

영한 기계번역의 한국어 생성 시스템에서 조동사의 생성

안동언, 조정미, 김길창
한국과학기술원 전산학과

Generation of Auxiliary Verbs in a Korean Generation System of English-Korean Machine Translation

Dong Un An, Jeong Mi Cho, Gil Chang Kim
Department of Computer Science, KAIST

요 약

기계번역에서 기능어의 번역은 각 언어가 고유의 문법적 특성과 표현방법을 가지고 있기 때문에 기능어가 가지는 의미와 문법적 기능을 자질로 표현하고 이를 통하여 해석과 생성을 한다. 한국어는 첨가어로 기능어가 매우 발달되어 있으며, 특히 조동사가 많은 문법적 기능을 담당하고 화자의 양태적 관계를 나타낸다. 따라서, 기계번역의 한국어 생성에서 조동사가 번역의 질에 상당한 영향을 미친다.

본 논문에서는 중간언어방식을 이용한 기계번역시스템의 한국어 생성에서 문법적 기능어나 양태적 관계를 나타낸 중간언어로부터 어떠한 조동사가 생성이 되어야 하고, 또한 조동사들간의 결합순서는 어떠한지를 살펴본다. 이를 위하여 Corpus를 분석하여 각 조동사 분류에 대한 대표 표층어를 선정하고 조동사간의 결합순서를 밝힌다. 조동사 생성을 위해 필요한 처리 과정과 언어적 지식베이스를 제안하고 시스템을 구현한다.

I. 서론

기계번역시스템에서 구체적인 대상이나 동작을 나타내는 내용어(content word)는 대역사전에서 단어를 선택하여 번역을 한다. 그러나 문법적 기능을 나타내는 기능어는(function word)는 각 언어가 고유의 문법적 특성과 표현방법을 가지고 있기 때문에 의미자질(semantic feature), 구문자질(syntactic feature), 어순정보(word order) 등에 의해서 표현이 되고 이를 통하여 기능어를 해석하고 생성한다.

한국어는 첨가어(agglutinative language)이므로 조사, 어미, 조동사 등의 기능이 대단히 발달되어 있다. 여러 기능어 중에서 조동사는 동사와 결합하여 많은 문법적 기능을 담당하고 화자의 양태적 관계를 나타낸다[2][4]. 이 문법적 기능은 문장 전체에 큰 영향을 미치며 화자의 의미 전달에 조동사가 중요한 역할을 한다. 따라서, 기계번역시스템의 한국어 처리에서 조동사의 정확한 해석과 생성이 번역의 질에 상당한 영향을 미친다.

본 논문에서는 영한 기계번역 시스템 PIVOT E/K에서 한국어 조동사를 생성하는데 있어서 필요한 언어 지식(linguistic knowledge)과 생성 처리(generation processes)를 살펴보기로 한다. PIVOT E/K는 중간언어방식(interlingua strategy)을 채택한 기계번역 시스템으로 PIVOT E/J에 기반을 두고 있다[10].

이 시스템에서 조동사가 나타내는 문법적 기능과 양태적 관계는 특정 언어와 독립된 중간언어로 표현되어 보조 개념 CP(Conceptual Primitive)로 주어진다[11]. 이 CP로부터 하나의 적당한 한국어 조동사를 생성하기 위해서는 우선 유사한 의미와 기능의 조동사 중에서 하나의 대표 표층어를 선택하여야 한다. 또한, 둘 이상의 조동사들이 있을 때 일정한 순서로 결합을 하여야 한다. 이러한 언어적 지식을 얻기 위하여 세가지 종류의 Corpus를 분석한다. 분석된 자료로부터 의미기능에 따라 조동사를 분류하고 각 분류에 대한 대표 표층어를 선정하고 결합 순서를 밝힌다. 이러한 언어적 지식을 이용하여 시스템을 구성하고 한국어 조동사 생성을 위한 처리 과정을 제안한다.

II. 관련 연구

한국어에서는 조동사가 매우 발달되어 있어서 다양한 의미와 결합을 보이고 있고 문장에서 많은 문법적 기능을 담당하는 중요한 역할을 한다. 따라서, 국어학에서 조동사에 관한 많은 연구가 있어 왔다[1][2][3][5]. 이러한 연구들은 분석적 입장에서 가능한 모든 조동사를 제시하고 결합 가능성과 배열을 보여준다[3][5]. 그런데, 조동사를 생성하는 입장에서는 같은 의미와 기능을 가진 여러 표층어 중에서 하나만을 생성하여야 하고 가능한 결합 순서중에서 하나만을 선택하여야 한다. 그러나, 이러한 점에 대해서 국어학에서는 많은 정보를 제공하고 있지는 않다. 조동사의 분류는 학자들마다 다르지만 명칭은 공통적으로 조동사의 의미 기능을 나타내고 있다[3]. 본 논문에서는 Corpus에서 얻어진 조동사를 의미기능에 따라세부적으로 분류한다.

한국어를 생성하는 기계번역 시스템으로는 ATLAS J/K, MATES E/K, PIVOT E/K 등이 있다[6][7][8][10]. ATLAS J/K와 MATES E/K는 변환방식의 기계번역시스템으로 영어와 일어의 조동사를 일대일로 번역하여 생성을 하며 한국어 조동사에 관한 전반적인 연구 및 처리는 미흡하다. PIVOT E/K의 한국어 생성시스템은 일본어 생성 시스템에 기반을 두고 있다. 이는 한국어와 일본어의 문법적 유사성에 착안하여 일본어 생성시스템의 언

어적 지식과 개발 환경을 이용한 것이다. 문법을 그대로 이용하고 실질어를 일대일 번역을 하여 초기 시스템을 구축하였으나 기능어에서는 많은 문제점이 노출되었다(10). 조동사에 있어서 한국어 조동사 고유의 활용, 결합순서, 형태가 있음을 다시 확인하는 계기가 되었다.

III. Corpus 분석을 통한 한국어 조동사의 분류

한국어 조동사는 각기 특정한 문법적 기능을 나타내며 동사 어간에 일정한 순서로 첨가되면서 동사의 의미를 변화시키거나 보완하면서 동사를 이룬다. 조동사가 나타내는 문법적 기능은 태(voice), 상(aspect), 양상(modality), 부정(negative) 등이다. 각 기능을 가지는 조동사에는 어떠한 것이 있는지 Corpus를 분석하여 알아본다. 같은 의미와 기능을 가지는 조동사들을 분류하고 각 분류에 대하여 출현 빈도에 따라 대표 표층어를 선정한다. 또한, 분류들간의 결합관계를 조사하여 전체 조동사의 결합 순서를 결정한다.

3.1 Corpus의 선정

다양한 종류의 조동사를 얻기 위해서는 Corpus가 의미적으로나 구문적으로 아주 다양한 문장을 가지고 있어야 한다. 다음의 세가지 종류의 Corpus는 어느 정도 위의 전제를 만족한다고 본다.

Corpus 1: 국민학교 각종 교과서 73권

Corpus 2: 고등학교 국어 교과서 2권

Corpus 3: 고등학교 문학 교과서 2권

3.2 조동사의 분포

표1에서 보듯이 세가지 Corpus에서 나타난 조동사가 포함된 문장은 23,544개이다. 하나의 문장에 나타나는 조동사의 수효로 본 문장의 분포를 보면 1개 또는 2개의 조동사를 포함한 문장의 수효는 99% 이상으로 3개 이상의 조동사는 잘 나타나지 않는다는 것을 알 수 있다. 따라서, 조동사간의 결합순서를 분석할 때, 2개의 조동사간의 결합을 분석한다.

Corpus에서 얻어진 조동사들을 의미 기능에 따라 34종류로 분류하였다. 표2는 조동사의 분류별 분포를 보여준다. "시행 조동사"가 30% 이상 많이 나타난 것은 학생들에게 교과서의 내용을 따라서 해 보도록 하는 국민학교 교과서의 특성 때문이다.

3.3 조동사의 결합 순서

조동사의 결합 순서는 2개의 조동사를 포함하는 용언을 분석하여 얻는다. 표3은 두

표1. 조동사 수효에 따른 문장의 분포

	조동사 포함 문장		조동사 1개		조동사 2개		조동사 3개		조동사 4개	
	빈도수	%	빈도수	%	빈도수	%	빈도수	%	빈도수	%
전체	23,544	100%	21,010	89.2%	2,327	9.9%	194	0.8%	13	0.1%
Corpus 1	19,908	100%	18,165	91.2%	1,623	8.2%	114	0.6%	6	0.0%
Corpus 2	1,087	100%	836	76.9%	222	20.4%	26	2.4%	3	0.3%
Corpus 3	2,549	100%	2,009	78.8%	482	18.9%	54	2.1%	4	0.2%

표2. 조동사의 분류별 분포

조동사	전체			Corpus1			Corpus2			Corpus3		
	빈도수	%	순위	빈도수	%	순위	빈도수	%	순위	빈도수	%	순위
시행	8,449	32.1%	1	8,312	38.2%	1	51	3.7%	10	86	2.7%	14
진행	2,656	10.1%	2	2,227	10.2%	2	106	7.7%	5	323	10.3%	3
피동	1,666	6.3%	3	1,078	5.0%	5	173	12.6%	2	415	13.2%	1
가능	1,588	6.0%	4	1,141	5.2%	3	189	13.8%	1	258	8.2%	5
부경	1,536	5.8%	5	1,119	5.1%	4	95	6.9%	6	322	10.2%	4
재귀	1,309	5.0%	6	1,031	4.7%	6	89	6.5%	7	189	6.0%	6
규정	1,214	4.6%	7	734	3.4%	9	147	10.7%	3	333	10.6%	2
봉사	1,167	4.4%	8	986	4.5%	7	26	1.9%	15	155	4.9%	7
담위	1,104	4.2%	9	887	4.1%	8	142	10.4%	4	75	2.4%	15
완료	891	3.4%	10	722	3.3%	10	34	2.5%	13	135	4.3%	10
시인	729	2.8%	11	544	2.5%	11	30	2.2%	14	155	4.9%	7
예정	700	2.7%	12	542	2.5%	12	68	5.0%	8	90	2.9%	12
사동	683	2.6%	13	496	2.3%	13	41	3.0%	11	146	4.6%	9
불가능	511	1.9%	14	366	1.7%	14	57	4.2%	9	88	2.8%	13
종료	451	1.7%	15	334	1.5%	15	30	2.2%	16	97	3.1%	11
원인	411	1.6%	16	301	1.4%	16	36	2.6%	12	74	2.3%	16
주속	271	1.0%	17	199	0.9%	18	20	1.5%	16	52	1.7%	17
희망	233	0.9%	18	217	1.0%	17	3	0.2%	23	13	0.4%	22
의도	192	0.7%	19	142	0.7%	20	17	1.2%	18	33	1.0%	18
보유	230	0.9%	20	193	0.9%	19	6	0.4%	9	31	1.0%	19
일념	78	0.3%	21	48	0.2%	21	4	0.3%	21	26	0.8%	20
가합	48	0.2%	22	36	0.2%	22	4	0.3%	21	8	0.3%	24
불가피	44	0.2%	23	24	0.1%	24	2	0.1%	25	18	0.6%	21
불허	39	0.1%	24	26	0.1%	23	1	0.1%	26	12	0.4%	23
승관	21	0.1%	25	19	0.1%	25	1	0.1%	26	1	0.0%	31
경향	16	0.1%	26	16	0.1%	26	0	0.0%	28	0	0.0%	32
요령	14	0.1%	27	11	0.1%	27	0	0.0%	28	3	0.1%	25
시발	11	0.0%	28	11	0.1%	27	0	0.0%	28	0	0.0%	32
가치	10	0.0%	29	3	0.0%	30	5	0.4%	20	2	0.1%	29
가식	8	0.0%	30	5	0.0%	29	0	0.0%	28	3	0.1%	25
극한	6	0.0%	31	3	0.0%	30	0	0.0%	28	3	0.1%	25
처지	5	0.0%	32	2	0.0%	32	3	0.2%	23	0	0.0%	32
강세	4	0.0%	33	2	0.0%	32	0	0.0%	28	2	0.1%	29
파신	3	0.0%	34	0	0.0%	34	0	0.0%	28	3	0.1%	25
총빈도수	26,298	100%		21,777	100%		1,370	100%		3,151	100%	

표3. 두 조동사간의 결합분포

선행 조동사	후행 조동사(결합 빈도수)	종류	빈도수	%
피동	피동(7), 사동(2), 종료(13), 진행(127), 완료(175), 의도(1), 요행(1), 가능(34), 추측(10), 부정(75), 불가피(1), 불가능(5), 재귀(33), 시인(36), 당위(43), 규정(60), 원인(12), 예정(14), 일념(2), 확산(1)	20	652	28.0%
사동	종료(1), 진행(44), 봉사(41), 시행(35), 의도(7), 가능(17), 부정(15), 불가피(1), 불가능(3), 가합(2), 재귀(4), 시인(16), 당위(34), 규정(20), 원인(1), 예정(7)	16	248	10.7%
시발	진행(1)	1	1	0.0%
종료	종료(7), 진행(13), 시행(1), 의도(1), 가능(15), 추측(1), 부정(5), 불가능(7), 시인(7), 당위(3), 규정(18), 예정(11)	12	89	3.8%
진행	종료(2), 진행(54), 가능(6), 추측(9), 부정(14), 습관(1), 시인(4), 당위(7), 규정(54), 원인(25), 예정(6), 일념(5)	12	187	8.0%
완료	추측(6), 부정(8), 불가피(1), 불가능(1), 재귀(4), 시인(3), 당위(4), 규정(10), 원인(10), 예정(5), 일념(1)	11	53	2.3%
봉사	진행(51), 시행(1), 의도(3), 가능(14), 추측(2), 불가능(5), 부정(15), 습관(1), 재귀(2), 시인(36), 당위(29), 규정(26), 원인(2), 예정(20), 일념(3)	15	210	11.4%
시행	진행(1), 의도(3), 가치(1), 가능(28), 부정(19), 불가피(1), 재귀(2), 시인(3), 당위(17), 규정(3), 예정(2)	11	80	3.4%
희망	진행(1), 시행(1), 추측(1), 부정(11), 시인(1), 원인(8), 일념(1)	7	24	1.0%
의도	진행(2), 부정(13), 규정(7), 원인(2), 예정(1)	5	25	1.1%
요행		0	0	0.0%
가치	부정(1)	1	1	0.0%
가식	종료(1), 가능(1)	2	2	0.1%
가능	사동(13), 추측(8), 부정(6), 재귀(78), 당위(13), 규정(35), 원인(21), 예정(52), 일념(1),	9	227	9.8%
추측	부정(6), 시인(1), 원인(2)	3	9	0.4%
부정	사동(21), 종료(3), 진행(20), 봉사(1), 의도(2), 가능(3), 추측(4), 불가능(32), 불허(12), 가합(2), 습관(1), 재귀(15), 당위(21), 규정(16), 원인(13), 예정(37), 경향(2)	17	205	8.8%
불가피	원인(1)	1	1	0.0%
불가능	추측(2), 부정(2), 재귀(15), 규정(13), 원인(9), 예정(9), 극한(1)	7	51	2.2%
불허	규정(1)	1	1	0.0%
가합	부정(1), 예정(1)	2	2	0.1%
습관	불가능(1)	1	1	0.0%
재귀	종료(1), 진행(1), 완료(11), 가능(2), 부정(2), 시인(2), 규정(46), 원인(2), 예정(41)	9	108	4.6%
시인	당위(1), 규정(2)	2	3	0.1%
당위	부정(5), 규정(9), 원인(5), 예정(50)	4	69	3.0%
규정	시인(1), 원인(4), 예정(1), 일념(1)	4	7	0.3%
원인	시인(1), 예정(3)	2	4	0.2%
예정	원인(1)	1	1	0.0%
일념		0	0	0.0%
경향		0	0	0.0%
극한		0	0	0.0%
확산		0	0	0.0%

조동사간의 결합 분포를 보여 준다. 앞에 있는 조동사가 선행 조동사이며 뒤에는 후행 조동사들과 각 후행 조동사들간의 결합 빈도수를 보여준다. 가능한 결합쌍 1165개(34x34)중에서 실제로 나타난 결합의 종류는 176개로 약 15% 정도 된다. 이것으로부터 결합에 관한 precedence matrix를 작성할 수 있다. 이때, 2개의 조동사가 두가지의 결합순서를 모두 보일 경우에는 빈도수가 많은 쪽을 선택한다. 그러나, 이런 경우는 5쌍에 지나지 않았다. 생성을 위해서 2개의 조동사가 주어졌을 때, 이들 조동사 사이에 해당하는 결합순서가 있으면 이 matrix로부터 그 결합순서를 결정할 수 있지만, 결합순서가 matrix에 없으면 결합순서를 결정할 수 있는 방안을 마련하여야 한다. 이를 위하여 전체 조동사들간의 결합순서를 구해본다.

조동사들간의 명백한 linear order는 없지만, 이 precedence matrix로 부터 최선의 결합순서는 결정할 수 있다. 표4는 전체 조동사들간의 결합순서를 보여준다. 이 결합순서는 precedence matrix와 비교해 볼 때 약 14개의 오류만을 보여준다. 문법적 기능에 따른 분류에 의하면 태, 상, 양상, 부정, 양상의 순서로 조동사가 결합을 한다. 이것은 (1)과 (2)에서도 제시하였던 것과 같다. 다만, 부정 조동사와 결합에 있어서 일부분의 양상 조동사는 뒤에 결합을 한다. 또한 각 문법적 기능에 따른 분류안에서도 결합순서를 결정하였다.

3.4 조동사의 대표 표층어 선정

조동사의 의미 기능에 따라 34종류로 분류를 하였다. 하나의 분류에는 유사한 의미를 가지는 여러 표층어가 있을 수 있으며 이때에는 생성에서 하나의 표층어만을 생성하여야 한다. Corpus의 분석 자료로부터 하나의 대표 표층어를 선택하기 위해서 넓은 결합 분포를 보이는 사용빈도가 높은 표층어를 선택한다.

표5에서는 일부 조동사의 표층어 빈도수를 보여주고 있다. 위의 선택 기준에 따라 각 조동사 분류에 대한 대표 표층어를 선정하였으며 표4에서 조동사 분류에 대한 대표 표층어를 보여주고 있다.

3.5 한국어 조동사와 보조개념 CP와의 대응

한국어 조동사는 문법적 기능이나 의미에서 영어의 조동사보다 범위가 훨씬 크다. 한국어 조동사는 영어의 조동사뿐만 아니라 동사, 부사, 접속사 등과 대응이 되고 태와 상은 표현방법이 다르다. 따라서 기계번역에서 한국어 조동사를 처리하기 위해서는 문법적 기능과 의미를 나타내는 자질을 설정하고 이를 통하여 해석과 번역이 이루어져야 한다.

본 시스템인 PIVOT E/K에서는 이러한 역할을 보조 개념 CP(Conceptual Primitive)가 담당한다. 보조 개념 CP는 자립 개념 CP와 관계 개념 CP와 더불어 중간언어를 이루고 있는 것으로 양상과 시상과 보조상 개념으로 구성되어 있다(11). 이 보조개념 CP

표4. 조동사의 결합 순서, 대표 표층어, 보조개념 CP

종류	순서	조동사	대표 표층어	보조개념 CP	영어 표층어
태	1	피동	아,히,리,기,받,당하,어지,되	XPASS	be ~ed
	2	사동	이,히,리,기,우,시키,게 하	XCAUS	make, have, get
상	3	시발	어 들	XBEGIN	begin, start
	4	종료	어 버리	XFINISH	finish
	5	보유	어 놓		keep
	6	진행	고 있	PROG	be ~ing
	7	완료	어 있	PREF	have ~ed
양상	8	봉사	어 주		do something for
	9	시행	어 보	XTRY	try
	10	희망	고 싶	XWISH	wish, want, hope
	11	강세	어 태		heavily, hard
	12	외도	려고 하	XPLAN	intend, plan
	13	요행	ㄹ 편하		come near
	14	가치	ㄹ 만하	XESTIM	be worth of
	15	가식	ㄴ 척하		pretend
	16	가능	ㄹ 수 있	XPOSS	can, be able to
	17	추측	ㄴ 것 같	XSEEM	seem as if
부정	18	부정	지 않, 지 못하, 지 말	NEG	not
양상	19	불가피	ㄹ 수밖에 없	XNECE	cannot help ~ing
	20	불가능	ㄹ 수 없	NEG^XPOSS	cannot
	21	불허	면 안 되	NEG^XMUST	must not
	22	가합	면 되	XRECO	had better
	23	습관	곤 하	XRULE	used to
	24	재귀	게 되	XCHAN	came to, become, get
	25	시인	기도 하		really
	26	당위	어야 하	XNEED	must, have to
	27	규정	ㄴ 것이	XCONCLU	
	28	원인	기 때문에		because
	29	예정	ㄹ 것이	XINFE	will, may, might
	30	일념	ㄹ 뿐이	XONLY	only
	31	경향	ㄴ 편이	XTEND	tend to, be apt to
	32	극한	ㄹ 지경이		
	33	치지	ㄴ 터이		
	34	확신	기 마련이		certain

표5. 조동사의 표층어 빈도수(부분)

분류	표층어	전체	Corpus 1	Corpus 2	Corpus 3
추측	ㄴ 것 같	200	164	10	26
	ㄴ 듯 하	68	32	10	26
	ㄴ 가보	3	3	0	0
예정	ㄹ 것이	601	459	66	76
	ㄹ 모양이	8	8	0	0
	ㄹ 터	91	75	2	14

는 매우 세분이 되어 있어서, 한국어 조동사와 대응이 잘 되며 표4에서 그 대응관계를 보여준다. 이 보조 개념 CP는 영어와 일어를 중심으로 개발이 되었기 때문에 한국어 조동사와 정확히 일치하지는 않는다. 표4에서 보조 개념 CP가 공백이거나 영어 표충어가 공백인 것은 적절한 대응관계가 없는 것을 나타내며 이를 번역하기 위해서는 일종의 Paraphrasing이 필요하다.

IV. 한국어 조동사의 생성

한국어 생성 시스템은 중간언어로 구성된 개념구조(Conceptual Structure)로부터 한국어 문장을 생성한다. 생성단계는 다음과 같이 3단계로 나누어 진다(10)

- 1단계(문체 생성) 개념구조의 각 노드의 개념에 해당하는 사전 정보와 개념구조를 분석하여 대상문의 생성 계획을 수립하고 문체를 결정한다. 언어 독립적인 개념구조를 대상 언어의 의미 구조로 변환한다.
- 2단계(구문 생성) 의미 구조의 각 노드에 구문 정보를 결정한다. 또한 격, 양상, 태 등의 형태소 자질과 어순 정보를 생성한다. 의미 구조를 구문 구조로 변환한다.
- 3단계(형태소 생성) 어순 정보를 이용하여 구문 구조의 각 노드를 배열한다. 각 노드로부터 실질어와 기능어의 표충어를 생성하고 이 표충어를 결합하여 단어를 생성하여 문장을 완성한다.

그림 1은 한국어 조동사를 생성하기 위해서 필요한 처리과정과 지식베이스(Knowledge Base)를 보여준다.

4.1 지식베이스

조동사 결합 Corpus

Corpus에서 조사된 모든 조동사 결합을 가지고 있다. 이 결합이 해당 조동사들의 결합 순서를 보여 준다.

조동사 결합 순서 테이블

조동사 결합 Corpus에 없는 조동사 집합에 대해 결합 순서를 결정하기 위해서 사용한다. 이 테이블의 조동사 순서는 결합 precedence matrix에서 얻어진 것으로 표4에서 볼 수 있다.

형태소 사전

조동사의 보조 개념 CP에 대한 표충어를 가지고 있다. 세분되거나 결합된 조동사의

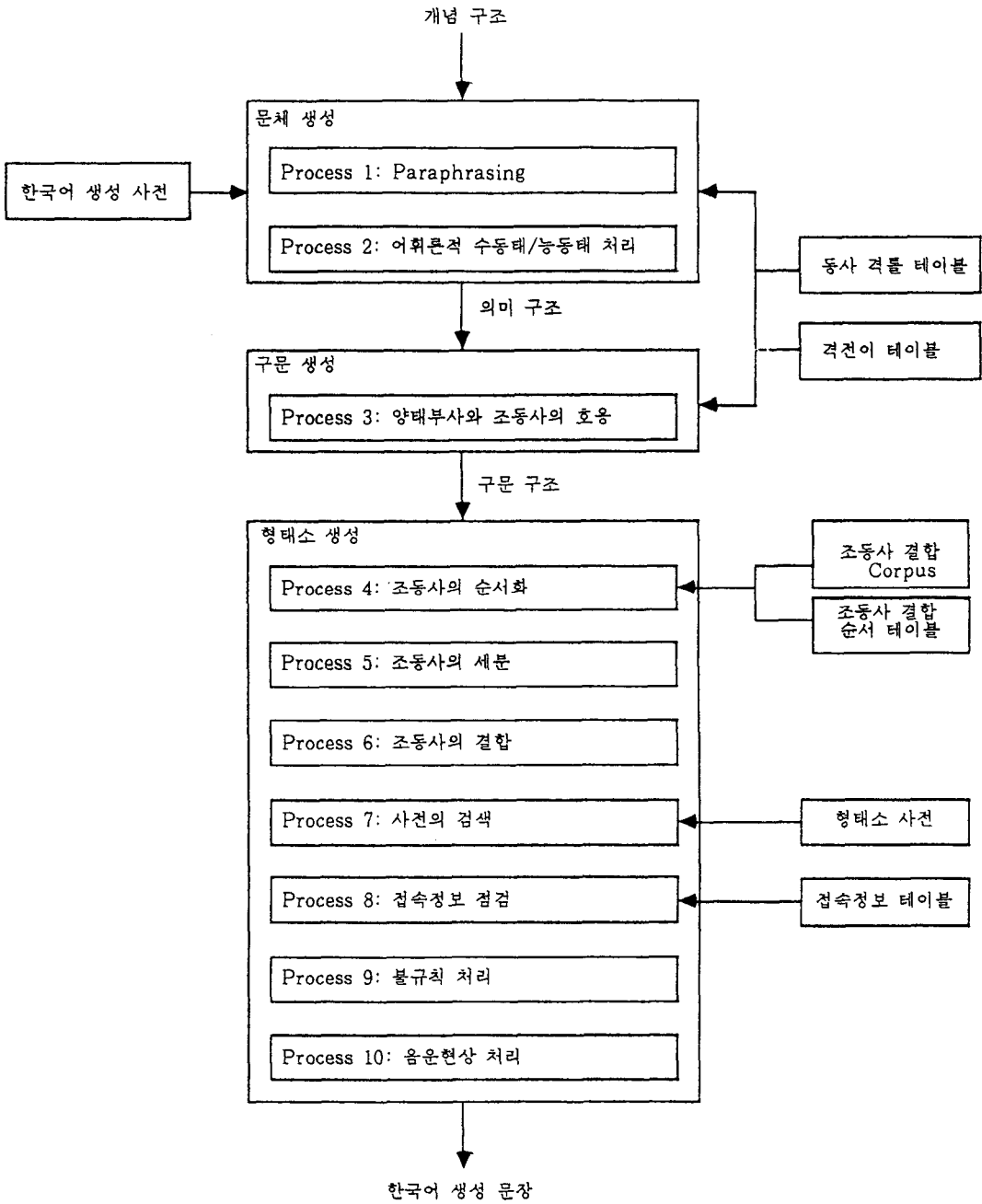


그림 1. 조동사 생성을 위한 처리 과정 및 지식 베이스

표층어가 있으며, 또한 이형태도 가지고 있다.

접속 정보 테이블

이형태의 여러 표층어 중에서 하나의 표층어를 선택하도록 한다. 이때, 필요한 정보는 선행하는 표층어의 가장 뒤 글자가 유종성인가 무종성인가, 혹은 양성모음인가 음성모음인가로, 이 정보에 따라 이형태중에서 하나를 선택한다.

4.2 처리 과정

Process 1: Paraphrasing

대응되는 보조개념 CP가 없는 한국어 고유의 조동사를 생성하기 위하여 개념 구조의 변경, 품사의 변경, 조동사의 추가 등을 행한다. 해당하는 영어의 표층어가 동사, 부사, 접속사 등이므로 각각의 처리는 다르다. “보유”, “봉사”, “강제”, “원인” 등의 조동사를 처리한다.

Process 2: 어휘론적 수동태/능동태 처리

어휘론적 수동태/능동태는 조동사에 의해서 수동태/능동태를 취하는 것이 아니라 대응되는 수동/능동의 의미를 가지고 있는 단어가 그 기능을 한다[9]. 따라서, 문장이 수동태/능동태를 요구하지만 해당 동사가 어휘론적 수동태/능동태를 취하면 사전에서 새로운 단어를 가져온다. 이 정보는 사전에 #KPASS라는 항목으로 필요로 하는 수동태/능동태 단어가 저장되어 있다.

Process 3: 양태부사와 조동사의 호응

일부 양태 부사는 특정한 조동사와 호응을 이루므로 필요한 조동사가 없으면 추가를 한다. 이 정보는 사전에 #REQUAUX라는 항목으로 저장되어 있다.

Process 4: 조동사의 순서화

우선 주어진 조동사가 조동사 결합 Corpus에 있으면 이에 따라 순서화를 하고 없으면 조동사 순서 테이블을 이용하여 순서화를 행한다.

Process 5: 조동사의 세분

수동태/능동태의 경우는 동사마다 필요로 하는 조동사가 다르다. 따라서, 생성을 위해서는 수동태/능동태 조동사를 세분한다. 이 정보는 #KPASS라는 항목으로 해당 동사에 저장되어 있다.

문장의 중결이 높임일때 조동사도 존칭어가 필요하다. “봉사”, “진행”, “완료” 조동사에서 이의 처리가 필요하다.

Process 6: 조동사의 결합

몇몇 조동사는 부정과 결합을 하면 부정을 포함하고 있는 표층어로 바뀐다. 특히, 조동사의 표층어에 존재를 나타내는 "있"이 있는 경우에 부정 조동사와 결합을 하면 부정 조동사는 없어지고 "없"이 된다.

또한, 관형형 어미를 포함하고 있는 조동사중에서 일부분은 관형형어미의 변화로 시제를 반영하므로 조동사에 시제 정보를 결합시킨다.

Process 7: 사전의 검색

형태소 사전에서 조동사의 표층어를 가지고 온다. 이 표층어는 이형태를 가지고 있다.

Process 8: 접속정보 점검

접속정보 테이블을 이용하여 여러 이형태 중에서 하나의 표층어를 선택한다.

Process 9: 불규칙 처리

조동사의 어간이 불규칙 활용을 할 때는 이의 처리를 한다.

Process 10: 음운현상 처리

조동사가 결합을 할 때 경우에 따라서 모음조화, 모음축약, 모음탈락의 음운현상이 일어난다.

V. 평가

PIVOT E/K를 평가하기 위해서 선택된 600개의 문장을 생성한 결과를 가지고 조동사의 생성을 살펴 보았다. 영어 600문장은 많은 종류의 문법적 점검을 할 수 있도록 선택되어진 것이며 200여개의 문장에서 조동사를 생성하고 있다. 대표 표층어의 선택과 조동사의 결합에서 만족스러운 결과를 보여 주었다. 그러나 부정 조동사에서는 "지 않"과 "지 못하"의 구분이 잘 되지 않았다. 또한, 영어의 진행형 형태가 한국어 조동사에서는 상태와 진행을 모두 나타낼 수 있는데 현재는 진행으로만 처리가 되고 있다.

VI. 결론

본 논문에서는 중간언어방식의 기계번역시스템의 한국어 생성에서 조동사를 생성하기 위해 필요한 처리 과정과 지식베이스를 살펴 보았다. 중간언어로 표현된 문법적 기능과 의미로부터 조동사 표층어를 생성하기 위해서는 대표 표층어의 선정과 조동사 결합순서의 결정이 필요하다. 이를 위해서 Corpus를 분석하여 지식베이스를 구축한다.

한국어 생성 시스템이 구현된 PIVOT E/K는 10여만 단어를 가진 기계번역시스템으로 시스템의 평가를 위해 문법적 점검을 위해 마련된 영어 600문장을 해석하고 한국어 문장으로 생성하였다. 생성된 결과에서 조동사의 생성이 대표 표층어의 선택과 조동사의 결합 순서에 별다른 문제점이 없는 것을 확인하였다.

앞으로의 과제는 평가에서 지적된 부정 조동사, 상 조동사의 문제점 해결과 더불어 조동사와 결합하는 보조사의 처리가 필요하다. 보조사는 조동사에 결합하면서 문장의 의미를 보완하거나 새로운 의미를 추가한다. 좀 더 정확한 의미 전달과 자연스러운 문장의 생성을 위해서는 보조사에 대한 처리가 이루어져야 한다.

참고 문헌

- [1] 고영근, "현대국어의 선어말어미에 대한 구조적 연구," 어학연구, Vol. 3, No. 1, PP. 51-63, 1967
- [2] 김민수, 국어문법론, pp. 268-302, 일조각, 1984
- [3] 김승렬, 국어어순연구, pp. 43-94, 한신문화사, 1990
- [4] 남기심, 표준 국어문법론, pp. 20-28, 탑출판사, 1985
- [5] 박형달, "현대한국어의 보조동사의 연구," 언어학, Vol. 1, pp. 43-72, 1976
- [6] 한국과학기술원, 영한기계번역시스템(III): 문법개발지원환경 및 해석문법 개발, 과학기술처, 1992
- [7] 한국과학기술원 시스템공학센터, 한일 일한 자동번역 시스템 개발에 관한 연구(III), 과학기술처, 1986
- [8] 한국과학기술연구원 시스템공학연구소, 기계번역을 위한 언어모델링 및 골격시스템 개발 (3), 과학기술처, 1992
- [9] J.H.Lee and G.C.Kim, "Voice Generation from Conceptual Representation: Syntactic, Semantic, and Pragmatic Aspects," Literary and Linguistic Computing, Vol. 3, No. 4, pp. 250-254, 1988
- [10] J.H.Lee and A.Okumura, K.Muraki, and G.C.Kim, "An English-Korean Machine Translation System: Korean Synthesis under the Environment of Japanese Generation System," Proc. The 2nd Japan-Australia Joint Symposium on Natural Language Processing, Iizuka, Japan, pp. 219-224, October 1991
- [11] K.Muraki, S.Kamei, and N.Nomura, "The Interlingua for a Machine Translation System," JIPS SIGNLP, Vol. 89, No. 54, pp. 99-106, 1989