



## 번역패턴 기반 한독 자동번역<sup>1)</sup>

최승권(한국전자통신연구원)

### 1. 머리말

인터넷의 빠른 보급으로 국가간에 각종 정보의 교류는 날로 증가하고 있으며 교류되고 있는 정보의 양적 팽창은 자국어에 대한 이해뿐만 아니라 외국어에 대한 빠른 이해를 요구하기 때문에 이러한 정보의 양적 팽창은 언어장벽을 해소하는 자동번역 연구를 더욱 필요로 하고 있다. 그러나 정보의 홍수 속에서 자동번역 연구에 대한 필요성은 누구나 공감하고 있지만 현재의 자동번역의 번역 수준은 일반인들이 기대하는 수준에는 아직 못 미치고 있다.

그럼에도 불구하고 자동번역의 필요성은 대용량 번역(문서의 양은 무제한임), 고속 번역(평균적으로 1초에 10문장), 대역어 번역의 일관성(다의성이 없는 하나의 단어는 하나의 대역어로 일관되게 번역), 일반대중의 혜택(외국어를 전혀 모르는 사람이라도 이용 가능)이라는 장점 때문에 전세계적으로 지속적으로 강조되고 연구되고 있는 실정이다. 그 동안 국내외 적으로 한국어와 관련된 자동번역 시스템으로 영어, 일본어간의 자동번역에 대해서는 집중적인 연구가 있었고 상용화된 번역시스템도 시장에 나오는 수준에 도달하여왔다.

그러나 한국어와 독일어간의 자동번역 연구는 기업들의 이윤추구라는 시장 논리에 의해 상용화된 번역 시스템으로 시장에 나온 것은 없고 언어학적인 관점에서 이론적으로만 연구되어 왔다.(최승권(1996), 이민행 외(1998)).

이러한 시점에서 본 논문에서는 향후 실제적인 한독 자동번역 시스템을 개발하기 위한 기초 작업으로써 한독 자동번역 시스템을 위해 없어서는 안 될 한독 번역지식으로써의 다양한 한독 번역패턴을 소개하고자 한다. 본 논문에서 기술되는 번역패턴은 향후 한독 자동번역 시스템에 직접 활용될 수 있으며 독

1) 이 논문에 대하여 논문 심사를 맡아 유익한 논평을 해주신 익명의 심사위원 세 분께 감사를 드린다. 세분 심사위원의 세심한 논평이 논문의 잘못된 부분과 모자라는 점을 개선하는 데 많은 도움이 되었다. 그러나 이 논문에서 비롯되는 오류는 전적으로 필자의 몫이다.

일어 교육 및 작문을 위한 독일어 교육 시스템이나 한독 전자사전에도 활용할 수 있을 것이다.

## 2. 기존 한국어를 출발언어로 한 자동번역 방법의 문제점

### 2.1. 의존문법 기반 한국어 구문분석의 문제

어순의 자유성, 생략 등 한국어의 특성과 관련하여 구구조 문법(*phrase structure grammar*)으로 한국어를 표현하면 분석 규칙의 수가 지나치게 많아지므로 구구조 문법은 한국어 구조 분석에 적합하지 않고(홍영국, 1993) 단어 사이의 의존 관계에 중심을 두는 의존 문법이 한국어 구조 분석에 적합하다는 주장이 있었다(류범모, 1996).

이러한 의존문법에 기반을 둔 한국어 구조분석 방법은 어순의 자유성, 문장 구성요소의 불연속성, 생략 등의 현상을 처리하는 데에는 적합하지만 구조 분석 결과가 너무 지나치게 많아지므로 해당 문장에 대한 가장 정확한 구조분석 결과를 찾아내는 어려움을 가지는 문제를 야기시켰다. 이러한 대량의 구조분석 결과 중에 정확한 분석구조를 찾아내는 일이 자동번역에서 중요한 데 그 이유는 정확한 분석구조를 선택하지 못하면 분석 이후의 절차를 아무리 잘 처리하여도 원문에 대한 번역문이 제대로 만들어 질 수 없어서 번역품질을 저하시키는 가장 중요한 요인이 되기 때문이다. 이러한 점에서 한국어 구문분석의 정확률을 높이기 위해 한국어 의존문법 기반 구문분석기에 의해 만들어진 분석 결과에 대해 ‘상호 정보를 적용한 방법2’(정석원, 2001)과 ‘말뭉치에서 직접 추출한 명사-조사 쌍과 동사에 의한 격틀 정보를 적용한 방법3’(진은희, 2001) 등이 제안되어 왔었다. 전자의 방법은 대량의 상호 정보 추출이 전제가 되어 상

- 
- 2) 문장 정확률(‘정답과 구조가 일치하게 분석된 문장의 수’를 ‘분석에 성공한 문장의 수’로 나눈 값, 분석에 성공한 문장이란 문장의 전체에 대한 파스트리가 생성된 경우를 말함)은 10어절 이내에서는 77%, 10어절 이상에서는 46%로 실험결과가 나왔음.  
3) 명사-조사쌍 의존 구문분석 처리 결과 정확률은 90%, 재현율은 74%의 성능을 보였다. 여기서 재현율이 낮은 이유는 격틀자료의 부족과 시소러스의 미흡함이 주요 원인이었다.

호 정보에 대한 통계적 처리가 주 해결책이나 대량의 정확한 상호 정보 추출 방법은 아직도 연구단계에 있기 때문에 실질적인 정확률 향상의 해결책은 되지 못하고 있다. 후자의 경우는 본 논문에서 시도하는 방법과 유사하나 말뭉치에서 추출된 명사-조사 쌍과 동사의 격틀정보가 임의의 문장 분석에 적용되는 적용영역이 너무 작기 때문에 이 또한 실질적인 정확률 향상의 해결책은 되지 못한다는 단점을 가지고 있다.

본 논문에서는 한국어 구조 분석에 적합한 의존문법의 장점을 취하면서 구조 분석 결과에 대한 정확률을 향상시키기 위해 어휘의미를 논항으로서 가지는 동사의 격틀을 적용하는 방법을 제안하고자 한다. 즉 의존문법의 이분법 규칙에 의해 가능한 모든 구조분석이 이루어진 후에 한국어 격틀이 각각의 구조 분석결과가 올바른지를 점검하도록 하여 올바른 구조분석결과가 찾아지도록 하는 것이다.

## 2.2. 복합문의 번역 문제

지금까지 한국어 복합문의 자동번역에 대한 연구는 단문의 자동번역과 비교할 때 연구가 미비하였다. 그 이유는 단문에 대한 구조 분석도 많은 성과를 이루지 못한 상태에서 복합문을 다룬다는 것이 너무 어렵기 때문이었다. 하지만 복합문에 대한 언어학적 연구는 많은 논문들에서 다루어져왔다(권재일(1985), 최재희(1992), 강현화(1998))

복합문이 특히 자동번역에서 어려운 이유는 단문들을 복합문으로 만드는 연결어미의 번역이 어렵기 때문이다. 연결어미는 단문들의 의미에 따라 다의성을 가질 수 있기 때문에 독일어로 번역될 때 연결어미의 의미를 정확히 파악해야만 독일어의 올바른 표현, 특히 올바른 접속사로 번역될 수 있다.

한국어 연결어미 “고”가 독일어로 “und”로만 번역되는 것이 아니라 단문들의 의미에 따라 다양하게 독일어로 번역되는 경우는 아래의 예에서 확인할 수 있다.

- 1) 그가 학교에 가고 나는 집에 있었다. (“고” : 나열)  
Er ist in die Schule gegangen und ich bin zu Hause geblieben.
- 2) 그녀는 그 소식을 듣고 매우 놀랐다. (“고” : 결과)  
Er war sehr überrascht, als er die Nachricht hörte.

위의 예문 1)과 2)에서 “고”가 독일어의 und나 als 부사절로 번역되는 것은 각각의 단문들의 의미간의 관계성 때문에 이루어지는 것이다. 본 논문에서는 복합문에서의 단문들의 의미간의 관계성으로 인해 한국어의 연결어미가 영어의 다양한 접속사나 해당 표현으로 번역되는 것을 동사의미를 포함하는 연결패턴에 의해 해결하는 방법을 제안하고자 한다.

### 3. 한독 자동번역을 위한 번역 패턴의 종류

한독 자동번역을 위해 사용하는 번역패턴은 크게 격틀 번역패턴, 연결 번역패턴, 명사구 번역패턴의 3가지 유형으로 나눌 수 있다. 각각의 번역패턴에 대해서는 다음절에서 상세히 기술할 것이다.

#### 3.1. 격틀 번역패턴

##### 3.1.1. 격틀 번역패턴의 정의

격틀 번역패턴이란 동사를 중심으로 하는 단문 중심의 한국어 격틀에 대한 독일어 격틀을 대응시킨 패턴을 말한다.

##### 3.1.2. 격틀 번역패턴의 기능

격틀 번역패턴은 한국어에 대해서는 의존문법 분석결과에 대한 올바른 단문 구조 파악을 위해 사용되며 독일어에 대해서는 한국어 격틀에 대한 독일어 대역문을 생성하는 역할을 한다.

##### 3.1.3. 격틀 번역패턴의 포맷

한독 격틀 번역패턴은 DB에 저장되는데 DB에 저장되는 형식은 격틀 내용을 찾기 위한 동사어휘인 key와 격틀 내용이 기술되어 있는 content로 기술된

다. 한독 격틀 번역패턴의 포맷은 다음과 같다:

항목	내용
Key	동사어휘+인덱스
Content	<p>&lt;격틀번역패턴&gt; →&lt;한국어토큰리스트&gt; &gt; &lt;영어토큰리스트&gt;                      &lt;한국어토큰리스트&gt; →&lt;실질형태소!형식형태소&gt; &lt;한국어토큰리스트&gt;   &lt;실질형태소!형식형태소:분석 정보&gt; &lt;한국어토큰리스트&gt;   &lt;메타 어휘=의미코드!형식형태소&gt; &lt;한국어토큰리스트&gt;   &lt;메타 어휘=의미코드!형식형태소:분석 정보&gt; &lt;한국어토큰리스트&gt;   NIL                      &lt;영어토큰리스트&gt; → &lt;일반 어휘&gt; &lt;영어토큰리스트&gt;   &lt;메타 어휘&gt; &lt;영어토큰리스트&gt;   &lt;메타 어휘:활용 정보&gt; &lt;영어토큰리스트&gt;   &lt;메타 어휘&amp;생성 정보&gt; &lt;영어토큰리스트&gt;   &lt;메타 어휘:활용 정보&amp;생성 정보&gt; &lt;영어토큰리스트&gt;   NIL                      &lt;메타 어휘&gt; → A   B   ... , X   Y                        (체언은 A,B 순으로, 용언은 X,Y순으로 기술한다.                      메타 어휘가 아닌 실제 문자 표시는 !A, !B, !C 로 표시한다.)                      &lt;분석 정보&gt; →&lt;시제 정보&gt;   &lt;양상 정보&gt;   &lt;문장 정보&gt;   &lt;화법 정보&gt;   &lt;품사 정보&gt;                      &lt;생성 정보&gt; →&lt;시제 정보&gt;   &lt;양상 정보&gt;   &lt;문장 정보&gt;   &lt;화법 정보&gt;   &lt;품사 정보&gt;                      &lt;활용 정보&gt; →[SUBJ   DOBJ   IOBJ   POSS   REFL  PRON]   [BSE   INF   GER   PSP   PRP   PAS   COMPRA   SUPRA]                      &lt;시제 정보&gt; →[PRES   PAST   FUT   PPAST]                      &lt;양상 정보&gt; →[PROG   PERF]                      &lt;문장 정보&gt; →[DECL   EXCL   INPR   QUES   SUGG ]                      &lt;화법 정보&gt; →한국어보조용언어휘                      &lt;품사 정보&gt; → v</p>

격틀 번역패턴 중 '먹다' 동사에 대한 격틀 번역패턴 예를 보이면 다음과 같다:

- 3) 범인이 고기를 먹다  
 Key: 먹다1  
 Content: A=사람!가 B=식품!를 먹!다 > A essen:v B
- 4) 범인이 소화제를 먹다  
 Key: 먹다2

Content: A=사람!가 B=약품!를 먹!다 > A nehmen:v B

5) 법인이 귀를 먹다

Key: 먹다3

Content: A=사람!가 귀!를 먹!다 > A sein:v taub

6) 그는 회사로부터 큰 이문을 먹다

Key: 먹다4

Content: A=사람!가 B=재화!를 C=조직!로부터 먹!다 > A ziehen:v aus  
C B

3)에서 5)은 ‘누가 무엇을 먹다’의 2항의 값을 가지는 “먹다”의 격틀 실례를 나타낸 것으로 3)이 “먹다”의 제일 일반적인 격틀 형태이며 4)는 목적어가 ‘약품’일 때의 격틀과 그것의 대역표현을 나타낸 것이고 5)는 ‘귀를 먹다’와 같이 속어를 표현하는 격틀 번역패턴을 기술한 것이다. 6)는 ‘누가 무엇을 어디로부터 먹다’의 3항의 값을 가지는 “먹다”의 격틀 실례를 나타낸 것이다.<sup>4)</sup>

### 3.1.4. 논항의 의미 정보

격틀 번역패턴에서 논항이 변수로 기술되는 부분은 논항의 어휘 의미 정보로 채워지고 있다. 이 논항의 어휘 의미 정보 체계와 관련해 시소러스를 활용코자 하였으나 아직 국내에는 한국어와 관련해 완성된 시소러스가 없기 때문에 현재의 어휘의미 정보 체계는 크게 구체명사(세상에 존재하는 구체물. 인간이 실제로 인지할 수 있는 대상), 추상명사(인간의 관념속에 존재하는 대상. 오감을 사용하여 인지할 수 없는 대상), 활동명사(동사적인 의미를 가지면서 시간의 선상에서 한 점으로 존재하는 것이 아니라 연속적으로 분포되어 있는 개념)로 구분하고 전체적으로는 150여개의 어휘의미 정보로 구성되어 있다. 그림 1은 번역패턴의 논항들에서 사용되는 의미정보 체계도를 나타낸다.

---

4) 격틀 패턴은 의존문법내의 논항구조와 차이가 나는데 필수 논항(obligatorisches Argument)뿐만 아니라 번역에 영향을 미칠 수 있는 수의 논항(fakultatives Argument) 및 부사 등과 같은 부가어가 기술될 수 있다는 점과 대역어 부분이 대응되어 있다는 점이 다르다.





### 3.2.2. 명사구 번역 패턴의 기능

명사구 번역패턴은 한국어에 대해서는 올바른 명사구 파악을 위해 사용되며 독일어에 대해서는 한국어 명사구에 대한 독일어 대역구를 생성하는 역할을 한다.

### 3.2.3. 명사구 번역 패턴의 포맷

항목	내용
Key	S와 한국어어휘
Content	<p>&lt;명사구번역패턴&gt;→&lt;한국어 토큰리스트&gt; &gt; &lt;영어토큰리스트&gt;          &lt;한국어토큰리스트&gt;→ &lt;실질형태소&gt; &lt;한국어토큰리스트&gt;            &lt;실질형태소!형식형태소&gt; &lt;한국어토큰리스트&gt;  &lt;메타어휘=의미          범주코드&gt; &lt;한국어토큰리스트&gt;  &lt;메타어휘=의미범주코드!형식형          태소&gt; &lt;한국어토큰리스트&gt;   NIL          &lt;영어토큰리스트&gt; → &lt;일반어휘&gt; &lt;영어토큰리스트&gt;   &lt;메타어          휘&gt; &lt;영어토큰리스트&gt;  &lt;메타어휘:활용정보&gt; &lt;영어토큰리스트&gt;            NIL          &lt;메타 어휘&gt; → A   B   ... , X   Y   Z          &lt;활용 정보&gt; →POSS   REFL   GER   COMPRA   SUPRA            DEF   NODEF   INDEF</p>

명사구 번역패턴은 DB에 저장되는데 DB에 저장되는 형식은 명사구 번역패턴을 찾기위한 명사어휘를 대체하는 S라는 메타기호와 관형격조사로 구성된 key와 그것의 내용이 기술되어 있는 content로 기술된다. 한독 명사구 번역패턴의 포맷은 다음과 같다:

DB에 들어갈 한독 명사구 번역패턴의 예를 보이면 다음과 같다:

7) 장수의 비결 > Geheimnis von Langlebigkeit

Key: S!의 S

Content: A=인생사!의 B=생각 > B von A

## 8) 정세의 변화 &gt; Veränderung der Lage

Key: S!의 S

Content: A=정치활동!의 B=변화 > B art:[*funct\_gen(A), funct\_num(A), 2*] A

## 9) 연정희와 정일순 등 7명 &gt; sieben Personen einschließlich Yon Jung-hee und Chong Il-soon

Key : S!와 S\_등\_N\_명

Content : A=인명!와 B=인명 등 N 명 > N personen:n einschließlich A und B<sup>5)</sup>

7)에서는 'A의 B'라는 한국어 명사구가 각각 A가 '인생사', B가 '생각'이라는 의미를 가질 때 독일어에서는 'B von A'라는 명사구로 번역된다는 것을 말하며, 8)에서는 'A의 B'라는 한국어 명사구가 각각 A가 '정치활동', B가 '변화'라는 의미를 가질 때 'B der A'식으로 번역이 된다는 것을 말한다.<sup>6)</sup> 여기서 소유격 정관사는 art:[*funct\_gen(A), funct\_num(A), 2*]식으로 기술되는데 *funct\_gen(A)*는 A 변수로부터 '문법적 성'을 가져오는 함수를 말하며, *funct\_num(A)*는 A 변수로부터 '문법적 수'를 가져오는 함수를 말하며 숫자 '2'는 '문법적 격'이 2격임을 기술한 것이다. 9)에서는 'A와 B 등 N 명'이라는 명사구가 독일어에서는 'N Personen einschließlich A und B'로 번역된다는 것을 말하는 것이다.

### 3.3. 연결 번역패턴

#### 3.3.1. 연결 번역패턴의 정의

연결 번역패턴이란 한국어의 복합문(등위접속문, 내포문 등)에 대한 독일어

- 
- 5) "A=인명!와 B=인명 등 N 명 > N personen:n einschließlich A und B"의 명사구 번역 패턴에서 독일어 부분의 *personen:n*이 명사입에도 소문자로 기술된 것은 대문자는 변수로 취급하기 때문에 변수와의 구별을 위해 소문자로 기술한 것이며 이러한 소문자 명사의 첫 번째 문자는 독일어 생성시에 대문자로 자동으로 생성될 것이다.
- 6) 'A의 B'와 같은 한국어 명사구를 해당 독일어 표현으로 변환하는 명사구 번역패턴은 이러한 예 이외에도 A와 B에 등록될 수 있는 의미정보에 따라 다양한 표현으로 변환될 수 있다. 하지만 모든 가능한 경우의 의미정보를 예측해서 기술할 수 없기 때문에 DB에 의미정보 조합이 없는 경우 'A의 B'는 'B von A'로 default로 번역된다.

표현을 대응시킨 패턴을 말한다.

### 3.3.2. 연결 번역패턴의 기능

연결 번역패턴은 격틀 번역패턴을 연결하여 번역하여 주는 기능을 한다.

### 3.3.3. 연결 번역패턴의 세부분류

연결 번역패턴의 종류는 크게 VP형 연결 번역패턴, 피수식 체언형 연결 번역패턴, 고정표현형 연결 번역패턴의 3가지 형태로 다시 나눌 수 있다. 각각의 종류에 대해 살펴보면 다음과 같다.

#### 3.3.3.1. VP형 연결 번역패턴

VP형 연결 번역패턴이란, 한국어의 연결 어미가 대응되는 독일어 접속사로 번역되는 경우의 연결 번역패턴을 말한다. 예를 들면 다음과 같은 문장들이 해당된다.

- 10) 그가 학교에 가고 나는 집에 있었다. > Er ist in die Schule gegangen und ich bin zu Hause geblieben.
- 11) 그는 그 소식을 듣고 매우 놀랐다. > Er war sehr überrascht, als er die Nachricht hörte.
- 12) 그가 학교에 가면 공부를 열심히 할 것이다. > Er wird fleißig lernen, wenn er in die Schule geht.

#### 3.3.3.2. 피수식 체언형 연결 번역패턴

피수식 체언형 연결 번역패턴이란, 한국어에서 VP절의 수식을 받는 체언이 독일어에서는 접속사의 형태로 번역되는 경우의 연결 번역패턴을 말한다. 이러한 체언의 예로는 “~ㄴ 때”, “~ㄴ 동안”, “~ㄴ 채”, “~ㄴ 결과”... 등이 있다.

13) 내가 들어갔을 때 그는 책을 읽고 있었다. > Als ich eintrat, las er.

### 3.3.3.3. 고정표현형 연결 번역패턴

고정표현형 연결 번역패턴이란, 한국어의 연결 어미에 대한 일반적인 독일어 접속사 정보가 아니라 고정 표현의 형태로 번역되는 경우를 나타내는 연결 번역패턴을 말한다.

14) 그여자와 결혼하느니 차라리 나는 죽어 버리겠다. > Eher will ich sterben als sie heiraten.

### 3.3.4. 연결 번역패턴의 포맷

항목	내용
Key	한국어 대표 연결어미:분석깊이
Content	$\langle tformat \rangle \rightarrow \{ \langle kpattern \rangle \} > \{ \langle epattern \rangle \}$ $\langle kpattern \rangle \rightarrow ( \langle nd-name \rangle   \langle nd-name \rangle [ \langle fnd-name \rangle ] ) +$ $\langle epattern \rangle \rightarrow ( \langle nd-name \rangle   \langle nd-name \rangle [ \langle fnd-name \rangle ] ) +$ $\langle nd-name \rangle \rightarrow \langle symbol \rangle \langle number \rangle$ $\langle fnd-name \rangle \rightarrow ( \langle symbol \rangle \langle number \rangle : [ \langle feature \rangle ( \langle feature \rangle ) * ] ) +$ $\langle feature \rangle \rightarrow \langle attribute \rangle \langle ops \rangle [ \langle value \rangle ]$ $\langle attribute \rangle \rightarrow \langle symbol \rangle ( \langle number \rangle )$ $\langle ops \rangle \rightarrow ==   !=   **   :=$ $\langle value \rangle \rightarrow ( \langle symbol \rangle ( \langle number \rangle )   \langle hangul \rangle )$ $\langle hangul \rangle \rightarrow ( 가   가 ... )$ $\langle number \rangle \rightarrow ( 1   ...   9 ) ( 0   ...   9 ) *$

한독 연결 번역패턴이 DB에 저장되는 형식은 연결어미들을 key로 하며 한국어 복합문에 대한 독일어 대역 패턴이 기술되어 있는 content로 구성된다. 한독 연결 번역패턴의 포맷은 다음과 같다:

연결 번역패턴 포맷에 따라 3가지 연결패턴의 예들을 기술하면 다음과 같다:

10') 그가 학교에 가고 나는 집에 있었다. > Er ist in die Schule gegangen und ich bin zu Hause geblieben.

Key: 고:2\_다:1

Content: { VP1[] VP2[] > VP1[ECONJ:[eroot := [und]]] VP2 }

11') 그녀는 그 소식을 듣고 매우 놀랐다. > Er war sehr überrascht, als er die Nachricht hörte.

Key: 고:2\_다:1

Content: { VP1[VERB1:[kframe == [듣다4-1]]] VP2[VERB2:[kframe == [놀라다1-1]]] > VP2 VP1[SCONJ:[eroot := [als]]] }<sup>7)</sup>

12') 그가 학교에 가면 공부를 열심히 할 것이다. > Er wird fleißig lernen, wenn er in die Schule geht.

Key: 면:2\_다:1

Content: { VP1[] VP2[] > VP2[] VP1[SCONJ:[eroot := [wenn]]] }

13') 내가 들어갔을 때 그는 책을 읽고 있었다. > Als ich eintrat, las er.

Key: 르때:2\_다:1

Content: { VP1[] VP2[] > VP2[] VP1[SCONJ:[eroot := [als]]] }

14') 그여자와 결혼하느니 차라리 나는 죽어 버리겠다. > Eher will ich sterben als sie heiraten.

Key: 기보다:2\_다:1

Content: { VP1[] VP2[] > VP2[CCONJ:[eroot := [eher]]  
VP1[SCONJ:[eroot := [als]]]}

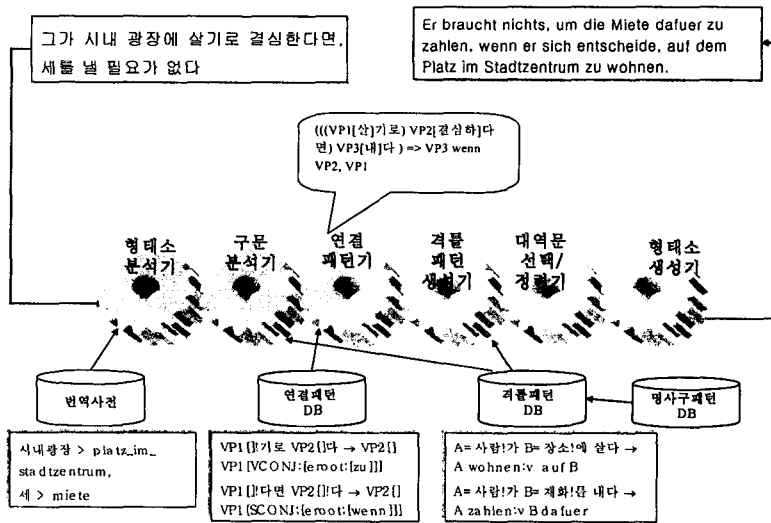
11')는 10')와 같이 “고”로 연결되는 등위문이지만 각 등위문내의 단문의 의미에 의해 10')는 나열의 의미를 11')은 이유의 의미를 나타내고 있다. 이에 따라 연결 번역패턴에서는 “고”를 “und”로 번역하는 것을 디폴트로 정하고 다른 대역어로 번역되는 경우에는 ‘VERB1:[kframe == [듣다4-1]]’와 같이 동사의미를 직접 기술함으로써 대역표현을 변별력있게 기술할 수 있다. 12')는 “면”의 접속사문을 “wenn”로 번역하는 것을 기술한 것이고, 13')은 ‘관형형 어미 + 불

7) 연결패턴에서 단문의 동사 의미는 ‘밀레니엄 새로나온 국어사전(민중서관)’에 있는 동사 어휘의 설명문을 기반으로 하여 의미코드를 부여하였다. 예를 들어 ‘듣다4-1’라는 것은 ‘듣다’ 표제어는 4가지 다의어를 가지는데 네 번째 다의어중 첫 번째 의미와 같다는 뜻이다.

완전 명사'를 독일어 접속사로 번역하는 것을 기술한 것이고, 14')는 “느니 차라리”가 “기보다”라는 대표형으로 기술되어 번역되는 예를 보인 것이다.

#### 4. 한독 자동번역 시뮬레이션

본 장에서는 한독 번역패턴이 적용되어 한국어 입력문이 독일어 대역문으로 번역되는 전체적인 번역절차를 소개하고자 한다. 전체적인 번역절차를 그리면 다음과 같다:



<그림 2> 번역 절차

한독 번역절차는 번역하고자 하는 한국어 문장에 대해 형태소 분석 및 태깅을 수행한 후, 의존 규칙과 격률 번역 패턴, 상호정보 등을 이용하여 구문구조를 분석하여 의존 구조를 생성해 낸다.(서영애 외, 2000) 생성된 의존 구조로부터 연결 번역패턴을 이용하여 한국어 격률간의 연결 관계에 대한 독일어 번역 정보를 생성해 낸다. 이렇게 해서 번역될 독일어 문장에 대한 독일어 단문

들간 연결을 결정한 후, 격틀 번역패턴과 명사구 번역패턴을 이용하여 각 격틀에 대한 독일어 문장을 생성한다. 이 후, 문장 정련과정을 거쳐 최종 독일어 문장을 생성함으로써 번역을 완료한다.

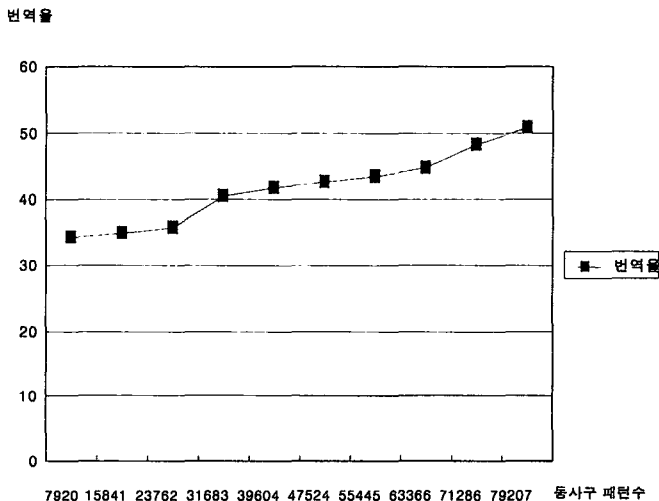
## 5. 번역 패턴 대비 한독 번역률 예측

1장에서 언급한 바와 같이 본 논문은 향후에 개발될 한독 자동번역 시스템의 필요 번역지식으로써의 한독 번역패턴에 대한 기초 연구이며 아직까지 시스템은 개발되지 않은 상태이다. 그러나 한독 번역패턴에서 기술된 번역패턴의 개념은 한국어를 출발언어로 할 때 독일어 뿐만 아니라 다른 외국어에도 적용될 수 있다. 이러한 한국어 격틀, 명사구, 연결 패턴의 재사용성은 개발 중에 있는 한영 자동번역 시스템에서 활용되었었다. 본 절에서는 한영 자동번역 시스템에서 활용한 한국어 격틀의 갯수를 응용하여 한독 자동번역 시스템의 번역률을 예측하고자 한다.<sup>8)</sup>

한영 자동번역 시스템에서 구축된 격틀 번역패턴 수는 약 7만 9천개였었다. 한영 자동번역시스템의 격틀 번역 패턴 증가 대비 번역률의 연관 관계는 KBS 9시 뉴스 방송자막 100문장(문장당 평균 단어수: 11,5단어)을 대상으로 실험하였으며 결과는 다음과 같았다.

---

8) 한영 자동번역율을 가지고 한독 자동번역율을 예측하는 것은 대역표현을 고려할 때 언어마다의 어순의 차이, 스타일의 차이로 인해 번역률에 문제가 있을 수 있다. 그러나 한영 자동번역 시스템에 활용한 한국어 격틀은 한독 자동번역 시스템에서도 동일하게 적용하기 때문에 독일어 대역표현이 모두 기술되었다고 가정할 때 한영 자동번역 시스템에서의 번역률 측정 방법은 한독 자동번역 시스템에서도 동일하게 적용될 수 있을 것이다.



<표 1> 번역 패턴 대비 번역률

<표 1>은 총 79,207개의 격틀 패턴을 대상으로 일정 시점에서 임의로 격틀 패턴을 10등분한 후에 격틀 패턴수가 증가함에 따른 번역률의 증가 현상을 보이고 있다. 초기 7,920 격틀 패턴에서의 번역률은 34%의 번역률을 보였으며 79,207개의 격틀 패턴에서의 번역률은 52%의 번역률을 보이고 있다. 번역률은 한국어 구조 분석의 애매성과 영어 생성의 오류등을 감안할 때 약 70%대에서 수렴하리라 예상하고 있으며 번역률 70%을 위해 약 100,000 격틀 패턴이 더 필요하리라 추정되고 있다(10%번역률을 향상시키기 위해 50,000패턴이 소요되고 있기 때문이다)

이상과 같은 한영 자동번역 시스템에서의 번역패턴은 한독 자동번역 시스템에서도 동일한 결과를 낼 것으로 예측되며 약18만 한독 격틀 번역패턴을 구축하게 되면 한독 자동번역도 70%의 번역률을 낼 수 있을 것이라 예측된다.



## 6. 결론

본 논문에서는 한국어 문장을 독일어로 자동번역하기 위한 한독 번역패턴의 종류를 살펴보았다. 이러한 번역패턴의 종류로서 격틀 번역패턴, 명사구 번역패턴, 연결 번역패턴을 살펴보았다. 번역패턴을 활용함으로써 얻을 수 있는 장점을 요약하면 다음과 같다:

- 높은 재사용성: 새로운 언어가 추가되더라도 출발언어에 활용한 번역패턴에 대응되는 목표언어의 번역패턴을 수정하기만 하면 됨.
- 편리한 유지 및 관리: 번역패턴에 대한 수정, 추가, 편집이 다중편집자에 의해 이루어져도 충돌되는 경우가 발생하지 않으므로 유지 및 관리하기가 편리함.
- 어휘단계에서의 기술가능: 속어 번역의 경우 어휘를 직접 기술해야 할 경우에 어휘단계에서의 기술이 가능함으로 언어쌍에 따른 언어현상을 직접 기술할 수가 있음.
- 자동번역율의 단조증가: 번역패턴의 증가에 비례하여 자동번역 시스템의 번역률 성능이 단조적으로 증가할 것임.<sup>9)</sup>

한독 자동번역 시스템의 번역률과 관련해서 한독 격틀 번역패턴이 약 18만 개가 구축되면 약70%의 한독 자동번역률을 낼 수 있다는 것도 기존의 한영 자동번역 패턴의 실험을 통해 예측해 보았다.<sup>10)</sup> 본 논문에서 기술된 다양한 번역패턴은 향후 독일어 교육 및 독일어 작문지원 시스템에 간접적으로 활용될 수 있으며 한독 전자사전에는 직접 활용할 수 있을 것이다.

9) 단조증가란 항상 증가상태에 있는 것을 의미한다.

10) 앞으로 20% 번역률 향상을 위해 100,000패턴이 더 구축되었을 때 단조적으로 번역율이 향상되는 이유는 현재 구축된 격틀은 약10,000개의 한국어 동사에 대해 구축된 격틀이며 1개의 동사 당 평균 7개 정도의 격틀이 구축되어 있는 상태이나 이 정도의 격틀은 각 논항에 들어가는 의미정보를 고려한다면 아직도 더 많이 구축되어야 한다. 그럼에도 불구하고 현재까지의 부족한 격틀을 가지고도 단조적인 번역율 향상이 실험적으로 보여졌다. 따라서 부족한 격틀의 수를 채우기 위해 앞으로 100,000개의 격틀이 추가될 때까지 번역율이 단조증가할 것으로 예측된다.

번역패턴의 장점에도 불구하고 현재 번역패턴 구축과 관련하여 계속해서 연구 및 보완해야 할 점으로는 다음을 들 수 있다.

- 격틀 번역패턴에서 필수격/수의격 논항의 정밀 정의 필요.
- 격틀 번역패턴에서 논항 어휘 의미 정보를 명사 시소러스로 대체.
- 명사구 번역패턴에서의 논항 어휘 의미 정보를 명사 시소러스로 대체.
- 연결 번역패턴에서의 격틀 의미를 동사 시소러스로 대체.
- 독일어 대역문에서 한국어 부사에 대한 독일어 대역어의 정확한 위치선정 문제.

앞으로 이상의 문제점을 중심으로 번역패턴을 계속 보완할 예정이다.

### 참고문헌

- 강현화(1998): 국어의 동사연결 구성에 대한 연구, 한국문화사.
- 권재일(1985): 국어의 복합문 구성 연구, 집문당.
- 류법모, 이태승, 이종혁, 이근배(1996): “술어중심 제약전파를 이용한 2단계 한국어 의존파서”, 한국정보과학회 1996 봄 학술발표논문집. 923-926쪽.
- 서영애, 김영길, 서광준, 최승권(2000): “용언구에 기반한 한영 기계번역 시스템: ‘CaptionEye/KE’”, 한국정보처리학회 추계 학술발표논문집 제7권 제2호.
- 전은희, 이성욱, 서정연(2001): “한국어 동사의 격틀 정보를 이용한 구문분석 후처리기”, 한글 및 한국어 정보처리학회, 445-449쪽.
- 정석원, 박의규, 나동열, 윤준태(2001): “격 관계와 상호정보를 이용한 한국어 의존 파서”, 한글 및 한국어 정보처리학회, 450-456쪽.
- 최재희(1992): 국어의 접속문 구성 연구, 탑출판사.
- 한국과학기술원(1996): 한·영 텍스트 번역 기술에 관한 연구, 시스템공학연구소.
- 홍영국, 이종혁, 이근배(1993): “의존 문법에 기반을 둔 한국어 구문 분석기”, 정보과학회 1993 봄 학술발표 논문집, 781-784쪽.
- 이민행, 최승권, 최경은(1998): “독-한 명사구 기계번역시스템의 구축”, 언어와 정보, 한국언어정보학회 제2권1호, 79-105.
- 최승권(1996): “한국어-독일어 자동번역”, 독일문학, 한국독어독문학회 37권1호, pp.323-337.

## Zusammenfassung

### Übersetzungspatternbasierte Koreanisch-Deutsche Automatische Übersetzung

Choi, Sung-Kwon(Electronics and Telecommunications Research Institute)

Seit dem Ereignis des Turmbaus zu Babel versucht man, Sprachbarriere zu überwinden. Eine dieser Bemühungen hat seit der 60er Jahre zur Anwendung von Computern in der Übersetzung geführt. Bei der automatischen Übersetzung übernimmt der Computer die Aufgabe, Texte von einer natürlichen Sprache in eine andere zu übertragen.

Die vorliegende Arbeit beschäftigt sich mit verschiedenen Arten von den sogenannten Übersetzungspatterns für die automatische Übersetzung Koreanisch-Deutsch. Zu den Übersetzungspatterns gehören u.a. Verbphrasen-Pattern, Nominalphrasen-Pattern und Verbindungspattern. Unter 'Verbphrasen-Pattern' versteht man einen gekoppelten Subkategorisierungsrahmen der Verben zweier Sprachen. 'Nominalphrasen-Pattern' verstehen sich als Übersetzungspattern der Nominalphrasen. Unter 'Verbindungs-Pattern' werden Übersetzungspattern von komplexen Sätze verstanden, wie z.B. Satzkoordinationen und Sätze mit eingebetteten Sätzen. Als Vorteil des Ansatzes zählen u.a. die Wiederverwendbarkeit bzw. das leichte Management der Patterns, die Beschreibungsmöglichkeit in der lexikalischen Ebene, die Generierung der natürlichen Zielsätze und die leichte Verbesserung der Übersetzungsqualität. Die in dieser Arbeit vorgestellten Übersetzungspatterns können als Hilfsmittel für Deutschunterrichte, Aufsatzhilfesysteme und elektronische Koreanisch-Deutsch Wörterbücher eingesetzt werden.

[검색어] 자동번역, 번역패턴

Automatische Übersetzung, Übersetzungspattern

최승권

(305-350)

대전광역시 유성구 가정동 161번지

한국전자통신연구원 언어처리연구팀

choisk@etri.re.kr