자동사인가 타동사인가의 구분에 있다. 그러나, 이미 기술한 바와 같이 '-받-, -당하-' 파생 동사들은 사전에 기술이 되어있지 않기 때문에, 이를 알기 위해서는 사람이 직접 타동사 여부를 기술하거나, 본 논문에서처럼 말뭉치로부터 정보를 추출해야 한다

능동 -하 -드리	피동 -받 -당하	능동 -> 피동
자동	자동	접사 변경
타동	자동	주격->부사격(에게,에의해) 목적격->주격
	타동	주격->부사격(에게,에의해) 부사격(에게)->주격

자동능동사(예:실직)인 경우는, 앞서와 마찬가지로 접사 교체만 수행하면 된다. 타동능동사인 경우는 자동피동사나 타동피동사로 변경이 가능하다. 자동피동사의 경우는 '-되-'의 경우와 동일하다. 타동피동사인

경우, 목적격은 그대로 유지되고 부사격(에게)이 주격으로 변경된다. 주격이 부사격으로 변경될 때의 부사격 조사는, 명사의 의미특성과 동사의 의미에 따라 '-에게,-에,-에게서,-에 의해,-로부터' 등으로 다양한 형태가 가능하다.

3.3 피동패턴의 피동화

'-받-, -당하-' 간의 상호 변경은 접사만 변경하면 된다. '-되-'와 '-받-, -당하-' 상호간의 변경은, '-되-'의 능동화 후 다시 피동화를 거치면 된다.

피동1 -받 -당하	피동2 -되	피동1 -> 피동2
자	자	접사 변경
타	자	주격->부사격(에게) 목적격->주격 부사격(에게) -> 부사격(-에의해)

3.4 피동패턴의 능동화

피동패턴의 능동화는, 앞서 기술된 능동패턴의 피동화 규칙을 역방향으로 적용하면 된다.

4. 접사 분포의 파악

동사패턴 간의 변환을 수행하기 위해서는 1)명사 어휘들의 접사 분포와 2) 구문정보가 필요하다. 서술성 명사들이 '-하-, -드리-, -되-, -받-, -당하-'에 대해 서로 다른 분포를 보이기 때문에, 개별 명사 어휘들의 접사분포를 파악해야 패턴간 변화를 적용할 수 있다. 예를 들어, '강탈'의 경우, '강탈하다, 강탈되다, 강탈당하다'는 가능하지만 '강탈드리다, 강탈받다'는 불가능하다. 일반 사전에는 '-되-, -하-'에 대한 정보만이 기술되어 있기 때문에, 본 연구에서는 말뭉치의 형태소 분석을 통해 명사 어휘들의 접사 분포를 획득하였다.

1. '명사+접사'로 분석된 모든 어절에 대해 (명사,접사) 쌍을 추출한다
2. '-를 {하,당하,받,드리}' 패턴에 대해 (명사,접사) 쌍을 추출한다. 명사가 한글자인 것은 제외한다.
3. '-이/가 {되}' 패턴에 대해 (명사,되다) 쌍을 추출한다. 명사가 한글자인 것은 역시 제외한다.

명사가 한글자인 것을 제외한 것은, '상을 당하다'에서 '상당하다'와 같은 것이 추출되는 것을 방지하기 위해서이다. 이렇게 해서 추출된 모든 (명사,접사) 쌍을 이용하여, 개별 명사들에 대한 '접사분포표'를 작성한다. 이때, 복합명사와 점두사를 고려한다. 복합명사 '전면+개방'의 경우, '개방'에 대한 접사분포가 존재하면 '개방'에 대한 접사분포 항목에 빈도수를 증가시키고, 없는 경우에만 '전면개방'에 대한 접사분포를 구성한다. 또한 '가+계약'과 같은 경우 접사 '가'

를 제외하고 '계약'에 대한 접사분포를 파악한다. 완성된 접사분포에서 빈도수 2 이상인 분포만을 이용한다.

접사분포표로부터 동사패턴 생성을 위한 대상을 선정하기 위해서 다음과 같은 절차를 거친다.

1. 2가지 이상의 서로 다른 접사를 취하는 명사들을 동사패턴 생성을 위한 후보로 기록한다. 따라서, '과잉되다'와 같이 하나의 접사와만 결합이 가능한 형태는 제외된다.
2. 동사패턴 생성 후보들에 대해 '-되-'나 '-하-'를 가지지 않은 경우는 후보에서 제외한다.
3. '-되-'나 '-하-' 형태가 사전에 등록되어 있는 것만을 최종 후보로 선정한다.

최종적으로 선정된 서술성 명사는 모두 4020개이며, 2개 이상의 접사와 결합하는 명사들의 수는 다음과 같다.

	2개	3개	4개	5개	계
개수	3296	638	84	2	4020

5. 실험

동사패턴으로부터 추출한 접사분포를 앞서 말문치로부터 추출한 접사 분포와 비교하면 다음과 같다.

	-하-	-되-	-당하-	-받-	-드리-
말문치	3995	3804	432	577	44
동사패턴	2641	900	30	61	9
누락율	.34	.76	.93	.89	.80

누락율은 말문치로부터 추출한 접사분포가 정확하다고 가정했을 경우, 동사패턴이 기술하지 않고 있는 접사패턴의 비율이다. '-하-'와 '-되-'의 발생 빈도가 말문치에서는 거의 비슷하나 동사패턴에서는 많은 차이를 보인다. 이것은, 실제 문장에서는 '-하-' 형태가 '-되-' 형태보다 많이 쓰이기 때문인 것으로 보인다. 자동 생성된 동사패턴 수에서도 '-되-'로부터 생성된 동사패턴이 가장 많았다. 자동 생성된 동사패턴의 정확도는 다음과 같다.

	정확	수정필요	오류
비율	36%	41%	23%

수정이 필요한 경우는 대부분 주어가 없는 경우이며, '사람'이나 '조직'을 주어로 생성시킬 수 있는 동사패턴들이 대부분이었다. '-되-'로부터 '-하-'로 생성되는 동사패턴들이 대부분을 차지한다. 오류 유형을 크게 분류해 보면 다음과 같다.

- 자타동사 판별 오류
- 부사격 조사 오류
- 예외적인 경우

자타동사 판별 오류는 생성 규칙의 오적용을 발생시킨다. 부사격 조사의 경우 어휘별로 차이를 보이기 때문에 좀 더 많은 자료의 분석이 필요하다. 기타 어휘별로 변형규칙이 적용되지 않는 경우들이 있다.

6. 결론

본 연구에서는 빈도수를 고려한 좀더 정확한 접사 분포의 파악과 동사의 의미 구분 등에서 부족한 점이 보인다. 더 많은 자료 확보를 통해 앞서 기술된 문제점을 해결하고, 피동접사뿐 아니라 사동접사에 관해서도 연구를 수행할 필요가 있다.

7. 참고 문헌

- [1]. Baldwin & Bond (2002): Alternation-based Lexicon Reconstruction in TMI 2002.
- [2]. Baldwin & Tanaka (2002): Verb Alternation and Japanese - How, What and Where?, in PACLIC 2000.
- [3]. Fujita & Bond (2002): A method of Adding New Entries to a Valency Dictionary by Exploiting Existing Lexical Resources, in TMI 2002.
- [4]. Kim, Hong, Huang, Kim, Yang, Seo & Choi (2002): Korean-Chinese Machine Translation Based on Verb Patterns, to appear in the Proceedings of AMTA 2002.
- [5]. Yang, Hong, Kim, Kim, Seo & Choi(2002): An Application of Verb-Phrase Patterns to Causative/Passive Clause, in IASTED2002.
- [6]. 이익섭 & 임홍배(1983): 국어문법론, 학연사
- [7]. 이정택(2001): 피동성 표현에 관한 연구-되다, 받다,당하다-를 대상으로, 한글 제251호.