

영어 특허문서 자동번역을 위한 특허번역패턴 연구

최 승 권

(한국전자통신연구원)

1. 서론

국제적으로 기술경쟁이 치열해지면서 특허의 중요성이 강조되고 있다. 각국의 특허청에서는 특허기술의 신속한 권리와 보호를 위해 특허 출원 및 등록 심사 처리 기간을 단축하려는 노력을 하고 있다. 이러한 노력은 심사제도 및 프로세스 혁신, 특허심사관 증원, 특허행정 정보화 시스템의 고도화 등에 의해서 이루어지고 있다.

특허행정 정보화 시스템의 고도화와 관련해 특허청에서는 특허 자동번역 시스템을 이용함으로써 심사관 및 특허출원자들이 선행기술조사를 더욱 편리하고 빠르게 수행할 수 있도록 지원하고 있다.

본 논문에서는 정보통신부의 지원 하에 2005년부터 2006년도까지 개발된 영한 특허문서 자동번역 시스템¹⁾에서 영어 특허문장을 한국어로 자동번역하기

1) 본 연구는 정보통신부 및 정보통신연구진흥원의 IT 신성장동력 핵심기술개발사업의

위해 사용된 영한 특허번역패턴에 대해 기술하는 것을 목표로 한다.2) 본 논문에서 기술되는 영한 특허번역패턴은 기존의 자동번역 시스템에서 사용되었던 일반적인 번역패턴(Takeda 145, Watanabe 1370)과는 달리 특허 전문 번역가들이 특허문서 번역시 사용하는 특허 고유의 번역패턴에 대한 것이다.

영한 특허문서 자동번역 시스템은 2005년도에는 전기전자 분야를 대상으로 개발되었으며, 2006년도에는 영어 특허문서의 모든 분야³⁾를 대상으로 개발되었고, 2006년 3월부터 산업자원부의 특허지원센터(<http://www.ipac.or.kr>)에서 중소기업의 변리사 및 특허 심사관을 대상으로 영한 특허문서 자동번역 서비스를 제공하고 있다.

본 논문은 다음과 같이 구성된다: 2절에서는 영어 특허 문서에 고유하게 나타나는 언어학적인 특성을 살펴본다. 3절에서는 2절에서 기술된 특허문서의 특성 중에서 특허 전문 번역가들의 지식이 담긴 특허 번역 패턴에 대해 특허 번역 패턴의 유형 분류, 형식화, 구축 방법에 대해 기술한다. 4절에서는 특허번역 패턴의 효과를 살펴보기 위해 특허번역패턴이 적용되기 전과 후를 비교하는 실험을 기술하고자 한다.

2. 영어 특허 문서의 특징

2.1. 영어 특허 문서의 구조

영어 특허문서는 미국 특허청의 심사지침인 Manual of Patent Examining

일환으로 수행하였음 ‘ 2007-S-037, 응용특화 한중영 자동번역 기술 개발’.

- 2) 본 논문에 대해 정확하면서도 세밀하게 심사해 주신 세분의 익명의 심사위원께 감사의 말씀을 드리며, 본 논문에서 나오는 데이터 오류나 논문 전개상의 논리의 부정확함이 있다면 순전히 저자의 책임임을 밝힙니다.
- 3) 특허문서의 분야 분류는 국제적으로 표준이 되어 있는 IPC(International Patent Classification)을 따르는데 그 분류수가 59,759에 이르고 있다. IPC 분류가 너무 세분화되어 있어 산업적으로 분류할 수 있도록 국내 특허청에서는 12개의 산업 분류법을 사용하고 있다. 본 논문에서는 이 12개의 산업 분류법을 따르도록 하였다. 12개의 산업분류법은 다음과 같다(전기전자, 화학일반, 기계, 의료위생, 컴퓨터, 채광금속, 농림수산, 섬유, 음료식품, 잡화, 토목건설, 사무용품).

Procedure(MPEP)의 §608.01(a)에 따라 다음의 항목별 순서에 의거 기재할 것을 명시하고 있다.(박찬용 29)

- 1) Title of the Invention (발명의 명칭): 특허의 제목을 기록하는 곳.
- 2) Cross-Reference to Related Application (if any) (관련 출원의 인용): 특정 관련 출원의 출원번호 및 출원일을 기재하는 곳.
- 3) Background of the Invention (발명의 배경)
 - 가) Field of the Invention (발명의 분야): 특허가 적용되는 분야 및 발명의 간단한 성질을 쓰는 곳.
 - 나) Description of the Prior Art (종래 기술의 설명): 본 특허에 앞서는 선행 기술에 대한 설명을 기술하는 곳.
- 4) Summary of the Invention (발명의 요약): 특허의 정확한 성질, 작용 및 목적을 기술하는 곳.
- 5) Brief Description of the Drawing(s) (도면의 간단한 설명): 특허의 도면에 대한 설명을 적는 곳.
- 6) Description of the Preferred Embodiment(s) (바람직한 실시예의 상세한 설명): 특허에 속하는 기술 분야에서의 숙련자, 또는 해당 기술에 밀접한 관련이 있는 숙련자라면 누구든지 실시할 수 있도록 특허의 내용을 기술하는 곳.
- 7) Claim(s) (특허 청구의 범위): 발명을 특허권으로 하여 법적인 보호를 받기 위해 발명을 명확히 한정하는 부분. 따라서 특허에서 가장 중요한 부분이라고 할 수 있다.
- 8) Abstract of the Disclosure (개시 내용의 요약): 특허 관계자 뿐 아니라, 일반 기술자도 신속하게 출원 발명의 내용을 이해할 수 있도록 발명의 요약을 기술하는 곳.

2.2. 영어 특허 문서의 단어 특성

웹과 같은 일반 영어 문장과 비교할 때, 영어 특허 문장에 나타나는 단어들은 영어 특허문서 특유의 특징을 나타낸다. 이와 같은 영어 특허 단어의 형태적 특성을 나열하면 다음과 같다:

9) 특히 문장에서는 기호 단어, 수식, 약어, 고유명사와 하이픈 연결 단어가 많이 나타난다. 예를 들어, 다음의 (1)에서 key-frame은 하이픈 연결 단어의 예이며, src = field는 하나의 수식으로써 인식되어야 하는 경우이다.

(1) The filename of ach key-frame is specified by an src = field.

10) 거의 모든 ‘~ed’ 형태의 단어들이 과거형보다는 과거분사로 사용된다. 예를 들어 다음의 (2)에서 arranged는 주동사가 아니라 앞의 명사를 수식하는 과거분사로써 사용된 경우이다.

(2) Constituent devices arranged in parallel in each-arm of the second power converter 2.

11) 대명사 출현빈도가 일반 도메인보다 적으며, 특히 인칭대명사는 거의 나타나지 않는다.

12) ‘said’가 특히 문서에서는 거의 모두 형용사로 사용되며 ‘die’는 동사보다는 명사로 대부분 사용된다. 예를 들어 다음의 (3)과 (4)에서 said는 형용사로 die는 동사가 아닌 명사 ‘다이’로 사용된 경우이다.

(3) Typically, the feed comprises at least 20 wt % of said alkene, such as from 20 to 80 wt % of said alkene.

(4) Preferably the die 32 is a generally rectangular bar

13) 일반 도메인에서는 명사 앞에 수사가 오는 반면에, 특허에서는 명사 뒤에 수사가 오는 형태가 빈번하다. 예를 들어 다음의 (5)에서 212는 algorithm의 청구항 번호를 나타내는 경우이다.

(5) the selection algorithm 212 processes 3 selection

2.3. 영어 특허 문장의 구문적 특성

웹과 같은 일반 영어 문장과 비교할 때, 영어 특허 문장의 구문적 특성은 다음과 같은 특성을 갖는다:

병렬 및 복잡한 수식에 의한 장문이 많음: 특허에서 40단어 이상의 장문이 자주 등장하며, 특히 특허에서 중요한 정보를 담고 있는 요약문이나 청구항의 경우 평균 약 35 단어이다. (6)은 컴퓨터 분야에서 69 단어로 구성되어 있는 청구항의 예이다.

(6) logic to derive a plurality of sets of image attribute data from a test region, each set relating to said test region scaled by a respective different scaling factor from a geometric progression of scaling factors, said progression being such that each scaling factor is related to a next scaling factor by a predetermined multiplicative factor, said factor being the same across the whole progression of scaling factors;

15) 전치사구, 분사구의 부착에 있어서 편향성을 나타냄: for 전치사구 및 현재 분사구 수식의 경우 동사구보다는 명사구 부착의 편향성을 나타냄. (7)에서 employing으로 구성되는 구문이 shows를 수식하기보다는 apparatus를 수식하는 편향성을 보이고 있다.

(7) FIG. 8 shows a typical image forming apparatus employing a developing apparatus of the above described rotary type.

16) 독립 분사 형태의 표현을 많이 사용함: NP VP-ing 형태의 문장이 앞에서 언급된 문장에 대해 더욱 상세히 설명하기 위해 자주 사용됨. (8)에서 밑줄 친 the label comprising은 앞에서 언급된 one spectrally-encoded label을 더욱 상세히 설명하고 있다.

(8) at least one spectrally-encoded label, the label comprising a spectrally-selective optical element having a label spectral signature, wherein the label spectral signature is determined according to a spectral-encoding scheme so as to represent predetermined label information within the spectral encoding scheme, and the label emits output light in response to input light selected by the label

spectral signature of the optical element;

17) 일반 도메인에서 많이 사용되는 시간부사구가 사용되지 않고, 도치, 생략 등의 특수한 구문이 비교적 적게 사용됨.

2.4. 영어 특허 문서의 고유한 번역 패턴들

특허문서에서는 일반 웹문서와 달리 특허문서에서 고유하게 나타나는 상투적인 표현들이나 반복적으로 나타나는 표현들이 많다.(Shinmori 57) 이러한 상투적이거나 반복적으로 나타나는 특허 고유의 표현들은 특허문서를 구성하는 항목별로 다르게 나타나는데 대표적인 특허문서 항목별 예들을 살펴보면 다음과 같다:

18) Background of the Invention (발명의 배경): 도면을 참조하여 내용을 간략히 설명하는 문장이 등장한다

(9) This invention relates to ~, and more specifically ... => 본 발명은 ~에 관한 것이며, 보다 상세히 설명하면 ...에 관한 것이다.

(10) This invention is applicable particularly, but not exclusively to ~ => 본 발명은 특히 ~에 적용 가능하지만, 그것에만 한정되지 않는다.

(11) As shown in FIG.~, => 도 ~에서 도시된 바와 같이,

19) Summary of the Invention (발명의 요약): 제안되는 발명의 목적을 설명하는 문장이 일반적으로 제일 처음 나온다.

(12) Accordingly, it is an object of the present invention to provide ~ : 따라서, ~을 제공하는 것이 본 발명의 목적이다.

20) Brief Description of the Drawing(s) (도면의 간단한 설명): 도면 부호 및 그에 대한 설명이 반복적으로 나타난다.

(13) Fig.~ illustrates ... : 도 ~는 ...을 나타낸다.

21) Claim(s) (특허 청구의 범위): 주로 현재분사/동명사를 동반한 명사구가 대부분을 이룬다

(14) What is claimed is : 청구항

(15) The method of ~ comprising the steps of: 다음의 단계들로 구성되는 ~는 방법

22) Abstract of the Disclosure (개시 내용의 요약): 특허 대상 기술을 소개하는 문장이 특허문서마다 등장한다

(16) The present invention relates to ~ => 본 발명은 ~에 관한 것이다.

(17) A method for ~ comprises the step of, 1) ..., 2) ... => ~기 위한 방법은 1) ..., 2) ...의 단계로 구성된다.

본 논문에서는 위에서 소개된 특허 문서의 언어학적 특징 중에서 특허 문서에 고유하게 나타나는 번역패턴 표현들에 대해 다음 절에서 상세히 기술하고자 한다. 특허 문서의 다른 언어학적 특징들은 본 논문의 논의에서는 제외하고자 한다.

3. 영한 자동번역용 특허번역패턴

자동번역 시스템에서 번역패턴이 유용한 이유는 크게 다음의 두 가지를 들 수 있다. 첫째는 번역패턴에 원문과 번역문이 동시에 기술되기 때문에 원문의 패턴이 선택되자마자 번역문이 만들어지는 동시성(Synchronization) 효과가 있다(Shieber 253). 둘째는 여러 개의 단어나 구를 하나의 분석 단위로 보기 때문에 원문의 패턴이 일치만 되면 문장을 분석하는 데 단어를 대상으로 하였을 때보다 분석 단위가 훨씬 작아지므로 문장 분석 속도가 빨라져 자동 번역 속도도 빨라진다(Watanabe 1370).

본 절에서는 특허 전문 번역가들이 특허문서 번역시 사용하는 특허번역패턴에 대해 특허번역패턴의 유형 분류, 자동번역 시스템에서의 형식화 방법, 특허번역패턴을 특허문서로부터 자동으로 추출하여 반자동으로 구축하는 방법에

대해 기술하고자 한다.

3.1. 특허번역패턴의 유형 분류

특허번역패턴은 패턴이 담당하는 표현 범위에 따라 크게 세 가지로 분류할 수 있다. 첫 번째는 구문번역패턴으로써 명사구, 전치사구, 부사구 등 동사를 포함하지 않는 구 단위의 특허번역패턴을 말하며, 두 번째는 동사구 번역패턴으로써 하나의 동사를 포함하는 동사 숙어와 같은 특허번역패턴을 말한다. 세 번째는 문장 전체를 대상으로 하는 문장번역패턴이다.

3.1.1. 구문번역패턴

구문번역패턴은 명사구, 전치사구, 부사구 등 동사를 포함하지 않는 구문 단위에 적용되는 번역패턴을 말하며, 구문 단위의 자연스러운 번역을 제공하는 것을 목적으로 한다.

- (18) ‘영어 원문’ In one embodiment, the present braking system comprises a pair of brake members 12 coupled to a frame 14 of a wheelchair for selectively abutting opposed rear wheels.
(<http://www.patyellow.co.kr/>)

‘번역가 번역문’ 일 실시예에 있어서, 본 발명의 브레이크 장치는, 프레임(14)에 결합되어 대향하는 뒷바퀴(후륜)와 선택적으로 접하는 한 쌍의 브레이크 부재(12)를 구비하고 있다.

‘구문 패턴’ in one embodiment : 일 실시예에 있어서

(18)에서 in one embodiment가 구문번역패턴의 하나의 예로 볼 수 있는데, 그 이유는 in one embodiment가 특허 번역에서만 ‘일 실시예에 있어서’로 번역되기 때문이다.

- (19) ‘영어 원문’ As is well known in the art, a conventional wheelchair has two large, rear wheels on either side of a seat while small caster

wheels are positioned on either side of a footrest.
(<http://www.patyellow.co.kr/>)

‘번역가 번역문’ 해당 기술분야에서 잘 알려진 것처럼, 종래의 휠체어는 시트의 양측에 2개의 큰 후륜을 가지면서, 발판의 양측에는 작은 바퀴들이 장착되어 있다.

‘구문 패턴’ as is well known in the art : 해당 기술 분야에서 잘 알려진 바와 같이

(19)에서 ‘as is well known in the art’가 구문번역패턴의 하나의 예로 볼 수 있는데, 그 이유는 특허 문서에서 반복적으로 자주 사용되며 일반 번역 패턴에서는 볼 수 없는 번역패턴이기 때문이다.

3.1.2. 동사구 번역패턴

동사구 번역패턴은 동사로 시작하는 동사구에 적용되는 번역패턴을 말하며, 동사구 단위의 자연스러운 번역을 제공하는 것을 목적으로 한다.

(20) ‘영어 원문’ The threads may be formed on the outside of the cross member, so that the cross member screws into the brake member.
(<http://www.patyellow.co.kr/>)

‘번역가 번역문’ 나사산이 크로스부재의 외측에 형성됨으로써, 이 크로스부재가 브레이크부재에 나사결합될 수 있다.

‘패턴’ screws into : 에 나사결합되다.

(20)에서 ‘screw into’가 동사구 번역패턴의 하나의 예로 볼 수 있는데, 그 이유는 특허 문서에서 반복적으로 자주 사용되며 일반 번역 패턴에서는 볼 수 없는 번역 패턴이기 때문이다.

(21) ‘영어 원문’ Chair 10 is optionally provided with a central opening 11 for facilitating toilet use. (<http://www.patyellow.co.kr/>)

‘번역가 번역문’ 의자(10)는 화장실 사용을 용이하게 하기 위하여 필요에 따라 중앙개구부(11)를 구비하고 있다.

‘패턴’ be provided with : 를 구비하다.

(21)에서 ‘be provided with’가 일반 번역패턴에서와는 달리 특히 문서에서 반복적으로 자주 사용되며 특히 번역시 고유한 대역어를 필요로 하기 때문에 동사구 번역패턴으로 간주할 수 있다.

3.1.3. 문장 번역패턴

문장 번역패턴은 문장 전체에 대해 적용되는 번역패턴으로써, 문장 전체에 대해 자연스럽게 번역하는 것을 목적으로 한다.

(22) ‘영어 원문’ There is a need for an image analysis technique that is more easily implemented. (<http://www.patyellow.co.kr/>)

‘번역가 번역문’ 보다 용이하게 실현되는 화상분석수법이 요구되고 있다.

‘패턴’ there is a need for ~ -> ~ 이 요구되고 있다

(22)에서 ‘there is a need for ~’는 특히 문서에서 반복적으로 자주 나타나기 때문에 문장 전체에 대해 자연스러운 번역을 위해 문장 번역패턴으로 구축하는게 좋다.

(23) ‘영어 원문’ This invention relates to the field of wheelchairs and more particularly to a kit which allows unadjustable straight backed wheelchair, such as a folding wheelchair, to be converted into an adjustable back wheelchair, wherein the back of wheelchair can be positioned along a continuum between a fully reclined position and an upright position. (<http://www.patyellow.co.kr/>)

‘번역가 번역문’ 본 발명은, 휠체어의 분야에 관한 것으로, 더욱 상세

하계는 접이식 휠체어처럼 등받이가 직각으로 조절할 수 없는 휠체어를 휠체어의 등받이가 완전하게 눕혀지는 위치와 직각으로 일어선 위치 사이를 연결한 부분을 따라 휠체어 등받이 위치를 결정할 수 있는 등받이 조절가능 휠체어로 변환할 수 있도록 한 시트에 관한 것이다. ‘패턴’ this invention relates to ~ and more particularly to ... : 본 발명은, ~에 관한 것으로, 더욱 상세하게는 ...에 관한 것이다.

(23)은 패턴을 구성하는 요소가 불연속적으로 나타날 수 있음을 보이는 문장 패턴으로써, ‘this invention relates to’와 ‘and more particularly to’가 특허 문서에서 동시에 자주 나타나는 표현이므로 특허 번역의 자연스러움을 제공하기 위해 문장 번역패턴으로써 구축할 수 있다.

3.2. 특허번역패턴의 형식화

앞서 언급한 특허번역패턴을 컴퓨터가 인식하기 위해서 영한 특허번역패턴은 데이터베이스에 등록되고 있다. 특허번역패턴의 데이터베이스는 다음과 같은 형식으로 이루어져 있다:

<표 1> 특허번역패턴의 데이터베이스 형식

Key	@상위노드_하위노드들
Content	{ 원문부 } -> { 대역부 }

표 1에서 ‘@상위노드_하위노드들’은 패턴을 찾기 위한 key를 의미한다. 상위노드는 해당 패턴이 어떤 패턴의 부류에 속하는 가를 알 수 있는 통사적 노드(syntactic node)로써 구문 번역패턴에 속하는 상위 노드는 명사구(NP), 전치사구(PP), 부사구(ADVP), 형용사구(ADJP)이며, 동사구 번역패턴에 속하는 상위 노드는 동사구(VP)이며, 문장 번역패턴에 속하는 상위 노드는 문장(S)이다. 하위노드는 상위노드를 형성하는 요소들로써, 어휘, 품사, 구문노드들로 구성되며, 반복적으로 사용할 수 있다. 하위노드들의 연결은 ‘_’로 이루어진다. 어휘는

영어 단어를 의미하며, 품사에는 명사(NOUN), 수사(NUM), 동사(VERB), 부사(ADV), 형용사(ADJ), 대명사(PRON), 한정사(DET), 접속사(CONJ), 전치사(PREP)들이 속하며, 구문 노드에는 명사구(NP), 전치사구(PP), 부사구(ADVP), 형용사구(ADJP), 동사구(VP), 문장(S)이 속한다. 여기서 품사와 구문 노드는 변수(variable)로 간주된다.

표 1에서 Content는 Key에 대한 번역패턴 내용을 기술하는 부분으로써 ‘{ 원문부 } -> { 대역부 }’의 형식으로 이루어진다. 원문부는 영어 원문 패턴을 말하며, 대역부는 원문부에 대응되는 한국어 패턴을 말한다. 원문부에는 Key에서 기술된 하위노드들이 ‘_’없이 기술되며, 가장 핵심이 되는 어휘에 !를 표시하여, 원문부의 핵심 어휘의 정보가 대역부의 핵심 어휘로 전달되도록 하였다. 대역부에 기술되는 한국어 패턴은 원문부에 기술된 어휘들에 대응되는 한국어 어휘를 기술하며, 원문부에서 품사나 구문노드로 기술된 변수들은 대역부에서 반드시 반복되어 적절한 위치에 기술되어야 한다.

3.2.1. 구문 번역패턴

3.1.1.절에서 구문 번역패턴 후보로 기술되었던 ‘in one embodiment : 일 실시예에 있어서’는 전치사로 시작하는 표현이므로 상위노드는 전치사구(PP)가 되며, 하위노드들은 ‘in_one_embodiment’가 된다. 원문부에는 KEY에 기술된 하위노드들에서 ‘_’를 지운 형태를 기술하고, 핵심어는 전치사이므로 in에 핵심어 표시인 !를 기술한다. 대역부에는 ‘일 실시예에 있어서’의 표현을 기술하고 ‘in!’에 대응되는 표현으로써 ‘있어서!’로 핵심어 정보를 이동할 수 있도록 기술한다. 이와 같은 기술은 다음의 (24)에서 볼 수 있다.

(24) @PP_in_one_embodiment

{ in! one embodiment } -> { 일 실시예에 있어서! }

3.1.1.절의 (19)에서 언급하였던 구문 번역 패턴인 ‘as is well known in the art : 해당 기술 분야에서 잘 알려진 바와 같이’도 위의 (24)에서와 같이 동일한 방식으로 구문 번역패턴을 만들면 (25)와 같이 부사구(ADVP)로 만들 수 있다.

(25) @ADVP_as_is_well_known_in_the_art

{ as! is well known in the art } -> { 해당 기술 분야에서 잘 알려진
바와 같이! }

3.2.2. 동사구 번역패턴

(20)에서 언급되었던 동사구 번역패턴 ‘screws into : 에 나사결합되다’를
데이터베이스에서는 다음과 같이 등록된다:

(26) @VP_screw_into_NP

{ screw! into NP } -> { NP:‘에’ 나사결합되! }

앞서 언급한 것처럼 동사구 패턴의 상위노드는 VP이며 ‘screw into’의 목적
어로 명사구(NP)를 기술하여 원문부와 대역부를 기술하면 (26)처럼 만들어진다.
대역부에서 조사는 명사구 변수 옆에 ‘:’에 기술하도록 한다.

앞서 기술된 (21)은 수동태 형의 동사구 패턴에 대한 것이다. 수동태 패턴
은 ‘VN’이라는 표식을 하위노드의 동사에 부착함으로써 이것을 통해 수동태
동사구 패턴임을 알 수 있다. (21)을 형식화하면 (27)과 같은 수동태 동사구 패
턴으로 된다.

(27) @VP_provide‘VN’_with_NP

{ provide! with NP } -> { NP:‘를’ 구비하! }

3.2.3. 문장 번역패턴

앞서 기술되었던 문장 번역패턴인 (22)와 (23)을 데이터베이스 형식으로 기
술하면 각각 다음과 같다:

(28) @S_there_be_a_need_for_NP

{ there be! a need for NP } -> { NP:‘가’ 요구되고 있! }

(29) @S_this_invention_relate_to_NP_and_more_particularly_to_NP

{ this invention relate! to NP1 and more particularly to NP2 } ->

{ 본 발명은 NP1:'에' 관한 것으로, 더욱 상세하게는 NP2:'에' 관한 것이! }

(28)에서 명사구(NP)는 'for' 이하에 나오는 전체 문자열을 표현하는 변수이며, (29)에서 NP1, NP2는 두 개의 명사구를 구분하기 위한 변수를 말하며, NP1 은 'to'와 'and' 사이에 나타나는 명사구를 NP2는 to 이하에 나오는 명사구를 의미한다.

3.3. 특허번역패턴의 구축

앞서 기술된 특허번역패턴은 2001년-2005년 사이에 미국에서 출원.등록된 100만 건 정도의 영어 특허 문서를 대상으로 특허 전문 번역가 3인에 의해서 2년 동안 구축된 것이다. 구축 방법은 1) 100만 건의 영어 특허 문서로부터 3단어 이상의 연속된 문자열을 자동으로 추출하고, 2) 추출된 3단어 이상의 연속된 문자열을 고빈도순으로 정렬한 후, 3) 기존에 보유하고 있는 일반 번역패턴과 일치하는 문자열과 일치하지 않는 문자열로 구분한다. 4) 일반 번역패턴과 일치하는 문자열은 일반 번역패턴에 부착되어 있는 대역어를 부착한 후, 특허 전문 번역사가 특허에 맞는 대역어 감수를 실시하여 특허 번역 패턴으로 구축하며 5) 일반 번역 패턴과 일치하지 않는 문자열은 특허 전문 번역사가 대역어를 수동으로 부착하였다. 구축된 특허 번역 패턴 수는 다음과 같았다:

<표 2> 특허 번역 패턴의 구축수

구분	특허 번역 패턴
구문 번역 패턴	1,257
동사구 번역 패턴	255
문장 번역 패턴	642
총계	2,154

특허 번역 패턴은 특허 문서 번역 경험이 없는 일반 번역가나 사전 편집자들에 의해 구축될 수가 없다. 왜냐하면 발명과 관련된 특허의 특성상 해당 전문 분야에 대한 전문 지식을 가지고 있어야 하며 또한 특허 표현법에 맞게 기술할 수 있어야 하기 때문이다. 이와 같은 이유로 표 2에서 제시된 특허번역패턴은 특허 전문 번역가에 의해서만 구축된 것이다.

4. 특허 번역 패턴 적용 전후의 번역률 평가

본 절에서는 특허번역패턴과 영한 특허문서 자동번역 시스템의 번역률과의 상관 관계를 살펴보고자 한다. 특허 문서의 자동 번역된 번역 결과를 평가하기 위해 사용한 평가 문장, 평가 방법, 평가 기준은 다음과 같다:

23) 평가 문장: 영어 특허문서 100만 여건 중에서 주요 5개 산업분야(기계, 전기전자, 화학일반, 의료위생, 컴퓨터)에 대해 각 산업 분야별로 임의로 1,000개의 문서를 선정하고, 각 산업 분야별로 100 문장을 각각 자동 추출하였다.

24) 평가 방법:

7인의 특허 전문 번역가에게 평가 점수 부여 기준을 교육한 후 평가 기준에 따라 각자 자동 번역 결과에 대해 문장별로 평가 점수를 부여하게 하고 각 문장별로 최고 최저점을 제외한 5개 점수의 평균 합에 대해 번역률을 계산하였다. (예: 334점 / (4점 x 100문장) = 83.50%)

25) 평가 기준

평가 기준은 다음의 표에서와 같은 점수 부여 기준을 따랐다.

<표 3> 평가 점수 부여 기준

점수	평가 기준
4.0	원어문의 의미가 그대로 전달된 경우
3.5	복문에서, 문장의 동사구가 정확히 전달되어 문장의 전체적인 의미의 골격이 전달되지만, 동사를 제외한 1-2단어의 대역어가 잘못된 경우
3.0	문장의 동사구가 정확히 전달되어 문장의 전체적인 의미의 골격이 전달되지만, 동사를 제외한 3-4단어의 대역어가 잘못된 경우
2.5	하나의 동사구라도 정확히 번역되어 부분적으로 문장의 의미를 전달할 경우
2.0	하나 이상의 구가 정확히 번역되지만 전체적인 문장의 의미를 파악하기 어려운 경우
1.0	문장 중에 하나의 단어 또는 구라도 정확히 번역된 경우
0.0	번역문 출력이 안 된 경우

위와 같은 평가 기준에 따라 각 분야별로 선정된 100 문장에 대해 영한 특허 자동번역 시스템을 평가한 결과는 다음과 같았다:

(표 4에서 특허번역패턴 포함 여부를 제외하고는 자동번역 시스템을 구성하는 각 모듈(예: 형태소 태거, 구조분석기, 대역어 선택기 등)은 이미 특허문서에 특화되어 있는 상태에서 번역률을 평가하였다)

<표 4> 번역률 평가 결과 (평가일:2006.12.13)

분야	문장당 평균단어수	전체번역률		체감번역률 ⁴⁾	
		특허 번역 패턴 포함	특허 번역 패턴 미포함	특허 번역 패턴 포함	특허 번역 패턴 미포함
기계	30.34	83.50%	81.28%	85.00%	79.00%
전기전자	28.19	82.20%	80.96%	88.00%	74.00%
화학일반	29.67	82.20%	81.50%	91.00%	76.00%
의료위생	26.75	81.63%	80.63%	86.00%	72.00%
컴퓨터	25.49	82.63%	81.80%	88.00%	79.00%
평균	28.09	82.43%	81.23%	87.60%	76.00%

표 4에서 알 수 있는 것은 특허번역패턴을 포함했을 때가 포함하지 않았을 때보다 전체번역률에서는 1.20%가 향상되었으며, 체감번역률에서는 11.60%가 향상되었다는 것이다. 이것은 특허 자동번역에서 특허 번역 패턴이 사용자에게 실제적인 번역의 이해도를 높여주는 중요한 번역 지식이라는 것을 의미하는 것이다.

100문장의 평가문에서 특허 번역 패턴이 적용되었을 때와 적용되지 않았을 때의 평가 예문의 번역 예를 보면 다음과 같다:

(30가) ‘영어원문’ 3. Detonator according to claim 1 or 2, characterized in that the plastic of the igniter support (4) is doped with mica.

(30나) ‘특허번역패턴 미적용 결과’ 3. 이그나이터 서포트 (4)의 플라스틱이 운모로 도핑되는 점에서 특징지워지는 청구항 1 또는 2에 따른 기폭장치.

(30다) ‘특허번역패턴 적용 결과’ 3. 제 1 또는 2 항에 따른 기폭장치에 있어서, 이그나이터 서포트 (4)의 플라스틱은 운모로 도핑되는 것을 특

4) 표 4에서 전체번역률은 0~4점까지의 평가 점수의 총점에 대한 번역률을 말하며, 체감번역률은 평가 점수에서 3점 이상의 문장에 대한 번역률을 말한다. 체감번역률은 일반 사용자 입장에서 자동 번역 결과를 초벌번역용으로 활용할 수 있는지의 수준을 수치화 한 것으로, 번역 결과가 이해가 되느냐 안 되느냐의 O, X로 크게 구분하는 번역률이라 할 수 있다.

징으로 한다.

(30라) ‘적용된 특허번역패턴’

```
@S_NP_COMMA_characterized_in_that_S
{ NP1 COMMA characterized! in that S1 } -> { NP1:‘에’ 있어서
COMMA S1:‘는_것을’ 특징으로 하! }
@NP_NP_according_to_claim_NUM
{ NP1! according to claim NUM1 } -> { 제 NUM1 항에 따른
NP1! }
```

(30가)의 원문에 대해 (30나)와 (30다)는 ‘characterized in that’의 앞뒤 어절을 어떻게 번역했는가에 따른 차이를 보이고 있다. (30나)와 (30다)는 특허 전문 번역사들에게 모두 이해 가능한 번역 결과이다. 그러나 특허 전문 번역사들은 (30나)보다는 (30다)를 선호한다. 왜냐하면 특허 전문 번역사들은 특허 번역시에 ‘A, characterized in that B’ 구문에 대해서는 (30라)의 특허번역패턴과 같이 ‘A에 있어서, B는 것을 특징으로 한다’와 같은 번역에 익숙해 있기 때문이다.

(31가) ‘영어원문’ 2. The data reproduction device as claimed in claim 1, wherein the recording medium has recording tracks each having a phase detection region and an information data start detection region to which regions a predetermined pattern is written, and

(31나) ‘특허번역패턴 미적용결과’ 2. 휴대용 저장 매체가 예정 패턴이 쓰여진 영역에 각각 위상 검출 영역과 정보 데이터 시작 검출 영역을 가지는 리코딩 트랙을 가진 청구항 1에서 청구된 것처럼 데이터 재생 장치, 그리고

(31다) ‘특허번역패턴 적용결과’ 2. 제 1 항의 데이터 재생 장치에 있어서, 기록 매체는 리코딩 트랙을 가지며, 각각의 리코딩 트랙은 예정 패턴이 쓰여진 영역에 위상 검출 영역과 정보 데이터 시작 검출 영역을 구비한다, 그리고

(31라) ‘적용된 특허번역 패턴’

```
@S_NP_COMMA_wherein_S
```

{ NP1 COMMA wherein S1! } -> { NP1:‘에’ 있어서 COMMA S1! }
 @S_NP_have_NP_each_having_NP
 { NP1 have! NP2 each having NP3 } -> { NP1:‘은’ NP2:‘을’ 가지
 며, 각각의 NP2:‘은’ NP3:‘을’ 구비하! }
 @NP_the_NP_as_claimed_in_claim_NUM
 { the NP1! as claimed in claim NUM1 } -> { 제 NUM1 항의
 NP1! }

(31가)의 원문에 대해 특허 전문 번역사는 특허번역패턴이 적용되지 않은 번역 결과인 (31나)에 대해서는 이해하는데 어려워하였으나, 특허번역패턴이 적용되어 나온 번역 결과인 (31다)에 대해서는 쉽게 이해를 하였다. 왜냐하면 ‘A each having B’, ‘A, wherein B’와 같은 문장을 번역할 때 특허 전문 번역사들은 위에서 언급한 특허번역패턴과 같이 습관적으로 번역하기 때문이다.

5. 결론

본 논문에서는 영한 특허문서 자동번역에서 영한 특허번역패턴이 특허 전문 번역사들이 번역 결과를 이해하는 데 있어 중요한 역할을 한다는 것을 살펴 보았다. 그에 대한 확인은 영어 특허 문서를 대상으로 번역한 번역 품질에서 알 수 있었다. 5개의 특허 분야에 대해 각각 100 문장을 선정하여 특허번역패턴을 포함했을 때와 포함하지 않았을 때의 번역률의 차이를 분석하였을 때, 5개 특허 분야 전체 평균은 전체 번역률에서는 포함하였을 때가 82.43%, 포함하지 않았을 때가 81.23%로써 1.20%의 차이가 나는 것을 확인하였으며, 체감번역률에서는 포함하였을 때가 87.60%, 포함하지 않았을 때가 76.00%로써 11.60%의 차이가 오로지 특허번역패턴에 의해 나는 것을 확인할 수 있었다.

특허번역패턴은 대용량의 특허 문서를 대상으로 특허번역패턴 가능한 후보들을 자동으로 추출한 후, 특허 번역의 전문적 지식을 가지고 있는 특허 전문 번역사에 의해 수동으로 구축되고 있다. 향후 더욱 개선되어야 할 부분은 특허 번역 패턴 후보를 자동으로 추출하는 방법이 향상되어야 하는 것이다.

참고문헌

- 박찬용. 1996. 『미국특허청구범위에 관한 고찰-영문특허명세서를 중심으로-』.
연세대학교 특허법무대학원 석사논문.
- 특허포탈. 2007. <http://www.patyellow.co.kr>
- Shieber, S.M. and Schabes, Y. 1990. "Synchronous Tree-Adjoining Grammars".
Proc. of the 13th International Conference on Computational Linguistics.
253-258.
- Shinmori, A., Manabu, O., Yuzo, M. and Makoto, I. 2003. "Patent Claim
Processing for Readability - Structure Analysis and Term Explanation",
ACL-2003 Workshop on Patent Corpus Processing.
- Takeda, K. 1996. "Pattern-Based Context-Free Grammar for Machine
Translation", *Proceedings of 34th ACL.* 144-151.
- Watanabe, H., & Takeda, K. 1998. "A Pattern-Based Machine Translation
System Extended by Example-Based Processing", *COLING 1998.*
1369-1373.

[Abstract]

A Study of Patent Translation Patterns for Machine Translation of English Patent Documents

Choi, Sung-Kwon

(Electronics and Telecommunications Research Institute)

It is the purpose of this paper to describe English-Korean patent translation patterns making the automatic translation of patent documents possible in English-Korean machine translation system for patent documents.

According to the representation scope that a pattern covers, English-Korean patent translation patterns can be classified into three types.

First, it is the phrasal translation pattern. We have now 1,257 phrasal translation patterns as the pattern covering a phrase. The second pattern is a verbal phrase translation pattern. There are 255 verbal phrase translation patterns as a kind of the idiom of verb phrase. The last pattern is a sentence translation pattern. It consists of 642 patterns as the pattern covering a sentence whole.

The significance of the patent translation pattern was shown in the translation quality of sentences translated from English patent documents into Korean patent documents.

When sentences were translated by the machine translation system including patent translation patterns, the whole translation accuracy from 0 to 4 as evaluation grade was 82.43% and the translation rate over 3 grade 87.60%. But when the patent translation patterns were not included, the whole translation rate was 81.23% and the translation rate over 3 grade 76.00%.

From experimental result of translation quality with or without the patent

translation patterns, we could know that the difference 1.20% of the whole translation rate and the difference 11.60% of the translation rate over 3 grade were made only by the patent-specific translation patterns.

Presently, the English-Korean machine translation system for patent documents described in this paper has been installed in International Patent Assistance Center (<http://www.ipac.or.kr>) of the Ministry of Commerce, Industry and Energy and provides small and medium-sized enterprises with English-Korean patent translation service.

▶ Key Words: machine translation, patent translation, patent translation patterns, English-Korean translation, English-Korean machine translation, English-Korean patent translation, English-Korean machine translation for patent documents

최승권

한국전자통신연구원 책임연구원

choisk@etri.re.kr

관심분야: 자동번역, 자동통역

논문투고일: 2007년 4월 9일

심사완료일: 2007년 5월 30일

게재확정일: 2007년 6월 13일