

두단계 대역어선택 방식을 이용한 구단위 패턴기반 한영 기계번역 시스템

김정재, 박준식, 최기선
한국과학기술원 전산학과 전문용어언어공학센터
대전시 유성구 구성동 373-1, 우:305-701
jjkim@csone.kaist.ac.kr, {jspark, kschoi}@world.kaist.ac.kr

Phrase-Pattern-based Korean-to-English Machine Translation System using Two Level Word Selection

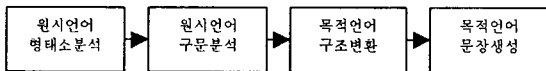
Jung-jae Kim, Junsik Park, Key-Sun Choi
Department of Computer Science, KORTERM
Korea Advanced Institute of Science and Technology
373-1 Kusong-dong, Yusong-gu, Taejon, 305-701

요약

패턴기반기계번역방식은 원시언어패턴과 그에 대한 대역언어패턴들의 쌍을 이용하여 구문분석과 변환을 수행하는 기계번역방식이다. 패턴기반 기계번역방식은 번역할 때 발생하는 애매성을 해소하기 위해 패턴의 길이를 문장단위까지 늘이기 때문에, 패턴의 수가 급증하는 문제점을 가진다. 본 논문에서는 패턴의 단위를 구단위로 한정시킬 때 발생하는 애매성을 해소하는 방법으로 시소러스를 기반으로 한 두단계 대역어 선택 방식을 제안함으로써 효과적으로 애매성을 감소시키면서 패턴의 길이를 줄이는 모델을 제시한다. 두단계 대역어 선택 방식은 원시언어의 한 패턴에 대해 여러 가능한 목적언어의 대역패턴들이 있을 때, 첫번째 단계에서는 원시언어 내에서의 제약조건에 맞는 몇가지 대역패턴들을 선택하고, 두번째 단계에서는 목적언어 내에서의 제약조건에 가장 적합한 하나의 대역패턴을 선택하는 방식이다. 또한 본 논문에서는 이와 같은 모델에서 패턴의 수가 코퍼스의 증가에 따른 수렴가능성을 논한다.

1. 서론

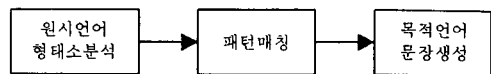
기계번역 방식은 크게 직접 번역 방식, 변환 방식, 중간 언어 방식으로 나뉜다. 이 방식들 중 초기의 변환 방식은 규칙기반 방식을 이용한다. 규칙기반 방식에서는 원시언어를 구문분석하고, 분석된 구문구조를 목적언어의 구문구조로 변환하고 변환된 목적언어의 구문구조트리를 바탕으로 목적언어에 맞는 문장을 생성하는 방식을 사용한다.



[그림 1] 전통적인 규칙기반 기계번역 방식

이 과정에서 원시언어의 구문구조를 목적언어의 구문

구조로 변환하기 위해서 사용되는 변환패턴들은 대개 번역의 질을 높이기 위해 고도로 어휘화된 형태로 기술된다. 이 어휘화된 변환패턴들을 이용하여 원시언어의 구문구조를 목적언어의 구문구조로 변환하는 것뿐만 아니라 원시언어의 구문분석까지 하는 방식이 패턴기반 기계번역방식이다. 이 방식은 기존 4단계 중에서 원시언어 구문분석과 목적언어로의 구조변환, 두 단계를 변환패턴을 이용하여 한 단계로 처리가 가능하기 때문에 속도를 향상시킬 수 있을 뿐 아니라 어휘화된 패턴을 이용하기 때문에 정확도도 향상시킬 수 있다.



[그림 2] 패턴기반 기계번역 방식

1.1 패턴의 형식에 따른 구분

이러한 패턴기반 기계번역방식은 패턴의 형식에 따라 몇 가지로 나누어볼 수 있다. 패턴의 형식은 크게 선형 구조와 트리구조로 나눌 수 있다. 선형구조는 하나의 패턴 내에 있는 요소들의 구문관계가 외부적으로 표현되어 있지 않고, 트리구조는 외부적으로 표현되어 있다.

- 선형구조의 예 : 그림+을 그리+는 사람
- 트리구조의 예 : (NP 사람
(VP 그리+는
(NP 그림+을)))

일반적으로 선형구조패턴에서는 구문구조가 외부적으로 표현되어 있지는 않지만 내부적으로는 패턴 내 요소들 간의 구문구조를 가진다. 또한 구축이 용이하다. 트리구조는 선형구조에 비해 구축이 어렵지만 외부적으로 구문구조가 표현되어 있어서 구문구조관계를 이용하고자 할 때 자동적인 탐색이 가능하다.

기존의 기계번역방식에서의 변환패턴은 대개 원시언어 패턴과 목적언어패턴이 모두 트리구조로 표현된다. (김영식, et al. 1989, 임철수, et al. 1997) 이때 원시언어패턴을 트리구조로 표현하는 것은 구조변환 이전에 원시언어에 대한 구문분석을 수행하기 때문이었다. 하지만 패턴기반 기계번역방식에서는 원시언어의 구문구조트리를 이용할 수 없기 때문에 원시언어 패턴은 트리구조일 필요가 없다. 오히려 원시언어 패턴들이 결합될 때 상위패턴에서 필요한 정보는 하위패턴의 의미자질과 중심어뿐이므로 이 정보들만 각 패턴이 가지고 있으면 된다. 하지만 목적언어에 맞는 문장을 생성할 때는 패턴내의 각 요소들 간의 구문관계를 알아야하기 때문에, 만약에 목적언어 패턴이 선형구조이면 정보의 손실의 우려가 있다.

본 시스템에서는 위와 같은 이유로 원시언어인 한국어의 패턴형식은 선형구조로, 목적언어인 영어의 패턴형식은 트리구조로 정하였다.

1.2 패턴의 단위에 따른 분류

기존 패턴기반 기계번역에서 패턴의 길이가 짧을 경우, 애매성이 증가하기 때문에 번역의 정확성을 높이기 위해 패턴의 길이를 문장단위까지 길게 하고 코퍼스에서 자동추출하는 방식을 사용하였다. (한국과학기술원 1998) 하지만 패턴의 단위가 문장이 될 경우, 번역영역이 특정분야로 한정되더라도 패턴의 수가 수렴할 것인지가 확실하지 않고, 패턴의 길이가 길어질수록 패턴부족문제 때문에 분석이 실패할 가능성이 많아진다.

본 시스템에서는 패턴의 단위를 구단위로 한정했을 때, 발생하는 애매성, 원시언어 구문분석시의 애매성과 대역어 선택시 애매성을 해소할 수 있는 모델을 제안한다. 2장에서는 기존 패턴기반 기계번역 모델들에 대해서

설명하고, 3장에서는 구단위 패턴들의 유형을 분류하고 코퍼스 증가에 따른 패턴 수의 수렴가능성을 논한다. 4장에서는 본 논문에서 제안하는 두 단계 대역어 선택방식에 대해 설명하고, 5장에서는 구축된 시스템에 대해 설명한다.

2. 기존연구

대역어 선택에 대한 기존의 연구는 선택과정이 한 단계인지, 여러 단계인지로 크게 구분할 수 있다.

대역어 선택과정이 여러 단계로 나뉘는 연구 중, Palmer, M. (1998)에서는 각 언어가 가지는 체계도가 언어마다 매우 다르기 때문에 언어간의 번역시 두 체계를 병합하는 것은 매우 어렵다는 것을 지적하고, 언어의 선택제약은 각 언어 고유의 체계도를 이용한다.

예를 들어 'wear'의 경우, 한국어에는 'wear'가 차지하는 범위에 해당하는, 즉 'wear'가 의미하는 바와 정확하게 일치하는 한국어 대역어가 없기 때문에 'wear'를 한국어로 번역하려면 'wear'의 목적어를 한국어의 체계도에 따라 구분하여 사전에 기입해야 한다. 하지만 이를 일일이 구분하여 사전에 기입하는 것은 거의 불가능하기 때문에 'wear'의 번역어로 '쓰다, 입다, 매다, 신다 등'을 사전에 등재하고 (coarse-mapping), 그 중 어떤 대역어를 선택할 것인지는 각 동사가 목적어에 대해서 가지는 한국어 고유의 선택제약을 이용하여 번역한다. 그러면 두 언어간의 서로 다른 의미체계를 모두 분석하는 부담을 덜게 된다.

Lee, H.A. et al. (1999)에서는 영어 동사에 대한 한국어 대역동사 선택시 원시언어동사의 의미분별, 의미 대 단어 맵핑, 가장 적절한 목적언어동사의 선택의 3단계과정을 제안하였다. 원시언어동사의 의미분별을 위해서 각 원시언어의 의미마다 대표예제를 사용하는데, 주어진 명사와 대표예제의 WordNet상에서의 거리를 이용하여 가장 가까운 대표예제의 의미를 선택한다. 가장 적절한 목적언어동사 선택은 한국어의 구문공기정보를 이용하여 통계적으로 선택한다.

3. 구단위 패턴의 유형분류

패턴기반 기계번역에서는 패턴을 이용하여 전체문장을 구문분석해야 하기 때문에 패턴의 유형에 대한 자세한 분류가 필요하다. 본 시스템에서 구단위 패턴을 크게 용언의 기본문형패턴, 명사구 패턴, 접속절 패턴, 문장종결 패턴, 부사구 패턴으로 나눈다.

3.1 용언의 기본문형패턴

용언의 기본문형패턴은 'NP1+가 NP2+를 NP3+에게 주다'와 같은 용언의 기본문형들에 대한 패턴을 의미한다. 기존의 용언의 하위범주화에 대한 연구가 많이 진행

되어서 그 결과를 이용할 수 있지만, 지금까지는 용언의 기본문형에 대한 대역영어패턴 구축이 많이 되어 있지 않았으므로 이에 대한 연구가 절실하다. 본 시스템에서는 경제분야 서신분야에서 빈도수가 높은 한국어 300여 개의 용언에 대한 기본문형패턴 및 대역영어패턴을 구축하였다.

3.2 명사구 패턴

명사구는 크게 지시사나 일반 형용사 등으로 수식받는 명사구와 관형절의 수식을 받는 명사구로 나눌 수 있다. 그런데 한국어에서의 지시사는 그 수가 한정되어 있고, 일반 형용사에 의해 수식받는 명사구에 대한 패턴은 'ADJ + NP'와 같은 형태로 표현이 가능하기 때문에 패턴의 수가 한정되어 있다. 하지만 관형절에 의해 수식을 받는 명사구의 경우 어떤 논항이 관형절에서 생략되었는지에 따라 패턴이 달라지므로 그 수가 급증하게 된다. (Park, et al. 1997)

- 그림을 그리는 사람 : a person who draws a picture
- 내가 선물을 준 친구 : a friend to whom I gave a present
- 사건이 일어난 대학 : a college in which the accident took place
- 이 장군이 적을 무찌른 칼 : a sword with which General Yi defeated enemies

하지만 관형절에 의해 수식받는 명사구에 대한 패턴은 용언의 기본문형패턴에서 비롯된 것이므로 용언의 기본문형패턴에서 자동변환으로 비교적 쉽게 구축이 가능하다.

3.3 접속절 패턴

접속절 패턴은 'VP1+고 VP2'와 같이 연결어미, 접속조사 등에 의해 둘 이상의 패턴이 접속되는 패턴이다. 연결어미와 접속조사의 수가 한정되어 있으므로 접속절 패턴의 수도 한정되어 있다.

3.4 문장종결 패턴

문장종결패턴은 'VP+바니+다+'와 같이 절 또는 구가 선언어말어미, 종결어미와 문장마침표(':', '!', '?' 등)와 결합한 패턴이다. 선언어말어미, 종결어미의 수가 한정되어 있으므로 문장종결 패턴의 수도 한정되어 있다.

3.5 부사구 패턴

부사구 패턴은 시간부사구, 장소부사구 등 여러 종류로 나누어볼 수 있는데 그 중 시간부사구는 다른 부사구에 비해 'A+년 B+월 C+일'과 같이 일정한 패턴으로 인식할 수 있는 경우가 많이 있다. 본 시스템에서는 패턴으로 미리 인식할 수 있는 부사구, 특히 시간부사구를

전처리기에서 미리 인식하여 구문분석 시에는 하나의 요소로 받아들여도록 하고 (박준식, 최기선 1999), 나머지 부사구는 되도록 용언의 기본문형패턴의 논항으로 들어가게 하고 시스템의 강건성 (robustness)을 위해 각 부사구를 처리할 수 있는 추상적인 패턴을 구축하는 방향으로 한다. 이와 같은 관점에서 부사구 패턴의 수도 한정되어 있다.

지금까지 용언의 기본문형패턴과 관형절의 수식을 받는 명사구 패턴을 제외하고는 패턴의 수가 한정됨을 보였다. 따라서 코퍼스의 증가에 따른 용언의 기본문형패턴의 수가 수렴될 것으로 기대되므로 전체적인 패턴의 수도 코퍼스의 증가에 따라 수렴할 것으로 기대된다.

4. 두단계 대역어 선택방식

그 동안 대역어 선택에 대한 연구가 많이 진행되어 왔는데 사용하는 정보는 크게 예제나열 (김나리, et al. 1996), 대표예제나열 (Lee, et al. 1999), 시소러스 (문경희, et al. 1998), 의미자질 (Palmer 1998) 등을 이용한다. 예제나열 자체는 자료부족문제가 심각하게 발생할 수 있고, 대표예제를 이용하는 방식은 대표예제를 선택하기가 어렵고 대표예제가 각각 동일하게 취급되므로 크기가 서로 다른 영역을 묘사하는 것이 어렵다. 시소러스는 그 자체만으로 대역어 선택의 높은 정확성을 기대하기 어렵다. 의미자질은 모든 단어마다 각 언어의 모든 자질에 대한 찬반 여부를 결정해야 하는 작업이 뒤따르게 되므로 현실적으로 아직 사용하기 힘든 정보이다.

본 시스템에서는 Palmer (1998), Lee (1999)에서 주장한 것처럼 두 언어 모두의 의미구분을 하나의 체계 안에서 하는 것이 불가능하므로 각 언어의 의미구분은 각 언어의 의미 체계 내에서 해결한다. 한국어의 구문분석시의 애매성은 한국어 시소러스와 예제나열방법을 사용하여 해결하고, 영어대역어선택시의 애매성은 영어단어들간의 구문공기정보와 영어시소러스를 사용하여 해결한다. 그리고 상호 연결은 1:1 대응 방식이 아닌 대략 대응 (coarse-mapping) 방식을 이용한다.

4.1 첫번째 대역어 선택

예를 들어 한국어 용언기본문형패턴 'NP+를 쓰다'에 대한 대역영어패턴은 다음과 같이 여러 가지가 존재할 수 있다.

- NP+를 쓰다
- write NP, compose NP, use NP, wear NP, ...

본 시스템에서는 위의 영어패턴들을 NP의 한국어 중심어에 따라 한국어 시소러스와 예제나열방법을 혼합한 혼합 (hybrid) 방식을 이용하여 대역영어패턴을 몇개의

그룹으로 나눈다.

- NP=[글,문서] – write NP, compose NP, ...
- NP=[의류] – put on NP, wear NP, ...
- NP=[도구,기계] – use NP, employ NP, handle NP, ...

대역영어패턴을 나누기 위하여 시소러스를 이용할 때, 예를 들어 NP의 한국어 중심어가 시소러스 상에서의 [의류]와 상위어관계 또는 하위어관계에 있을 때, 'put on NP, wear NP, ...'의 패턴 그룹이 선택된다.

다음은 'NP+를 내리다'에 대한 패턴구축의 예이다.

- NP='결정' – make decision
- NP='결론' – have come to the conclusion
- NP=[가치,가격] – lower NP, cut down NP, reduce NP
- NP=[인공물] – lower NP, drop NP, take down NP, ...

예제나열방법은 '결정+을 내리다', '결론+을 내리다'와 같이 NP가 특정 몇 개의 명사로 제한되어 나타날 때 사용한다.

4.2 두번째 대역어 선택

NP의 한국어 중심어에 의해 일련의 대역영어패턴들이 결정되면 그 중에 가장 적절한 대역영어패턴을 선택하는 것은 NP의 대역영어명사와 대역영어패턴의 동사와의 구문공기정보를 이용하여 결정한다. 이때 자료부족문제가 발생할 수 있으므로 WordNet과 같은 영어 시소러스를 사용할 수 있다.

다음은 'NP+를 쓰다'에 대한 번역패턴의 예이다.

한국어 패턴	NP의 제약조건	영어 대역패턴
NP+를 쓰다	[의류]	wear NP put on NP cover NP
	[문서]	write NP compose NP put down NP describe NP
	[인공물]	use NP employ NP handle NP

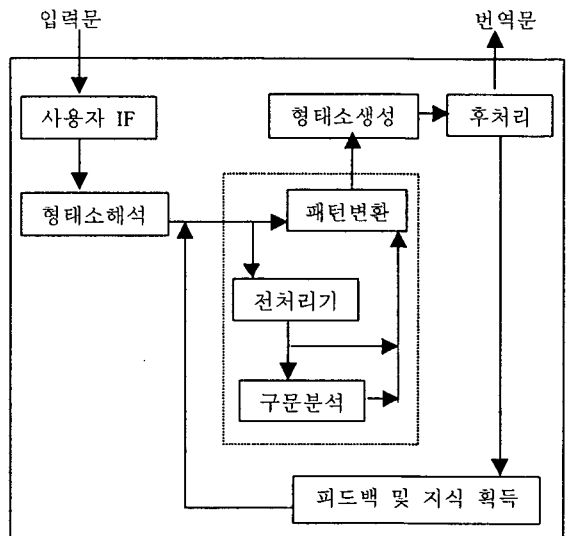
[입력예1] 입력으로 '종이를 쓰다'가 들어오면 '종이'에 대역어가 'paper'이므로 'write paper'와 'use paper'가 받아들여질 수 있는 번역이 될 수 있다. 하지만 '종이'가 [문서]가 아닌 [인공물]에 속하므로 'use paper'로 번역된다.

[입력예2] '기계를 쓰다'라는 문장이 들어오면 '기계'가 인공물이기 때문에 'use machine', 'employ machine', 'handle machine'으로 번역될 수 있다. 그런데 'use, employ, handle'과 'machine'이 동사-목적어로 나타난 횟수를 보면

'use-machine'이 7번 나타나고 'employ-machine', 'handle-machine'이 한번도 나타나지 않으므로 'use machine'을 최종 번역으로 선택하게 된다.

5. 두단계 대역어 선택방식을 이용한 한영 구단위 패턴기반 기계번역 시스템

기존의 한영기계번역은 1) 한국어 형태소 분석, 2) 한국어 구문분석, 3) 한영 구조변환, 4) 영어문장생성의 4단계로 이루어졌다. 하지만 한국어 구문분석이 아직 기계번역에 이용할 만큼 좋은 결과를 내지 못하고 있고, 오히려 한영 구조변환에서 번역의 정확도를 높이기 위하여 어휘화된 패턴을 이용한다. 이러한 변환패턴을 이용하여 동시에 원시언어 구문분석과 구조변환을 하는 방식이 패턴기반 기계번역 방식이다. 이 방식은 1) 한국어 형태소분석, 2) 패턴매칭 (한국어 구문분석 + 한영 구조변환), 3) 영어문장생성의 3단계로 이루어진다.



[그림 3] 구단위 패턴기반 기계번역 시스템 흐름도

5.1 전처리

한국어 형태소분석의 결과를 바로 패턴매칭에 이용할 경우 이형태로 인한 패턴수의 급증과, 조사 '-에'가 가지는 애매성 때문에 패턴의 잘못된 매칭이 많아진다. 이와 같은 이유때문에 본 시스템에서는 패턴매칭 전에 전처리 단계를 두어 이형태의 통일, 시간부사구 인식, 수치/시간 표현 인식, 수량사의 처리, 패턴을 이용한 복합명사인식 등을 수행한다.

5.2 패턴매칭

본 시스템에서는 패턴매칭 알고리즘으로 Generalized LR Parsing (GLR) 방식을 이용하였다. GLR은 애매성이 존재하는 CFG 문법에 LR 방식을 적용하기 위하여 Tomita에 의해 개발된 방법이다. 보통 LR 방식은 선형 스택을 사용하는데 이 방법은 파싱결과를 그래프 구조화 스택(Graph Structured Stack; GSS)을 이용하여 저장한다. 그래프를 이용하므로 Local Ambiguity Packing (LAP) 기법과 Shared Parse Forests (SPF) 기법이 가능하다. LAP 기법은 한국어 문장에서 같은 범위의 구가 같은 범주를 가지면 하나의 결과처럼 패키징하는 방법이고, SPF 기법은 한 노드를 여러 부모가 자식노드로 공유할 수 있도록 하는 기법이다.

한국어 패턴구축시 사용한 시소러스는 일본 NTT 사건(1997)에 있는 명사와 용언에 대한 체계 (이하 NTT 체계)를 사용하였다. NTT 체계는 전체 레벨이 10까지 내려가지만 본 시스템에서는 NTT 체계의 레벨 6까지만 사용하여 1700여개 단어에 대한 의미코드를 부착하였다. 패턴매칭이 끝나고 대역어를 선택할 때 필요한 영어구문공기정보는 PennTreeBank에서 서로 다른 주어-동사 쌍 6만여개, 동사-목적어 쌍 8만여개를 추출하여 사용하였다.

5.3 영어문장생성

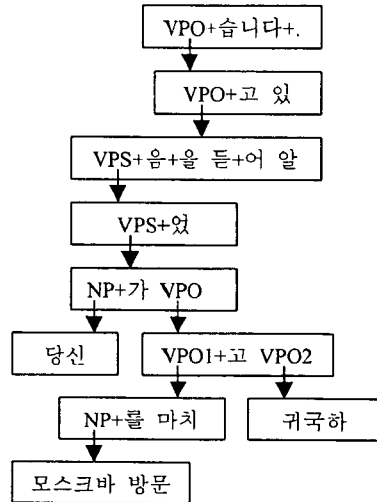
영어문장 생성시 어순, 동사의 시제, 성·수 일치, 명사의 수·관사 등을 정확하게 처리해야 한다. 기존의 패턴기반 기계번역에서는 대역패턴을 선형구조로 표현함으로써 이와 같은 문제에 대해 적절히 대처할 수가 없었다. 본 시스템에서는 대역영어패턴을 트리구조로 표현함으로써 기존의 영어문장생성에 관하여 연구된 보다 정교한 방법들을 사용할 수 있는 환경을 제공한다.

5.4 실험결과

다음은 본 시스템에서 수행한 번역의 한 예이다.

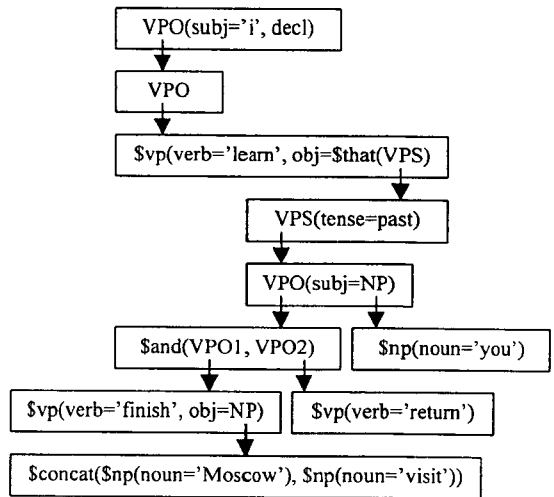
[한국어 문장] 당신이 모스크바 방문을 마치고 귀국하였음을 들어 알고 있습니다.

다음은 위 문장을 패턴들을 이용하여 구문분석한 결과인 패턴트리들 중의 하나이다.



[그림 4] 패턴매칭 결과의 예

다음은 각 한국어 패턴에 대한 여러 가지 대역영어패턴들 중에서 영어구문공기정보를 이용하여 가장 적절한 대역영어패턴을 선택한 영어패턴트리이다



[그림 5] 대역어선택이 끝난 영어패턴트리의 예

다음은 위 영어패턴트리로부터 생성한 영어문장의 한 예이다.

[영어 문장] I learn that you finished Moscow visit and returned.

6. 결론

본 논문에서는 문장단위로 패턴의 길이가 길어졌을 때 발생하는 패턴 수의 급증을 해결하고자 패턴의 길이를 구단위로 한정하는 방안을 제시하였다. 패턴의 길이가 짧아지면서 대두되는 번역시의 애매성 문제는 두단계 대역어선택방식을 이용하여 해결하고자 하였다. 그리고 구단위 패턴의 유형들을 분류조사하여 코퍼스 증가에 따라 패턴수가 수렴가능함을 보였다.

앞으로는 구구조로 표현되는 변환된 영어구구조를 이용하여 기존의 관사/소유격, 동사의 시제, 성수 일치, 필수격의 복원, 재배열 등에 대한 연구를 적용해야 한다.

7. 참고문헌

- [1] 김나리, 김영택, 한국어 동사 패턴에 기반한 한국어 문장 분석과 한영 변환의 모호성 해결, 한국정보과학회논문지 제23권 제7호, 1996
- [2] 김영식, 최기선, 김길창, 영한 기계번역에 있어서 변환과 역어선택, 한국정보과학회 가을 학술발표논문집, 1989
- [3] 김중혁, 한영기계번역에서의 예문을 이용한 변환, 한국과학기술원 석사학위논문, 1994
- [4] 문경희, 이종혁, 김정인, 양기주, 일한 기계번역 시스템: 언어패턴을 이용한 어휘 다의성 해소, 한국정보과학회논문지 제25권 제8호, 1998
- [5] 박준식, 최기선, 한영 기계번역에서의 효율적인 구문 분석과 번역을 위한 FST 전처리의 설계 및 구현, 한글 및 한국어정보처리학회, 1999 (발표예정)
- [6] 서병락, 김영택, 한영 기계번역을 위한 번역 패턴에 기반한 영어 문장 생성기, 한국정보과학회논문지 제23권 제5호, 1998
- [7] 임철수, 이현아, 최명석, 이공주, 김길창, "어휘화된 규칙에 기반한 영한 기계 번역 시스템", 한국정보과학회 추계학술대회 발표 논문집, Serial 24, 161-164, 1997
- [8] 한국과학기술원, 예제기반 한영 기계번역 시스템 개발에 관한 연구, 최종보고서, 한국전자통신연구원, 1998
- [9] 日本語語彙大系 / NTTコミュニケーション科 學研究所, 1997
- [10] Lee, H.A., Park, J.C. and Kim, G.C., Lexical Selection with a Target Language Monolingual Corpus and an MRD, Proceedings of the 8th International Conference on Theoretical and Methodological Issues in Machine Translation, 1999
- [11] Palmer, M., Constraining Language Specific Constraints on Lexical Selection with Feature-based LTAG, TAG+ Workshop, 1998
- [12] Park, S.B. and Kim, Y.T., Semantic Role Determination in Korean Relative Clauses Using Idiomatic Patterns, Proceedings of ICCPOL 97, pp. 1-6, 1997
- [13] Takeda, K., Pattern-based Machine Translation, COLING-96 pp. 1155-1158, 1996
- [14] Tomita, M., Generalized LR Parsing, Kluwer Academic Publishers, 1991
- [15] Watanabe, H. and Takeda, K., A Pattern-based Machine Translation System Extended by Example-based Processing, COLING-98 pp. 1369-1373, 1998
- [16] Zhang, M. and Choi, K.S., Pipelined multi-engine Machine Translation: accomplishment of MATES/CK system, Proceedings of the 8th International Conference on Theoretical and Methodological Issues in Machine Translation, 1999