

# 번역수월성의 관점에서 본 통제독일어, 통제영어, 통제한국어의 비교연구

홍문표 (성균관대)

## 1. 서론

본 논문은 통제독일어, 통제영어, 통제한국어를 번역수월성 Übersetzbarkeit의 관점에서 비교, 분석하고, 통제언어 Kontrollierte Sprache 연구의 궁극적인 목표인 통제한국어 설계에 대한 방향을 설정하는 것이 연구의 목표이다.

한국에서도 제조물책임법 Product Liability<sup>1</sup>이 2002년 7월부터 시행된 이후로 제품에 수반되는 문서의 품질에 대한 관심이 날로 높아지고 있다. 문서의 품질이란 문서의 가독성, 번역수월성 및 심미성과 주로 관계되는 개념이다. 한국에서 개발되어 외국으로 수출되는 제품에 수반되는 문서의 경우, 한국어 원문의 품질이 번역의 품질을 결정하는 주요 요인이 된다.<sup>2</sup> 한국어 원문의 품질은 문서내의 형식/포맷의 일관성, 용어의 일관성, 문법성, 스타일의 명료성 등에 의해 결정된다. 문서의 형식이나 포맷의 일관성 등을 위해서 현재는 주로 XML기반의 'single-sourcing' 기법 등을 사용하고 있으나, 이것은 문서의 형식이나 포맷에만 관련될 뿐, 문서를 구성하는 텍스트, 또는 개별문장의 언어학적 수준을 보장해주는 것은 못한다.<sup>3</sup> 이러한 문제를 해결하기 위해 90년대 중반부터 주로 미국과 유럽에서 통제언어에 대한 논의가 시작되었다.<sup>4</sup> 다국어 사회인 유럽 내의 언어간 번

\* 이 논문은 성균관대학교의 2007학년도 성균학술연구비에 의하여 연구되었음.

1 제조물 책임법은 제품의 제조, 설계, 표시 등의 결함으로 인해 발생하는 사용자의 손해에 대해 제조사가 손해배상을 하도록 하고 있다. 통제언어가 직접적으로 적용될 수 있는 부분은 제품의 표시부분이라고 할 수 있다.

2 몇몇 제품의 경우, 한국에서 생산되더라도 아예 원문을 영어로 작성하는 경우가 있다. 이러한 경우는 본 논문의 연구대상에서 제외된다.

3 Esselink (2000) S.211.

역문제 해결을 위해 본격적으로 개발이 시작된 통제언어는 통제영어로 출발해, 통제독일어, 통제스페인어, 통제덴마크어, 통제스웨덴어 등의 개발로 이어졌다.

한국에서도 2000년대 중반 이후로 주로 산업계와 학계를 중심으로 통제언어에 대한 관심이 높아지고 있다. 본 연구에서는 현재 국내에서도 그 필요성이 점차 부각되고 있는 통제한국어를 개발하기 위해, 번역수월성의 관점에서 어떤 언어적 특성을 고려해야 하는지를 독일어와 영어와의 비교를 통해 알아보게 될 것이다. 이를 위해 본 논문에서는 저자가 ETRI 한영 기술논문 자동번역시스템을 위해 개발한 통제한국어 모델과 Lehrndorfer(1996)의 통제독일어, 'Microsoft'사(이하 MS사)의 통제영어를 번역수월성의 관점에서 비교하게 될 것이다. 통제한국어, 통제독일어, 통제영어의 비교를 통해 어떠한 규칙이 번역수월성의 관점에서 언어 보편적이며, 어떤 규칙이 개별언어 의존적인지 알아보게 될 것이다. 또한 통제언어가 적용되는 기계번역시스템의 방법론에 따라 어떤 차이가 발생할 수 있는지에 대해서도 논의하게 될 것이다. 이 연구결과는 현재 독어학자들을 중심으로 논의가 활발하게 진행되고 있는 범용 통제한국어의 개발에 응용될 수 있을 것으로 기대한다.

## 2. 통제언어와 번역수월성

Lehrndorfer/Mangold(1997)는 통제언어가 기계번역시스템 뿐만 아니라 번역메모리 Translation Memory 시스템과 연계해서도 큰 효과를 보인다고 밝혔다. 번역메모리 시스템은 컴퓨터가 문장의 번역을 완전히 자동으로 수행하는 것이 아니라, 번역자가 번역하고자 하는 문장을 입력하면 이 문장과 가장 유사한 번역문쌍을 퍼지매칭 Fuzzy Matching 등의 방법을 사용하여 찾아낸 후 사용자에게 제시하는 기능을 갖고 있다. 이 경우, 통제언어에 기반하여 작성된 번역메모리를 사용할 경우, 문장 매칭율이 획기적으로 증가하여 결국 번역비용의 절감효과를 가져올 수 있다고 주장하였다.

4 Göpferich(1995), Ley(2005), Möller(2003), O'Brien(2005, 2006), O'Brien&Rotourier(2007) 참조.

통제언어를 기계번역에 적용할 경우 번역률 향상이 기대됨은 이미 Lehrndorfer가 통제독일어를 개발할 때 이미 밝힌 바 있다. 그녀는 지멘스사에서 개발한 ‘METAL’ 기계번역시스템을 사용하여 통제독일어가 기계번역시스템의 번역수월성 향상에 도움을 줄 수 있음을 밝혔다.

통제언어가 일반적으로 기계번역시스템의 번역수월성에 좋은 영향을 끼칠 것은 비교적 예측 가능한 일이다. 왜냐하면, 문법적으로 완벽하고 매끄러운 스타일로 작성된 문장이 기계적으로 처리하기가 수월하기 때문이다. 이는 대부분의 상용 기계번역시스템이 번역원문의 형태소, 구조 분석을 수행한 후 목표언어로 변환하는 방법론을 채택하고 있기 때문이다. 그러나 Aikawa(2007) 등의 연구에 따르면 이러한 방법론을 채택하고 있지 않은 통계기반 기계번역시스템의 경우에도 통제언어가 번역수월성에 긍정적인 영향을 미친다는 것이 보고되고 있다.

### 3. 통제언어

#### 3.1 통제영어

Ogden의 ‘Basic English(1930)’에서 출발한 읽기 쉽고 배우기 쉬운 영어에 대한 연구는 ‘AECMA Simplified English (이하 AECMA SE)’를 통해 산업계에 널리 퍼지기 시작하여 오늘날에 이르고 있다. 현재 논의되고 있는 현대적 개념의 통제언어로는 항공산업분야에서 사용되는 텍스트의 작성을 위해 고안된 ‘AECMA SE’가 최초라고 할 수 있다. 1995년도에 배포된 ‘AECMA SE Issue I’의 경우, 1,565개의 일반어휘가 수록된 사전과 57개의 문장작성 규칙으로 구성되어 있다. ‘AECMA SE’는 항공산업분야에서 작성되는 기술문서의 가독성과 번역수월성을 높이기 위해 고안되었으나, 번역수월성의 경우 애초부터 기계번역을 염두에 두고 개발된 것은 아니다.

이에 비해 ‘Microsoft사(이하 MS사)’ 통제영어의 경우, 애초부터 기계번역의 수월성을 위하여 고안되었다. MS사에서 출시되는 주요 제품들의 경우, 정식버전이 출시되기 이전 독일과 일본 등과 같은 주요시장을 대상으로 베타버전 Beta Version이 먼저 출시되어, 시장의 반응을 살피게 된다. 제품의 베타버전이 출시될

때, 제품에 수반되는 주요 문서들도 해당국의 언어로 번역이 되어 출시된다. 이때 기한의 촉박함, 번역비용의 절감 등을 위해 이러한 문서들은 전문번역가에 의해 번역되는 것이 아니라, MS사에서 직접 개발한 ‘MSR-MT System’을 사용하여 독일어와 일본어 등으로 번역한다.<sup>5</sup> MS 통제영어로 작성된 기술문서는 다른 언어로 기계번역될 때, 그렇지 않은 경우와 비교하여 월등한 성능향상의 효과를 보인다고 한다. 이에 대해서는 6장에서 좀 더 자세히 다루도록 한다.

MS 통제영어의 경우, ‘AECMA SE’와는 달리 사전에 대한 제약은 알려져 있지 않고, 21개의 쓰기규칙 Schreibregel로 구성되어있다. 쓰기규칙의 각 카테고리 및 이에 대한 설명은 다음 표에서 찾아볼 수 있다.

카테고리	설명
Formal Style	속어나 구어체 표현 사용금지
Spelling	스펠링 오류금지
Long Sentences	문장의 길이를 25단어 이하로 할 것
Short Ambiguous Sentences	6단어 미만의 짧은 문장을 금지함
Sentence Breaks	문장의 끝에 구두점을 부착할 것
Commas	컴마규칙을 따를 것
Hyphen	하이픈을 통한 새로운 복합어 사용의 금지
Abbreviations	등록되지 않은 약어사용의 금지
Parentheses	괄호를 이용한 코멘트의 금지
Capitalization	강조를 위한 대문자 사용의 금지
Relative Pronoun	관계대명사를 사용할 것
Attachment	외치된 관계절의 사용금지
Relative Clauses	관계대명사가 생략된 관계절 사용금지
-ed Verbs	-ed 형태 동사를 모호하게 사용하지 말 것
Ambiguous VP conjunct	모호한 병렬관계의 동사구 사용금지
Ambiguous VP conjunct2	명사구를 동사구의 병렬로 시작하지 말 것
Ambiguous NP/AP conjunct	명사구/형용사구의 모호한 사용금지
Ambiguous NP conjunct	명사구 병렬구조를 동사로 시작하지 말 것
Adjective/Verb Ambiguity	동사구 병렬구조를 형용사로 시작하지 말 것
-ing clauses	명백한 주어가 없는 -ing형 동사의 사용금지
-ing ambiguity	-ing형태 동사의 모호한 사용금지

표 1 : MS CL의 쓰기규칙

5 Aikawa et al. (2007) S.1

위 쓰기규칙들은 대부분 원문의 작성시 모호성을 최대한 배제하도록 하기 위한 규칙들이다. ‘MS CL’은 ‘AECMA SE’와는 달리 규칙들이 철저하게 기계번역 시스템이 다루기 어려운 현상들을 배제하기 위해 고안되었음을 알 수 있다.

### 3.2 통제독일어

통제영어의 성공으로 통제언어에 대한 산업적, 학문적 관심이 높아진 상황에서 Lehrndorfer(1996)는 처음으로 독일어에 통제언어의 개념을 도입하였다. ‘AECMA SE’가 문서의 가독성 측면을 주로 강조하여 개발된 반면, 통제독일어는 문서의 가독성 및 번역수월성을 고려하여 개발되었다. Lehrndorfer는 ‘AECMA SE’가 지나치게 어휘부를 제한하는 점을 비판하였는데, 제약이 심한 어휘부의 학습이 어렵다는 것이 그 비판의 주된 논지였다. 따라서 그녀는 통제독일어를 개발할 때 어휘부는 최대한 자유롭게 하였으며 통사부에서 쓰기규칙의 강화를 통해 독일어의 사용을 통제하고자 하였다.

Lehrndorfer는 자신이 개발한 통제독일어 규칙의 타당성 검증을 위하여 ‘METAL’이라는 기계번역시스템을 사용하였다. 통제독일어 규칙이 반드시 기계번역시스템에 긍정적인 영향만을 미친다고는 할 수 없지만, 대체적으로 기계번역시스템의 성능향상에 기여할 수 있음을 보였다. 또한 Lehrndorfer/Schachtl(1998)에서는 독일 지멘스사를 위해 개발한 ‘Controlled Siemens Documentary German’과 또 다른 기계번역시스템인 ‘TopTrans’시스템과의 관계를 언급하였다.

통제독일어가 통제영어와 비교하여 갖는 가장 큰 특징은 어휘부의 제약이 상대적으로 약하다는 점이다. Lehrndorfer가 통제영어의 가장 큰 문제점으로 지적한 것은 제약이 심한 어휘부의 학습수월성 Lernbarkeit이다. 통제언어를 사용하게 될 기술문서 작성자들이 사용 가능한 어휘와 사용 불가능한 어휘들을 모두 학습하고 자유롭게 문서를 작성하기는 매우 어려운 것이라는 비판이었다. 따라서 어휘부에서 적용할 수 있는 제약을 통사부로 전이시키는 방법을 채택하였다. 특정 어휘의 사용을 금지하기 보다는 이 어휘가 사용되는 특정 구문의 작성을 금지하는 것이 학습수월성의 측면에서 보다 더 효율적일 수 있다는 것이다. Lehrndorfer가 제안한 문장작성 규칙 중 가장 대표적인 규칙을 정리하면 다음과 같다.

- 기능동사구문 사용금지
- 화행에 알맞은 문장형식으로 작성
- 행위수동 금지
- 관계절 사용금지
- 병렬구조 작성규칙 준수
- 접속사 명시
- zu-분사형 금지
- 주어문두 배치

‘MS CL’과 비교해보면, 통제독일어 역시 문장구조의 모호성을 불러일으키는 구조의 사용을 금지하고 있음을 알 수 있다. 특히 독일어 문장에서 자주 사용되는 기능동사구문, 접속사가 생략된 문장, zu-분사형 등의 특정 구조를 금지하고 있음을 확인할 수 있다. 통제영어와 비교하여 특이한 점은 관계절 Relativsatz의 사용을 아예 금지하고 있다는 점인데, 이는 ‘한 문장 당 한 개의 명제’라는 기술 문서 작성원칙을 충실히 따르고 있다고 볼 수 있다.

## 4. 통제한국어

### 4.1 연구배경

통제언어를 한국어에도 적용하려는 시도는 기계번역분야에서 최초로 이루어졌다. 국내의 기계번역분야 연구는 한국전자통신연구원(ETRI) 등과 같은 정부출연 연구소를 중심으로 1990년대 후반부터 본격적으로 수행되어왔다. 특히 한국어를 출발언어 Quellsprache로 하는 경우에는 한국어 분석의 어려움으로 인해 프로토타입 Prototyp 시스템은 일찍이 개발되었으나 상용화에는 실패해왔다.<sup>6</sup>

한국어를 출발언어로 하는 기계번역시스템으로 최초로 상용화에 성공한 시스템은 한영 특허문서 자동번역시스템이라고 할 수 있다.<sup>7</sup> 한영 특허번역시스템은 기존의 일반분야 한영번역시스템을 특허문서를 대상으로 특화하여 현재 유럽, 미

6 Yang et al. (2001), 최승권 외 (2005) 참조.

7 Hong et al. (2005) 참조.

국, 일본 등의 특허심사관들에게 한국특허문서를 실시간으로 영어로 번역하여 제공하는 서비스를 위해 사용되고 있다. 한영 특허번역시스템은 이미 수년전부터 유럽특허청(EPO)의 특허심사관을 대상으로 자동번역서비스를 시행해오던 일본특허청(JPO)의 일영 번역시스템과 비교하여 그 성능이 거의 동등하거나 오히려 약간 앞서는 것으로 보고되고 있다.<sup>8</sup>

이를 기반으로 현재 'ETRI'에서는 이공계 전문직 종사자들을 위해 한영 기술 논문 자동번역시스템을 개발 중에 있다. 한국어 통제언어의 필요성은 바로 여기에서 출발한다. 단순히 중요정보만을 추출하기 위해 문장을 번역하는 경우와 영어로 논문을 작성하여 학회나 저널 등에 제출하는 경우는 번역의 목적이 완전히 다르다고 할 수 있다. 일반적으로 한영 자동번역시스템의 경우 번역률이 최고 75%~80%정도로 보고되고 있다. 이 정도의 번역률은 정보검색의 측면에서는 큰 불편 없이 사용될 수 있을 수준이나, 완성도 높은 영어문장을 만들어내야 하는 경우에는 매우 부족하다고 할 수 있다. 영어를 모국어로 하는 영문감수자에게 최종적으로 전달할 수 있을 정도의 수준이 되려면 적어도 번역률이 90%이상은 되어야한다. 그러나 현재 기계번역시스템의 수준에서 90%이상의 번역률을 단기간에 달성하는 것은 사실상 불가능하다고 할 수 있을 것이다.

이러한 문제를 해결하기 위한 하나의 방안으로 한국어 통제언어를 설계하는 것이 제안되었다. Kim et al. (2007)의 연구에 따르면 한국어 통제언어를 사용할 경우, 통제언어를 사용하지 않는 기존의 기계번역시스템보다 이론적으로 최대 13% 정도의 성능향상이 기대된다고 한다. 즉, 현재 최고수준의 한영 기계번역시스템에 한국어 통제언어를 적용했을 경우, 90% 이상의 번역률이 기대된다고 할 수 있다.

#### 4.2 문장작성 오류유형

통제한국어 개발에 대한 연구배경에서도 밝힌 바와 같이 본 논문에서 소개하는 통제한국어는 기계번역시스템의 번역수월성만을 고려하여 설계된 것이다.<sup>9</sup> 통

8 Hong et al. (2005) 참조.

9 번역수월성의 향상이 반드시 가독성의 향상을 가져온다고 볼 수는 없다. Möller(2003)의 연구에서는 통제덴마크어가 기계번역시스템의 번역수월성을 향상하는 효과는 있었으나,

제언어는 자연언어의 하위언어 *Subsprache*이다. 따라서 마치 인공언어 *künstliche Sprache*를 설계하는 것과 같이 공리 *Axiom*으로부터 출발하여 허용되는 모든 문장을 생성할 수 있는 생성기제와 사전, 의미해석규칙 등을 하향식 *Top-down*으로 정의할 수 있을 것이다.

그러나 한 언어의 문장들이 구체적으로 어떤 생성 매커니즘에 의해 만들어지는지 명확하게 알 수 없는 현실에서 이러한 생각은 매우 비현실적으로 보인다. 따라서 본 연구에서는 하향식 *Top-down*으로 통제한국어를 개발하지 않고, 상향 *Bottom-up* 방식으로 개발하고자 했다. 즉, 통제한국어의 모습이 어떠해야 한다면 정의하기 보다는, 어떤 문법구조는 피해야 한다는 식의 제약규칙 등을 설정함으로써 통제한국어의 모습을 간접적으로 정의하고자 했다.

이를 위해 본 연구에서는 한국어의 어떤 문법구조 및 어휘가 기계번역에서 문제가 될 수 있는지를 분석하였다. 기계번역은 일반적으로 입력문장의 분석 *Analyse*, 목표언어로의 변환 *Transfer*, 목표언어의 생성 *Generierung* 등의 단계로 수행된다.<sup>10</sup> 이 중 언어학적으로 가장 중요한 단계는 입력문장의 분석 단계라고 할 수 있다.

통제한국어를 설계하기 위해 우리가 주목한 부분은 기계번역시스템이 원문에 대해 완벽한 형태소/구조 분석을 수행하고 목표언어(영어)로 변환, 생성을 했음에도 불구하고 영어답지 못한 표현이 생성된 경우이다. 기계번역에서 일어나는 대부분의 오역은 분석이나 변환과정에서 오류가 생겼기 때문에 발생한다. 분석이나 변환과정에서 오류가 없었음에도 오역이나 어색한 번역이 생기는 이유는 한국어 원문에 문제가 있었기 때문이라고 말할 수 있을 것이다.

우리는 이러한 한국어 오류유형들을 파악하기 위해 기계번역된 4만5천개의 문장을 수작업을 통해 검토하였다. 4만5천개의 번역문중 분석과 변환은 주어진 문

독자(기술분야 전문가)들에게는 오히려 가독성의 측면에서 호응도 *Akzeptabilität*가 낮았음을 보여준다.

10 최근 기계번역의 연구동향은 문장에 대한 언어학적 분석을 거의 수행하지 않는 통계기반방식이 주류를 이루고 있으나, 상용화된 대부분의 시스템은 규칙기반 기계번역 혹은 패턴기반 기계번역 시스템이므로, 본 연구에서는 통계기반 기계번역시스템을 고려하지 않았다. 그러나 통제독일어의 예에서 본 바와 같이, 통제언어는 통계기반 기계번역시스템의 성능에도 긍정적인 영향을 미치는 것으로 알려져 있다. Aikawa et al. (2007), 홍문표(2008) 참조.



맥정보와 문장 내 정보를 최대한 이용해 성공적으로 수행되었으나 한국어 원문의 이상으로 어색한 번역이 나온 경우를 크게 다음과 같이 9개의 경우로 분류하였다.

- 주어-술어 불일치 문제
- 문두보조사 남용의 문제
- 모호한 표현의 사용 문제
- 구어체 표현의 사용 문제
- 동일표현/동일격 반복 문제
- 문장부호 미부착 문제
- 문법오류 문제
- 기능동사 남용의 문제

위의 8개 오류유형이 우리가 분류한 기계번역시스템의 번역수월성에 가장 큰 영향을 미치는 한국어 오류유형이다. 먼저 주어-술어 불일치 문제부터 살펴보기로 한다.

- 주어-술어 불일치 문제

예: 본 논문은 저대역 이동법의 단점 중 계산량을 크게 줄이고, 동일압축률에서 좋은 성능을 보이므로 보다 효율적인 동영상 부호화를 가능하게 한다

위 오류는 우리가 문장을 작성할 때 흔히 범하는 오류로서, 위 예문에서는 주어 '본 논문'이 3개의 용언, '줄이고', '보이므로', '가능하게하다'와 호응관계에 있다. 저자가 의도한 바는 '본 논문에서 제안하는 해결방안'이 3개의 용언과 의미적으로 호응하는 것이겠지만, 단어를 기반으로 번역을 시도하는 기계번역시스템의 경우에는 위와 같은 문제는 소위 '콩글리시 Konglish'를 생성하게 하는 원인이 된다.<sup>11</sup>

---

11 위 예문의 경우 this paper reduces... , shows..., and enables... 와 같은 오역이 발생하게 된다.

- 문두보조사 남용의 문제

예: 손상된 세그먼트의 표면은 그림 3에서 보는 것과 같이 큰 표면 삼각형을 관찰할 수 있다

한국어 문장의 문법적인 특징 중 하나는 ‘은’, ‘는’, ‘도’, ‘만’ 등과 같은 문두보조사를 매우 빈번하게 사용한다는 점이다. 이러한 문두보조사는 주제, 대조, 초점 등과 같은 담화적인 기능을 수행하지만, 전산처리에 있어서는 매우 큰 분석상의 문제를 초래한다. 일반적으로 구조분석단계에서 위와 같은 보조사가 등장하면 이러한 보조사의 문법상의 격 (주어, 목적어 등)을 복원하여 처리해야한다. 위 예문의 경우 ‘표면은’의 격은 의미상으로는 ‘표면에서는’과 같이 부사격이지만, 분석시스템은 위 명사구를 주격으로 처리하게 될 가능성이 높다. 따라서 위 예문을 “... the surface can observe ...”와 같은 비문을 생성하게 된다.

- 모호한 표현의 사용문제

예: 1588 기업체 음성다이얼 서비스는 1588에 등록된 회사의 명칭을 말로써 하면 그 회사의 1588 번호로 연결하여 주는 서비스이다

한국어 어휘 중에 그 사용역이 매우 넓어 독일어나 영어로 번역할 경우 여러 어휘로 번역될 수 있는 것들이 많다. 이에 속하는 대표적인 어휘가 ‘하다’, ‘으로’, ‘대하다’ 등이다. 이러한 표현들은 분석단계에서 모호성을 유발하여 결국 오역을 생성할 수 있는 가능성을 높게 한다. 위 예문의 경우에는 일반동사 ‘말하다’를 ‘말로써 하다’로 표현하였는데, ‘하다’ 동사는 상황에 따라 매우 다양하게 번역될 수 있는 동사이다. 번역의 합성성 Kompositionalität에 따라 번역을 수행하는 기계번역시스템의 경우 이러한 표현을 정확하게 번역하기는 매우 어렵다.

- 구어체 표현의 사용문제

예: 그리고, 그림 4의 결과와 survey 논문에서의 복원 결과와의 비교에서도 정확한 3차원 정보를 얻었음을 눈으로 확인할 수 있었다

기술논문 등과 같은 문어체 텍스트에서 구어체표현은 일반적으로 많이 사용되

지 않을 것으로 생각되나, 실제로는 이러한 구어체 표현이 매우 많이 등장함을 알 수 있다. 이러한 구어체 표현은 기계번역시스템의 성능 및 분야별 커스터마이제이션 Customization 정도에 따라 번역성능에 큰 영향을 안 미칠 수도 있으나, 현실적으로 수많은 구어체 표현을 처리할 수 있는 기계번역시스템은 존재하지 않는다. 따라서 이러한 표현들은 번역수월성의 측면에서 사용이 금지되어야 할 필요가 있다.

#### - 동일표현/동일격 반복 문제

예: 연구개발 단계에서 생성된 문서의 번호 부여는 ‘시스템 명 + L2 + 일련번호’와 같은 형태로 부여한다

한 문장 내에서 동일한 단어 혹은 구가 반복되는 경우가 흔히 있다. 한국어의 주제지향성 topikorientiert에 기인하는 것으로 보이는 이러한 특징은 특히 번역의 결과를 매우 어색하게 만든다. 주로 술어성명사 prädikatives Nomen와 ‘-하다’동사가 이러한 비문을 만드는 경우가 많은 것으로 분석되었다.

#### - 문장부호 미부착 문제

예: 현대 사회는 조직의 규모가 커지고 처리하는 업무가 복잡해지고 있다

독일어나 영어와는 달리 한국어 문장 작성시 문장부호의 사용법에 대한 교육이 비교적 소홀하게 다루어져왔던 것이 사실이다. 이 결과, 많은 작성자들은 문장을 작성할 때, 반드시 쉼표 등을 사용해야 할 부분에서 이를 사용하지 않는 경우가 빈번하다. 위 예문은 하나의 문장이지만 두 개의 명제가 전달되고 있다. ‘현대사회는 조직의 규모가 커진다’와 ‘처리하는 업무가 복잡해지고 있다’가 그것이다. 이렇게 두 개의 전혀 다른 명제를 하나의 문장으로 표현할 경우에는 쉼표의 사용이 필수적이다. 따라서 위 문장에서는 ‘커지고’와 ‘처리하는’ 사이에 쉼표가 사용되어야 한다. 이러한 문장부호의 미사용 및 부적절한 사용은 문장의 가독성을 떨어뜨릴 뿐만 아니라 기계번역시에도 분석의 오류를 유발하게 된다. 실제로, 위 문장은 ‘커지고’와 ‘처리하는’이 동시에 ‘업무’를 수식하는 구조로 분석되어 오역이 생성되었다.

### - 문법오류문제

예: 데이터 *압축이*란 원래의 데이터 보다 작은 부호로 *표현한다*

일반적으로 통제언어의 역할이 문장의 문법성/비문법성을 판별하는 것은 아니다. 문장의 문법성/비문법성은 통제언어의 적용 앞 단계에서 철자 및 문법오류 검사 등을 통해 이미 판별되어야 할 문제이기 때문이다. 그럼에도 불구하고 많은 경우에 작성자들은 자신이 작성한 문장이 비문법적인 요소를 가지고 있는 사실을 인지하지 못하는 경우가 있다. 위 예문의 경우가 바로 그러하다. '-이'란은 한국어의 인용격조사이다. 따라서 문법적으로 '-라고 한다' 등과 호응하게 된다. 그럼에도 불구하고 위 예문에서는 일반동사 '표현하다'와 결합되었다.

### - 기능동사 남용의 문제

예: 이중 유기 EL은 LCD, PDP, FED등과 비교해서 구동전압의 측면에서 약 10V 정도의 낮은 전압에서도 *구동이 가능하다*

독일어 텍스트에서와 마찬가지로 한국어 텍스트에서도 많은 기능동사구문 Funktionsverbgefüge이 사용된다. 한국어의 기능동사구문은 독일어나 영어의 그것과 완전히 일치하지는 않지만 대개 술어적 의미를 지닌 명사와 의미정보를 많이 잃은 기능동사가 결합한다는 점에서는 유사하다고 볼 수 있다. 또한 기능동사구문을 하나의 일반동사로 바꿔 쓸 수 있다는 점에서 유사하다. 위 예문을 보면 '구동이 가능하다'는 '구동된다'로 바꿔 쓸 수 있다. 위와 같이 기능동사구문을 사용하게 되면 기계번역시스템에서는 단어레벨에서 번역을 시도하므로 'drive is possible'과 같은 어색한 번역이 생성된다.

이상 본 바와 같이 번역수월성의 관점에서 본 통제한국어는 위의 주요 오류를 배제한 한국어의 하위언어라고 볼 수 있을 것이다.<sup>12</sup>

12 통제언어를 설계할 때에는 일반적으로 사용가능한 어휘부, 문장을 만들기 위한 통사부의 규칙 등을 고안하게 된다. 그러나 본 연구에서 개발한 통제한국어는 기술논문을 작성하는 불특정다수를 위하여 고안된 것이므로, 최대한 제약을 없애기로 하였다. 문장작성자는 통제언어를 염두에 두지 않고 문장을 자연스럽게 작성하게 되고, 기계번역시스템으로

### 4.3 통제한국어체커

통제언어가 성공적으로 사용되기 위해서는 통제언어체커 *Kontrollierte Sprache Checker*라고 부르는 도구가 반드시 지원되어야 한다. 통제언어가 제약하는 어휘 부 및 통사부 규칙들은 대부분 자신만의 문장작성 방식에 익숙한 기술문서 작성자들에게는 처음에는 매우 불편할 수 있기 때문이다.<sup>13</sup> 최근 통제언어관련 연구 동향은 문서작성자들에게 최대한 문서작성의 자유를 허락하고, 기계번역 등을 적용하기에 앞서, 작성된 문서가 통제언어 작성규칙에 위배되는 것이 없는지 점검하는 워크플로우를 어떻게 가장 효율적으로 설계할 수 있는가에 관한 것이다. 이와 관련하여 Mitamura(1999), Reuther(2003), Nübel(2004), Kim et al. (2007) 등의 연구에서는 각각 영어, 독일어, 한국어에 대해 통제언어체커의 적용가능성에 대해 언급하고 있다.

통제한국어체커를 개발하기 위해서는 한국어처리모듈이 반드시 필요하다. 특히, 형태소분석, 구문분석 모듈이 뒷받침되어야 실질적인 한국어체커를 개발할 수 있다. 한영 논문번역시스템의 경우, 통제언어에 대한 특별한 교육을 받은 저자, 즉, 일반인이 텍스트를 작성한 후, 영어로 기계번역을 수행하기에 앞서, 통제한국어규칙에 위배되는 것이 없는지 체커를 통해 검토하게 된다. 앞 장에서 제안한 통제한국어 규칙을 적용하기 위해서는 형태소분석, 구조분석을 수행한 후, 위 규칙에 위배되는 것이 없는지를 통제언어규칙 적용모듈을 구동하게 된다. 통제언어규칙 적용모듈은 메타규칙으로 적용할 수 있는 부분과 데이터 기반 체크 부분으로 나뉜다. 앞서 제시한 통제한국어 규칙 중 메타규칙으로 구현할 수 있는 규칙은 ‘동일격, 동일표현 반복금지 규칙’, ‘문장부호규칙’ 등이다. 그 외의 규칙들은 모두 메타규칙으로 표현하기가 매우 어려운 내용들이다. 따라서 이러한 규칙들의 성공적인 적용을 위해서는 규칙에 위배되는 데이터들의 대량 수집을 통한 데이터기반 방법론을 사용해야 한다. 그림1은 이러한 방법론을 적용한 통제한국

문장을 번역하기 직전에 문장이 통제언어 규칙에 위배되는 것이 없는지 통제언어체커를 통하여 검토하게 된다. 이에 대해서는 통제언어체커 부분에서 자세히 논의하도록 한다.

13 Lehmdorfer(1996)는 이러한 이유 때문에 통제독일어를 사용자의 학습이 어려운 어휘부보다는 통사부의 제약을 강화하는 방향으로 설계하였다. 그러나 통제언어체커가 잘 지원된다면 이러한 학습성에 관련된 문제는 쉽게 해결할 수 있다.

어체커 적용시나리오이다.

그림에서 보는 바와 같이 통제언어체커가 수행할 수 있는 기능은 개별 언어처리 모듈 (형태소분석, 구조분석)과 연결되어 있다. 또한 사용자의 편의를 위해 통제언어규칙을 엄격히 적용할 것인지, 아니면 필수적인 주요 규칙들만 적용할 것인지 사용자가 선택하게 함으로써, 통제언어에 대한 사용자의 거부감을 최소화하도록 하였다.

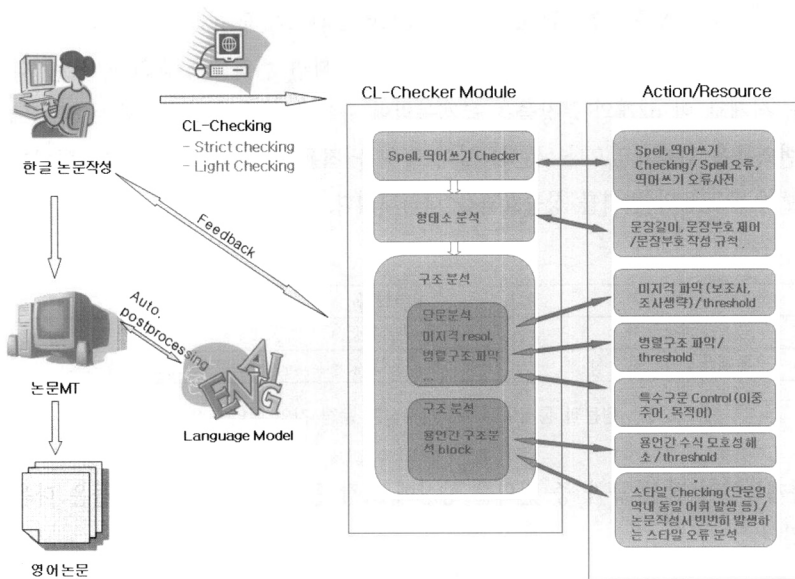


그림 1 : 통제한국어체커 모듈

## 5. 비교분석

### 5.1 통제독일어 실험

본 장에서는 통제언어를 사용한 경우와 사용하지 않은 경우, 번역수월성에 어떤 차이가 생기는지를 독일어, 영어, 한국어의 경우에서 살펴보고자 한다. 이를 위하여 한-영 기계번역, 독-영 기계번역, 영-아랍/중/불/네덜란드어 기계번역의 결

과를 비교 평가하였다. 한-영, 독-영 기계번역결과는 본 연구에서 직접 실험을 수행한 결과이고, 영-아랍/중/불/네덜란드어 기계번역시스템 번역률은 Aikawa et al. (2007) 연구의 결과이다.

먼저 통제독일어의 경우, Lehrndorfer(1996)가 제시한 통제독일어 작성규칙에 따라 기술문서에서 무작위로 추출한 264개의 문장을 통제독일어 규칙을 통해 재작성하였다. 이렇게 하여 실제로 264개의 문장 중 통제독일어 작성규칙에 의해 재작성된 문장은 총 62개의 문장이었다. 총 264문장을 기준으로 통제독일어를 적용한 경우 번역률은 73.17%로부터 75.58%로 2.41%가 향상되었다.<sup>14</sup> 실제로 통제독일어 규칙이 적용된 62개의 문장만을 고려한다면 번역률 상승의 폭은 좀 더 크다. 실제로 이 62개의 문장들은 통제독일어 규칙을 적용하기 이전에는 64.91%의 번역률을 나타내었으나, 통제독일어 규칙을 적용한 후에는 74.19%의 번역률을 보여, 9.28%의 번역률 상승효과를 나타내었다.

	번역률	비고
원문	73.17%	
통제독일어 재작성	75.58%	+2.41% 향상

표 2 : 원문과 통제독일어로 재작성된 문장 간의 번역률 비교

통제독일어 작성 규칙 중 번역률 향상에 가장 큰 영향을 미친 규칙들은 다음과 같다.

- 관계절의 사용대신 문장을 2문장으로 분리
- 25단어 이상의 장문을 단문으로 분리<sup>15</sup>
- 병렬구조 Koordination 규칙
- 주어를 문두로 한 일반적인 어순준수
- 접속사 생략의 금지

14 이 실험은 규칙기반 기계번역시스템인 'Systran'시스템에서 이루어졌다. 통계기반 기계번역시스템에서 이루어진 실험에 대해서는 홍문표(2008)에서 좀 더 상세하게 다루고 있다

15 이 규칙은 Lehrndorfer(1996)에서 제안된 규칙은 아니나, 통상적으로 장문의 경우 문장의 분석성공 확률을 낮게 하기 때문에, 통제독일어 적용규칙시 일괄적으로 25단어를 기준으로 이보다 긴 문장은 두 문장 이상으로 재작성하였다.

통제독일어의 통사규칙 중 번역률 향상에 가장 큰 영향을 미친 규칙들의 공통점은 문장의 구조를 단순화하는 규칙이라는 점이다. ‘관계절’ 및 ‘문장길이’와 관련된 규칙은 직접적으로 문장의 구조를 단순화하는 역할을 하고 있다. 또한 ‘병렬규칙’, ‘어순규칙’, ‘접속사’규칙 등은 모호성을 방지하는 규칙이라고 볼 수 있다.

## 5.2 통제영어 실험

다음은 통제영어가 영어를 출발언어로 하는 기계번역시스템에서 번역률 향상에 미친 영향에 대해 알아보기로 한다. Aikawa et al. (2007)에 따르면 영어를 출발언어로 하는 기계번역시스템의 경우 통제영어를 통해 아랍어의 경우 2.8%, 중국어의 경우 3.4%, 프랑스어의 경우 5.6%, 네덜란드어의 경우 2.45%가 향상되어, 평균 약 3.56% 정도의 번역률 향상이 보고되었다.<sup>16</sup> 통제독일어를 통해 영어로 번역했을 때 약 2.41%의 번역률 향상이 이루어진 것과 같이, 유사한 언어쌍인 영어=>네덜란드어의 경우 2.45%의 성능향상이 보고된 점은 흥미롭다 할 수 있겠다.

	일반영어	통제영어	비고
영어=>아랍어	64.75%	67.55%	+ 2.8%
영어=>중국어	63.63%	67.03%	+ 3.4%
영어=>프랑스어	59.63%	65.23%	+ 5.6%
영어=>네덜란드어	58.88%	61.33%	+ 2.45%

표 3 : 통제영어가 번역률 향상에 미치는 영향에 대한 실험결과

각 언어쌍별로 번역률의 향상에 가장 큰 영향을 미친 규칙들은 다음 표에서 나타내는 바와 같다.<sup>17</sup>

16 Aikawa et al.(2007)의 실험에서 사용된 기계번역시스템은 독=>영 기계번역에서 사용한 것과 같은 규칙기반 기계번역시스템이 아니라 통계기반 기계번역시스템이다. 그러나 홍문표(2008)에서 보인 것과 같이, 통계기반 기계번역시스템에서도 규칙기반 기계번역시스템의 경우와 유사하게 통제언어를 통해 비슷한 정도의 번역률 향상이 이루어진다. 따라서 본 실험에서 사용한 기계번역시스템의 방법론이 다르더라도 통제언어가 번역률 향상에 미치는 영향에 대한 비교평가는 큰 무리가 없다고 볼 수 있다.

17 O'Brien(2006)은 이러한 규칙들을 ‘Negative Translatability Indicators’라고 부른다.



순위	영어=>아랍어	영어=>중국어	영어=>프랑스어	영어=>네덜란드어
1	Formal Style	Formal Style	Short Ambiguous Sentences	Formal Style
2	Hyphens	Attachment	Formal Style	Capitalization
3	Short Ambiguous Sentences	-ing clauses	Spelling	Spelling
4	Capitalization	Spelling	Adjective/Verb Ambiguity	Short Ambiguous Sentences
5	Spelling	Long Sentences	Capitalization	Long Sentences

표 4 : 언어쌍별로 번역률 향상에 도움이 되는 규칙순위

영어를 출발어로 할 때, 거의 모든 경우에서 ‘Formal Style’ 규칙이 번역률 향상에 가장 큰 영향을 미치는 것으로 드러났다.<sup>18</sup> 이것은 영어 기술문서 상에 의외로 구어체 표현이 많이 사용된다는 것을 보여준다. 이러한 표현들은 번역시스템의 분석을 어렵게 한다기 보다는 사전이나 데이터베이스에 누락되어 있기 때문에 오역을 초래하는 것이다. 또한 ‘Hyphen’이나 ‘Capitalization’ 관련 규칙들도 주로 분석상의 모호성 등을 해결하기 위한 규칙이라기보다는, 하이픈이나 잘못된 대문자화로 인해 해당 어휘나 표현을 사전이나 데이터베이스에서 찾을 수 없기 때문에 발생하는 문제를 해결하기 위한 규칙들이다.

그 외 ‘Short Ambiguous Sentences’, ‘-ing clauses’, ‘Attachment’, ‘Long Sentences’ 규칙 등은 분석상의 모호성을 해소하기 위한 규칙들이다. 전체적으로는 모호성 문제를 해결하기 위한 규칙들이 많지만, 그 영향력에서는 데이터베이스나 사전에 미등록된 표현들의 방지를 위한 규칙들이 더 크다는 점이 통제영어의 경우 다른 언어들과 비교하여 흥미롭다 할 수 있다.

이 실험 결과는 우리에게 중요한 점을 시사하고 있다고 볼 수 있다. 즉, 통제 언어가 기계번역시스템의 전반적인 번역수월성 향상을 가져온다고 말할 수는 있지만, 기계번역시스템이 어떠한 방법론을 채택하느냐에 따라 번역수월성에 영향을 미치는 규칙이 바뀔 수 있다는 점이다. 통제독일어의 경우 실험에 사용했던 ‘Systran’ 시스템은 전통적인 규칙기반 기계번역시스템이었기 때문에 문장의 구조분석을 용이하게 하는 규칙들이 번역률 향상에 큰 효과가 있음을 알 수 있었

18 ‘Formal Style’ 규칙은 기술문서를 작성할 때 구어체 표현을 사용하는 문제와 관련된 규칙이다.

다. 그러나 통제영어 실험에서 사용된 ‘MSR-MT’ 시스템은 통계기반의 방법론을 채택하고 있으므로, 그 특성상 입력문의 구조분석을 수행하지 않는다. 따라서 구조분석을 용이하게 하는 규칙들보다는 오히려 문장 스타일을 일반적인 스타일로 제한하는 규칙이 가장 효과가 있음을 알 수 있었다.

### 5.3 통제한국어 실험

통제한국어가 한영 기계번역시스템의 번역률 향상에 미치는 영향을 파악하기 위한 실험을 수행하였다. 이 실험을 위해 한국어 기술분야 문서에서 무작위로 100문장을 추출하였고, 이 문장들을 앞 장에서 제시한 통제한국어 규칙에 따라 재작성하였다. 문장의 재작성은 통제한국어 규칙에 기반한 통제한국어 체커가 자동으로 제안하는 부분만 재작성하는 방식으로 수행하였다. 전체 100문장 중 앞서 소개한 통제한국어 작성 규칙에 따라 재작성한 문장은 총 41문장이었다. 통제한국어 체커의 성능을 향상시키고 데이터베이스 규모를 늘린다면 통제한국어의 적용범위는 향후 더 커지게 될 것이다. 이 실험결과 번역률은 70.5%에서 74.75%로 4.25%가 향상되었다.

	번역률	비고
원문	70.5%	
통제한국어 재작성	74.75%	+4.25% 향상

표 5 : 원문과 통제한국어로 재작성된 문장 간의 번역률 비교

통제한국어 규칙을 적용한 결과를 좀 더 구체적으로 살펴보면, 규칙이 적용된 총 41문장 중 25문장의 경우는 번역결과가 향상되었으며, 13문장의 경우는 번역결과와의 차이가 없었으며, 3문장의 경우는 오히려 번역결과가 나빠진 것으로 분석되었다. 통제한국어를 적용했음에도 번역결과가 개악된 이유를 살펴보면, 원문에 등장하는 한국어 용언구가 번역데이터베이스에 존재하여 번역이 잘 되었으나, 통제언어로 재작성한 표현은 오히려 번역데이터베이스에 존재하지 않는 경우였다. 그러나 이러한 경우는 보편적인 현상이라기보다는 데이터베이스의 일부 오류로 보는 것이 합당할 것이다.

통제한국어의 각 규칙별로 번역률 변화에 영향을 미친 정도는 다음 표에 나타났다. 표6에서 굵은 글씨체로 표기된 규칙들이 번역수월성에 큰 영향을 미친 규칙들이다. 한영 기계번역시스템은 규칙기반 방법론과 유사한 패턴기반 방법론에 기반하여 작동한다. 따라서 이 시스템에서도 입력문의 구조분석 결과가 번역 성능에 큰 영향을 미친다. 이 규칙들을 좀 더 자세히 살펴보면 독일어에서는 볼 수 없었던 ‘주어-술어 불일치’ 관련 규칙, ‘문두보조사 남용’ 관련 규칙, ‘동일표현, 동일격 반복’ 관련 규칙 등이 한국어의 언어적 특성상 번역수월성에 큰 영향을 미치는 것을 알 수 있었다. 그 외에 ‘모호한 표현’ 관련 규칙, ‘문장 부호 미부착’ 관련 규칙, ‘구어체표현’ 관련 규칙 등도 번역수월성에 비교적 큰 영향을 미치는 것으로 분석되었다.

오류명칭	번역률변화
주어-술어 불일치	+21%
문두보조사 남용	+12.5%
모호한 표현	+14.5%
구어체 표현	+7.5%
중의적 표현	+25%
동일표현, 동일격 반복	+20%
문장부호 미부착	+12.5%
문법오류	+4.2%
기능동사표현	+4.2%

표 6 : 규칙별 번역수월성 변화수치

## 7. 결론

본 연구에서는 통제언어가 기계번역시스템의 번역수월성에 미치는 영향을 통제독일어, 통제영어, 통제한국어를 통해 살펴보았다. 또한 통제한국어를 개발할 때 고려해야 할 점에 대해 논의해보았다.

우리는 본 연구에서 진행된 3개 언어에 대한 실험을 통해 통제언어가 일반적으로 기계번역시스템의 번역수월성을 향상시킨다는 기존의 주장을 확인할 수 있었다. 그러나 번역수월성을 향상시키는 규칙은 언어적 특성에 기인할 뿐만 아니

라, 어떤 기계번역시스템을 사용하느냐에 따라서도 달라질 수 있음을 확인하였다. 즉, 한국어는 독일어나 영어와는 전혀 다른 어족에 속하므로, 그 언어적 특성에 기인하여 번역수월성에 가장 큰 영향을 미치는 규칙들이 통제독일어나 통제영어에서는 전혀 고려되지 않는 규칙들이었다 (주어-술어 불일치 제약규칙, 문두보조사 남용계약 규칙 등).

또한 거의 유사한 언어적 특성을 지니는 영어와 독일어의 경우, 어떤 기계번역방법론을 채택한 시스템을 사용하느냐에 따라 문장의 구조분석을 수월하게 하는 규칙들이 더 큰 효과가 있는지, 아니면 문장의 스타일을 단순화하는 규칙들이 더 큰 효과가 있는지 알 수 있었다. 본 연구의 실험결과, 통제독일어의 경우 그 적용대상이 규칙기반 기계번역시스템이었으므로 문장구조의 모호성을 해소하는 규칙들이 번역수월성을 높이는 주요한 규칙들이었다. 통제영어의 경우 그 적용대상이 통계기반 기계번역시스템이었으므로 구어체 표현 등을 일반화하는 규칙들이 번역수월성의 향상에 가장 큰 영향을 미친다는 것을 알게 되었다.

우리는 이 연구를 통해 통제언어를 개발할 때 해당언어의 언어학적 특성을 충분히 고려해야할 뿐만 아니라, 어떤 기계번역시스템에서 통제언어가 사용될 것인지도 고려해야할 필요가 있음을 알게 되었다. 어쩌면 이러한 결론은 통제언어의 유용성에 대한 반대 주장이 될 수도 있을 것이다. 그러나 규칙기반 기계번역시스템에 유용한 통제언어 규칙과 통계기반 기계번역시스템에 유용한 규칙들이 서로 완전히 상충되는 것이 아니라, 대부분의 규칙들은 양 시스템에 공통적으로 유용하게 적용되므로, 통제언어 규칙의 적용 우선순위를 기계번역시스템의 유형에 따라 적절히 배치하는 작업이 필요할 것이다. 이러한 규칙적용의 우선순위를 매개 변수화함으로써 범용통제언어 모델을 개발할 수 있을 것으로 기대한다.

## 참고문헌

- 최승권, 홍문표, 박상규 (2005): 다국어 자동번역 기술, 전자통신동향분석, 제20권 제4호, 16-27
- 홍문표 (2008): 통제독일어가 번역수월성 향상에 미치는 영향에 대한 연구, 독일언어문

학 제 40집

- AECMA (1995): AECMA Simplified English, A Guideline for the Preparation of Aircraft Maintenance Documentation in the International Aerospace Maintenance Language, Issue I
- Aikawa et al. (2007): "Impact of Controlled Language on Translation Quality and Post-editing in a Statistical Machine Translation System", Proceedings of MT-Summit XI, 1-8
- Esselink, B. (2000): A Practical Guide to Localization, John Benjamins Publishing Company, Amsterdam/Philadelphia
- Göpferich, S. (1995): Textsorten in Naturwissenschaft und Technik. Pragmatische Typologie - Kontrastierung - Translation. Tübingen: Narr.
- Kim, Y., Hong, M. & S. Park (2007): CL-Guided Korean-English MT System for Scientific Papers, Lecture Notes in Computer Science, vol.4394, 409-419, Springer Verlag
- Lehrndorfer, A. (1996): Kontrolliertes Deutsch: linguistische und sprachpsychologische Leitlinien für eine (maschinell) kontrollierte Sprache in der technischen Dokumentation. Tübingen: Narr.
- Lehrndorfer, A. / R. Mangold (1997): "How to Save Money in Translation Cost", TC-Forum 97-2, URL: <http://www.techwriter.de/tc-forum/pdf/editions/tcf972s.pdf>
- Lehrndorfer, A. / S. Schachtl (1998): Controlled Siemens Documentary German and TopTrans, TC-Forum 98-3, URL: <http://www.tc-forum.org/topictr/tr9contr.htm>
- Ley, M. (2005): Kontrollierte Textstrukturen. Ein (linguistisches) Informationsmodell für die Technische Kommunikation. Dissertation, Justus-Liebig-Universität Gießen.
- Mitamura, T. (1999): "Controlled Language for Multilingual Machine Translation". Proceedings of MT-Summit 1999
- Möller, M. (2003): Grammatical Metaphor, Controlled Language and Machine Translation, Proceedings of EAMT/CLAW 2003
- Nübel, R. (2004): "Evaluation and Adaptation of a Specialised Language Checking Tool for Nonspecialised Machine Translation and Non-expert MT Users for Multi-lingual Telecooperation", Proceedings of LREC 2004
- O'Brien, S (2005): Methodologies for Measuring the Correlations between Post-Editing Effort and Machine Translatability, Machine Translation, vol.19, no.1, 37-58
- O'Brien, S (2006): Pauses as Indicators of Cognitive Effort in Post-editing Machine Translation Output, Across Languages and Cultures vol.7, no.1, 1-21
- O'Brien, S. & J. Roturier (2007): How Portable are Controlled Language Rules? A

- Comparison of Two Empirical MT Studies, Proceedings of MT Summit XI, 345-352
- Ogden, C. K (1930): Basic English: A General Introduction with Rules and Grammar. London. Paul Treber & Co., Ltd.
- Reuther, U. (2003), "Two in one - can it work? Readability and translatability by means of controlled language". Proceedings of the 4th International Workshop on Controlled Language Applications, Dublin, Ireland
- Yang, S., Kim, Y., Seo, Y., Choi, S. & S. Park (2001): Korean to English TV caption translator: CaptionEye/KE, Proc. 6th Natural Language Processing Pacific Rim Symposium (NLPRS), 639-645

## Zusammenfassung

### **Komparative Untersuchung über Kontrolliertes Deutsch, Kontrolliertes Englisch und Kontrolliertes Koreanisch im Hinblick auf die Übersetzbarkeit**

Hong, Munpyo (Sungkyunkwan Uni)

Generell wird angenommen, dass Kontrollierte Sprache (=KS) der Übersetzbarkeit von Texten für maschinelles Übersetzen(=MÜ) durch restringiertes Lexikon und vereinfachtes Syntax dient. In dieser Arbeit versuchen wir herauszufinden, welche Schreibregeln die Übersetzbarkeit jeweils für Deutsch, Englisch und Koreanisch am meisten verbessern und welche linguistischen Charakteristiken dahinter stecken. Darüber hinaus wollen wir untersuchen, ob der MÜ-Ansatz, auf den sich ein maschinelles Übersetzungssystem stützt, die Schreibregeln der KS beeinflussen.

Um das herauszufinden, wurden zwei Experimente durchgeführt. In dem ersten Experiment wurde ein deutscher Text anhand von dem kontrollierten Deutsch, das von Lehdorfer(1996) eingeführt wurde, neu erfasst. Anschließend wurden der originelltext und der mit dem neu erfasste Text jeweils maschinell übersetzt Das MÜ-System, das für

dieses Experiment bereitgestellt wurde, war das Systran System, das sich auf den traditionellen regelbasierten Ansatz stützt. Danach wurden zwei Übersetzungen von Übersetzungsexperten evaluiert. In diesem Experiment wurde gezeigt, dass die wichtigsten Regeln für das Deutsche solche Regeln sind, die die strukturelle Ambiguität des Satzes vermindern.

Das zweite Experiment für das Koreanische verlief auch in ähnlicher Weise. Um einen koreanischen Text neu zu erfassen, wurden die Schreibregeln, die für das ETRI MÜ-System entworfen sind, herangezogen. Unser Experiment zeigte, dass die wichtigsten Regeln auch wie bei der deutschen Sprache, solche Regeln sind, die die Ambiguität des Satzes beseitigen. Aber wegen der linguistischen Charakteristiken der koreanischen Sprache, spielen in der koreanischen Sprache solche KS-Regeln wie ‘Subjekt-Prädikat Kongruenz Regel’ oder ‘Verbot des Topikmarkers am Satzanfang’ eine sehr wichtige Rolle. Dies spiegelt die linguistischen Verschiedenheiten zwischen Koreanisch und Deutsch wider.

Der Vergleich von unserem Experiment mit dem von Aikawa et al. (2007) zeigt, dass bei der KS-Entwurf die Arten von MÜ-Systemen mit berücksichtigt werden sollen. Bei dem statistikbasierten Ansatz scheinen eher die Stilregeln als die strukturbezogenen Regeln wichtigere Rolle zu spielen.

주제어: 통제언어, 통제독일어, 기계번역, 번역수월성

Schlüsselbegriffe: kontrollierte Sprache, kontrolliertes Deutsch, maschinelles Übersetzen, Übersetzbarkeit

필자 이메일 주소: skkhmp@skku.edu

투고일: 2008.07.05 / 심사일: 2008.07.30 / 심사완료일: 2008.08.19