

## 통제독일어가 번역수월성 향상에 미치는 영향에 대한 연구

홍 문 표(성균관대)

### I. 머리말

최근 전자제품을 구입한 경험이 있는 사람이라면 누구나 한번쯤은 제품의 실제적 사용법에 어려움을 겪었을 것이다. 그럼에도 불구하고 대부분의 사람들은 사용설명서를 읽기 싫어한다. 그 이유는 무엇인가? 여러 가지 원인을 찾을 수 있겠지만, 특히 사용설명서 등에 나타나는 복잡한 문장구조 때문이 아닐까 생각해본다. 사용설명서 등에 나타나는 문장이 복잡하고 이해하기 어려운 경우, 사용자들은 설명서 읽기를 아예 기피하거나, 읽더라도 내용을 잘못 이해할 가능성이 높다. 또한 이런 문장을 다른 외국어로 번역할 경우에는 오역의 가능성이 매우 높아지게 된다.

삼성, LG, 현대 등과 같이 전세계를 대상으로 사업을 벌이는 소위 글로벌 플레이어 Global Player의 경우, 제품을 생산할 때 제품에 수반되는 사용설명서 및 수리안내서 등과 같은 기술문서도 적게는 영어, 독일어, 중국어 등과 같은 세계 주요 3~4개 언어에서부터 많게는 10개 이상의 언어로 번역되어 현지소비자들에게 공급되어야 한다. 제품의 기획, 생산, 판매 공정을 단축하기 위해 최첨단 경영기법이 도입되고 있는 지금, 문서의 생산 및 번역, 현지화 Lokalisation 작업은 아직도 종래의 방법을 답습하고 있는 실정이다. 따라서 이러한 프로세스를 최적화하여 시간과 비용을 단축하기 위한 각 기업들의 노력이 치열하게 전개되고 있다. 문서 및 제품의 효율적인 현지화 작업을 위해서 많은 기업들은 주로 번역메모리 Translation Memory 시스템 등을 사용하여, 기존 번역물을 재사용하는 방식으로 번역공정을 단축하고 있다.<sup>1)</sup> 번역메모리 방식은 기존의 번역물이 방대하게 존재하는 경우에는 상당히 높은 품질의 번역결과를 비교적 단시간 내에 얻을 수 있다는 장점이 있으나, 기존 번역물, 즉, 번역메모리가 존재하지 않을 경우에는 기존의 수동번역과 크게 다를 바 없다. 따라서 이러한 경우에는 완전 자동번역인 기계번역 Maschinelle Übersetzung (이하 MÜ)에 의존할 수밖에 없게 된다. 그럼에도 불구하고 많은 사람들은 기존의 수많은 기계번역시스템의 성능에 대해 회의적인 생각을 갖고 있다. 그 이유는 기계번역시스템의 실망스러운 성능 때문일 것이다. 그러나 미국과 유럽의 경우, 많은 기업들에서는 기계번역시스템을

---

1) 박주형 외 (2001) 참조

업무에 투입하여 번역공정을 단축하고 있다. 실제로 마이크로소프트사의 경우 영어로 작성된 문서를 독일어와 일본어로 번역하기 위해 기계번역시스템을 투입하고, 전문번역가가 컴퓨터로 번역된 결과를 후처리작업을 통해 완전하게 만드는 공정을 채택하고 있다.<sup>2)</sup> 이를 통해 5~25%의 시간과 비용을 절감하고 있다고 한다.<sup>3)</sup> Aikawa et al. (2007)는 이와 관련하여 기계번역 시스템의 성능을 최대한 이용하기 위해서는 원문을 문법과 스타일에 어긋남 없이 통제언어의 규칙에 맞게 작성해야 한다고 주장한다.

본 논문에서는 문서의 가독성 Lesbarkeit과 번역수월성 Übersetzbarkeit을 높이기 위해 고안된 통제독일어 Kontrolliertes Deutsch가 기계번역의 측면에서 실제로 어느 정도의 성능향상을 가져올 수 있는지를 실험을 통해 알아보고자 한다. 그 성능향상 정도가 통제언어 Kontrollierte Sprache 옹호자들이 주장하는 바와 같이 의미 있는 정도인지를 알아보게 될 것이다. 또한 Lehrndorfer(1996)가 제안한 통제독일어가 종래의 규칙기반 기계번역방식 Regelbasierte MÜ에서 뿐만 아니라, 현재 기계번역연구 패러다임의 주류를 이루고 있는 통계기반 기계번역방식 Statistikbasierte MÜ에도 번역수월성 향상에 도움이 될 수 있는지 알아보고자 한다.

Lehrndorfer가 통제독일어를 제안한지도 어느덧 10여년이 흐른 지금, 10년 전과 비교하여 자연언어처리 기술이 눈부시게 발전한 이 시점에서 통제독일어 규칙들은 어떤 의미를 가지는지 알아보기 위해 본 논문은 다음과 같이 구성된다. 먼저 II.1장에서는 통제언어의 개념에 대해 소개한다. 통제언어의 간단한 역사와 산업계에 널리 알려진 대표적인 통제언어인 'AECMA Simplified English'를 소개한다. 그리고 II.2장에서는 본 논문의 연구대상인 통제독일어에 대해 소개한다. II.3장에서는 통제독일어 규칙을 통한 기술문서 재작성에 대해 논의한다. III장 실험에서는 통제독일어로 재작성된 문장과 원문 그대로인 문장을 각각 규칙기반 기계번역시스템과 통계기반 기계번역시스템을 사용하여 번역한다. 그리고 번역된 결과에 대해 영어-독일어 번역 전문가에 의해 평가 Evaluierung를 실시한 결과에 대해 논의한다. IV장에서는 실험 결과의 의미에 대해 논의하고자 한다. 끝으로 V장은 향후 연구방향에 대해 기술한다.

## II. 본론

---

2) 마이크로소프트사의 경우, 오피스 등과 같은 제품을 전세계에 출시하기 전 시장의 반응을 살피기 위해 독일과 일본시장만을 대상으로 베타버전 Betaversion을 출시한다고 한다. 이 경우 소프트웨어와 이에 수반되는 문서의 현지화를 위해 영-독, 영-일 기계번역시스템을 사용하고 있는 것으로 알려져 있다.

3) Aikawa et al. (2007) 참조

## II.1 통제언어 Kontrollierte Sprache

통제언어란 자연언어에 내재되어 있는 복잡성 Komplexität과 중의성 Ambiguität 등을 없애거나 최소화하여 문장의 이해와 번역을 용이하게 하기 위해 자연언어를 인공적으로 제약한 일종의 하위언어 Subsprache이다.

통제언어의 개념과 같이 자연언어를 인공적으로 제약하여 의사소통의 문제를 해결하고자 했던 최초의 시도는 Ogden(1930)의 'Basic English'이다. 이 시도는 그러나 현대적 개념의 통제언어라기 보다는 850개의 핵심단어만을 사용하여 영어표현을 만들도록 한 단순화된 영어의 하위집합이었다.

보다 현대적 개념의 통제언어로 볼 수 있는 것은 'CFE Caterpillar Fundamental English'이다. 중장비 제조업체인 미국의 'Caterpillar'사에서 가독성과 번역수월성이 높은 기술문서를 작성하기 위해 'Basic English'에 기반하여 통제언어의 일종인 'CFE'를 고안하게 되었다. 'CFE'는 'Basic English'의 기본어휘에 중장비 분야에서 사용되는 필수 전문용어를 추가한 형태를 띠고 있다.

오늘날 통제언어를 언급할 때 가장 많이 인용되는 것은 'AECMA Simplified English'이다. 'AECMA Simplified English'는 항공분야에서 발생하는 기술문서의 작성을 위해 고안된 통제언어로서, 1970년대 후반 유럽의 'European Association of Aerospace Industries(AECMA)'와 미국의 'Aerospace Industries Association(AIA)'에서 민간항공기 제작분야 기술문서들의 가독성에 대한 연구를 시작하면서 개발되었다. 'AECMA Simplified English'는 1986년도에 첫 번째 공식버전이 출시된 이래 몇 번의 수정을 거쳐 현재는 'ASD Simplified English'라는 이름으로 유지되고 있다. (URL: <http://www.asd-stel00.org/>)

'AECMA Simplified English'는 통제된 어휘부 Lexikon와 몇 개의 쓰기 규칙 Schreibregel으로 이루어져있다.<sup>4)</sup> 어휘부는 1,565개의 일반어휘와 전문용어가 포함되어 있다. 쓰기 규칙은 57개로 이루어져 있다. 어휘부의 구성원칙은 최대한 모호성을 배제하는 것이다. 즉, 어떤 단어가 여러 개의 품사를 할당받을 수 있는 경우, 단 하나만의 품사를 허용한다. 또한 의미적으로도 어떤 단어가 여러 개의 의미를 가지고 있을 때, 하나만의 의미를 허용한다. 다음의 예를 보면서 설명하도록 한다.

4) 본 논문에서는 1995년도에 배포된 'AECMA Simplified English Issue I'을 기준으로 소개한다.

Keyword(POS)	Assigned Meaning/Use	Approved Examples	Not Acceptable
ACCURATE(adj)	"Exact"	THE ADJUSTMENT MUST BE ACCURATE	
act(v)	NOTE: Do not use this verb. Use a "specific action" verb	THE HYDRAULIC FLUID OPEN THE VALVE	The hydraulic fluid acts on the inlet valve.
active(adj)	ON	WHEN THE SYSTEM IS ON, YOU WILL GET A REPORT	When the system is active, you will get a report.

표 1: 'AECMA Simplified English'의 어휘부

위의 표는 'AECMA Simplified English' 어휘부의 일부이다. 위 표에서 'Keyword' 칼럼은 사전에 등재되는 표제어를 나타내고 있다. 이 중 표제어가 대문자로 표기되는 경우(예: ACCURATE)만이 'AECMA Simplified English'에서 허용되는 단어이다. 그 외에 소문자로 표기되는 경우(예: act, active)는 해당 단어가 허용되지 않음을 나타낸다. 이 경우 해당 단어 대신에 사용해야 할 단어들이 'Assigned Meaning/Use' 칼럼에 제시되어 있다. 어휘부에서는 이와 같은 방식으로 1,565개의 일반 어휘에 대해 허용되는 품사와 의미를 명시적으로 제시하고 있다.

쓰기 규칙의 경우는 총 57개의 규칙이 9개의 섹션으로 나뉘어져 있다. 이는 단어 사용 규칙부터 시작하여, 구구조 작성규칙, 문장 작성규칙 등에 관한 규칙들이다. 구구조 작성규칙에 관한 예를 살펴보면 다음과 같다. 다음의 규칙은 영어의 명사구 작성에 관한 규칙으로서, 하나의 명사구를 세 개 이상의 명사를 연속하여 작성해서는 안 됨을 제약하고 있다.

RULE: 2.1 Do not make noun clusters of more than three nouns.

Example:

WRITE: **The bolt that attaches the uplock to the nose landing gear is ....**

NOT: **The nose landing gear uplock attachment bolt is ....**

'AECMA Simplified English'는 이상에서 본 바와 같이, 어휘부를 상세하게 제약하

고, 주로 통사부에 해당하는 쓰기 규칙은 비교적 느슨하게 제약하는 특징을 보이고 있다. 이러한 특징은 추후 Lehrndorfer에 의해 통제언어의 학습수월성 Lernbarkeit 측면에서 비판을 받는 원인이 된다. 통제언어를 사용하게 될 사용자, 즉 주로 테크니컬라이터 Technical Writer들이 어휘부를 모두 염두에 두고 사용가능한 단어와 사용불가능한 단어를 구별해서 문장을 작성하기를 기대하는 것은 무리라는 것이다. 이 대신 Lehrndorfer는 어휘부는 비교적 느슨하게 제약하고, 학습이 어휘부에 비해 상대적으로 쉬운 통사부의 쓰기 규칙을 강하게 제약함으로써 어휘부의 제약을 통해 얻는 효과를 간접적으로 얻을 수 있다고 주장한다. II.3절에서 Lehrndorfer가 주장한 통제독일어에 대해 좀 더 자세하게 알아보기로 한다.

## II.2 통제독일어 Kontrolliertes Deutsch

Lehrndorfer는 ‘AECMA Simplified English’의 가장 큰 문제점은 기술문서 작성자들이 이것을 학습하기가 어렵다는 점에 있다고 보았다. ‘AECMA Simplified English’는 앞 절에서 본 바와 같이 간단한 57 개의 쓰기 규칙만이 존재하고 대부분의 언어 통제는 어휘부 Lexikon에서 이루어진다. 허용된 단어와 금지된 단어의 목록을 구비한 후, 이 지침에 맞게 작성자가 기술문서를 작성해야 하는 것이다. Lehrndorfer는 950개에 달하는 어휘가 각각 사용이 가능한 어휘인지 아닌지 사용자가 학습하기가 어려울 것이라는 점을 문제로 지적하였다.

이러한 문제를 고려하여 그녀는 통제독일어를 고안할 때 어휘부는 최대한 자유롭게 허용하면서 통사부를 비교적 엄격하게 제안하였다. 그녀의 주장에 따르면 특정한 통사규칙 혹은 구조를 제한하면, 그러한 구조에 필수적으로 동반되는 특정한 어휘들의 사용을 막는 간접적인 효과가 있을 것이다.<sup>5)</sup> 예를 들어 기능동사구문 Funktionsverbgefüge에서 자주 사용되는 ‘nehmen’ 동사를 어휘부에서 직접적으로 사용이 가능한지 아닌지를 제약할 것이 아니라, 아예 통사부에서 기능동사구문의 사용을 금지한다면, ‘nehmen’이 우선적으로 갖는 ‘취하다’ 의미를 제외한 그 다른 어떤 용법이나 의미로 사용되는 것을 막을 수 있다는 것이다.

이러한 접근방법을 취할 때의 장점은 통제언어를 사용자들이 좀 더 쉽게 배울 수 있다는 점이다. 왜냐하면 어휘부에서 사용이 가능한 단어의 목록을 학습하는 것보다 통사부의 문장작성 규칙을 학습하는 것이 사용자 입장에서 좀 더 용이하기 때문이다. 그러나 자연언어처리 Natürliche Sprachverarbeitung의 기술수준이 그녀가 통제독

5) Lehrndorfer(1996) S.139참조

일어를 처음 제안했을 당시(1996년)보다 월등히 높아진 지금, 그녀의 이와 같은 주장은 큰 설득력을 갖지 못한다. 왜냐하면 최근에는 통제언어 규칙을 자동으로 적용하여 사용자에게 통제언어 규칙에 위배되는 문장을 자동으로 알려주고 수정할 수 있도록 도와주는 통제언어체커 *Kontrollierte Sprache Checker*가 많이 개발되어 널리 사용되고 있기 때문이다. 즉, 더 이상 사용자는 허용되는 어휘가 무엇인지 특별히 학습할 필요가 없이 자신이 원하는대로 문장을 작성한 후, 통제언어체커를 통해 잘못 작성된 부분을 알게 되면 체커가 제안하는대로 문장을 재작성하면 되기 때문이다.<sup>6)</sup>

## II.4 통제독일어를 통한 문장 재작성

통제독일어가 번역수월성에 미치는 영향에 대한 실험에 앞서 일반 독일어로 작성된 기술문서를 *Lehrndorfer*가 제시한 통제독일어 작성규칙에 따라 재작성하였다. 본 실험에서 사용된 문장은 디지털 TV, 디지털 카메라, 컴퓨터 모니터의 사용설명서로부터 무작위로 추출되었다. 실험을 위해 추출된 문장을 읽어본 후, *Lehrndorfer*의 규칙을 적용할 수 있는 경우에는 적용하여 재작성하고, 어떠한 규칙도 적용되지 않는 경우는 그대로 남겨두었다. 이를 통해 일반 기술문서에 대한 *Lehrndorfer*의 통제독일어 규칙의 적용비율도 함께 조사하고자 했다. 이에 대한 결과는 다음 절에서 다시 언급하도록 한다. *Lehrndorfer*의 규칙 중 실험대상 기술문서에 대표적으로 적용된 규칙들은 다음과 같다.

### a. 기능동사 사용금지 (일반 동사로 표현)

원문: Ein kaltes Gerät darf in einem warmen Raum erst *in Betrieb genommen werden*, wenn ein eventuell vorhandener Feuchtigkeitsbeschlag auf der Bildschirmfläche verdunstet ist.

통제독일어: Ein kaltes Gerät darf in einem warmen Raum erst dann *benutzt werden*, wenn ein eventuell vorhandener Feuchtigkeitsbeschlag auf der Bildschirmfläche verdunstet ist.

*Lehrndorfer*는 기능동사구문의 사용을 금지하는데, 이와 같은 통사부의 제약을 통해 분석과 번역시 모호성을 일으키는 특정 어휘의 사용을 간접적으로 금지할 수 있다고 주장한다.<sup>7)</sup> 기능동사구문은 그 밖에도 기계번역시 문장분석의 부담을 가중하게

6) 통제언어체커에 관해서는 *Mitamura&Nyberg(1995)*, *Mitamura(1999)*, *Nübel(2004)*, *Reuther(2003)* 참조

7) *Lehrndorfer(1996)*, S.138참조

된다. ‘in Betrieb nehmen’이 하나의 의미단위로 함께 분석되어야 하기 때문이다.<sup>8)</sup>

b. 화행에 알맞은 문장형식으로 변경

원문: Ihr Gerät ist für den Betrieb in trockenen Räumen bestimmt.  
 통제독일어: Ihr Gerät in trockenen Räumen benutzen!

사용자 설명서와 같은 기술문서에는 제품의 올바른 사용을 위해 사용자가 취해야 할 행동지침 등에 대한 지시 등이 주로 표현된다. 이와 같은 지시 화행 Direktiva은 명령형 문장으로 표현하여야 정보가 왜곡되지 않고 전달된다.<sup>9)</sup> 위 예문의 경우는 어떤 기기를 건조한 공간에서만 사용하라는 내용이다. 원문에서는 이러한 정보를 ‘für et. bestimmt sein’으로 간접적으로 전달한다. 이러한 경우에는 직접적인 명령형으로 화행에 맞게 문장을 재작성하였다.<sup>10)</sup>

c. 행위수동 금지

원문: In das Batteriefach der Fernbedienung dürfen keine Akkus eingelegt werden.  
 통제독일어: Keine Akkus in das Batteriefach der Fernbedienung einlegen

기술문서에서 수동태를 사용해도 좋으나 아니냐에 관한 문제는 아직도 많이 논의되고 있는 문제이다. ‘AECMA Simplified English’의 경우, 기술문서에서 지시화행을 나타내는 경우에는 수동태의 사용을 무조건 금지하고 있다.<sup>11)</sup> Lehrndorfer(1996)는 제시화행의 경우에는 상태수동 Zustandspassiv를 허용하지만, 행위를 나타내는 지시화행의 경우에는 수동태를 허용하지 않는다. 위 예문의 예에서는 수동형 문장을 명령형을 사용하여 능동문장으로 표현하였다.

d. 관계절 Relativsatz 금지

8) 독일어 기능동사구문의 전산언어학적 처리에 대해서는 Krenn&Erbach(1994)참조  
 9) 한국어 통제언어의 관점에서 본 화행과 문장구조의 관계에 대해서는 최명원&최지영(2008) 참조  
 10) 지시화행의 올바른 전달을 위해 반드시 명령형구조를 사용해야 하느냐에 관한 것은 논쟁의 여지가 있다. 기업의 입장에서는 제품설명서 등을 읽는 사용자가 지시 혹은 명령을 받고 있다는 느낌이 들지 않게 하기 위해서, 의도적으로 명령형을 피하고 간접적으로 지시화행을 표현할 수도 있기 때문이다. 실제로 마이크로소프트사 소프트웨어의 매뉴얼 등을 살펴보면, 명령형보다는 간접적인 표현이 더 많음을 알 수 있다. 그러나 본 논문에서는 이 논의에 대해서는 더 이상 깊게 다루지 않고자 한다.  
 11) AECMA(1995), S.1-3-3 참조

원문: Entsprechende AV-Ebene einstellen, dazu die Taste „AV“ drücken und mit den Zifferntasten das AV --Gerät auswählen, welches das Bild liefern soll.

통제독일어: Entsprechende AV-Ebene einstellen. Dazu die Taste „AV“ drücken und mit den Zifferntasten das AV --Gerät auswählen. Es soll das Bild liefern.

대체적으로 문장이 길면 길수록 이해도가 떨어지고 전산적으로도 처리하기 어려운 문제가 있다.<sup>12)</sup> 문장이 길어지는 요인은 여러 가지가 있지만 특히 관계절의 사용이 주요 원인 중의 하나이다. Lehrndorfer(1996)는 관계절을 금지할 경우 문장 길이가 짧아지는 효과 뿐만 아니라 한 문장에 하나의 테마만을 전달하는 효과를 얻을 수 있다고 주장한다. 위의 예문에서도 원문의 ‘welches das Bild liefern soll’ 부분을 통제독일어 부분에서는 대명사 ‘es’를 사용하여 두 개의 문장으로 분리하였다.

e. 병렬구조 Koordination

원문: Mit der Taste der Fernbedienung in den Stand-by Zustand schalten oder, bei längerer Nichtnutzung, den Netzschalter am TV-Gerät drücken und das TV -Gerät vom Netz trennen.

통제독일어: Mit der Taste der Fernbedienung in den Stand-by Zustand schalten. Bei längerer Nichtnutzung, den Netzschalter am TV-Gerät drücken und das TV --Gerät vom Netz trennen.

병렬구조는 통사구조의 분석단계에서 많은 모호성을 일으킨다. 따라서 Lehrndorfer는 병렬구조를 만들 때에는 반드시 같은 품사끼리만 연결을 하도록 제한한다. 위 예문에서 등위접속사 ‘oder’는 전치사구로 이끌어지는 문장을 연결하고 있으나, 병렬구조의 앞 성분 내에 또 다른 전치사구 ‘in den Stand-by Zustand’가 등장하므로, 분석시 모호성이 발생할 수 있다. 따라서 이 경우에는 아예 문장을 새로 작성하여 병렬구조의 모호성도 없애고 문장의 길이도 줄일 수 있다.

f. 접속사 명시

원문: Ist kein Tastenbefehl der Fernbedienung erfolgt, so schaltet das TV-Gerät in den Stand-by Zustand.

통제독일어: Wenn kein Tastenbefehl der Fernbedienung erfolgt ist, so schaltet das TV-Gerät in den Stand-by Zustand.

12) 독일어 문장의 길이가 번역수월성에 어떤 영향을 미치는가에 관한 연구는 거의 이루어진 바 없다. 영-한 기계번역에서 긴 문장을 분할하여 처리하는 방법론에 대해서는 Roh et al (2003)참조



독일어 문장의 종속절 Hypotaxen을 작성할 때에는, 종속절의 의미에 따라 접속사를 명시적으로 사용하도록 권고된다. 예를 들어 인과관계를 나타낼 때에는 반드시 'durch', 'da', 'weil', 'so dass' 만을 사용할 수 있다. 위 예문은 원문에 조건접속사 'wenn'이 생략된 형태이다. 이러한 경우에는 반드시 'wenn'을 사용하여야 하며, 주절에서는 'dann'을 사용해도 되고 안해도 된다. 접속사 'wenn'이 생략되면, 기계번역시스템의 구조분석기는 이 문장을 동사가 문두에 위치한 의문문으로 잘못 분석할 가능성이 있다.

g. zu-분사형 금지

원문: Schalten Sie das zu decodierende Programm ein (z.B. PREMIERE).

통제독일어: Um ein Program (z.B. PREMIERE) zu decodieren, schalten Sie das Program ein.

통제독일어에서 수식어 Attribute는 통제독일어 사전에 정의되어 있는 단어만 사용이 허용된다. 특히 zu-분사형의 사용은 금지되고 있다. 따라서 이러한 경우에 사용자는 의미가 동일하도록 다른 문장을 사용하여 재작성해야 한다.

h. 주어문두

원문: Beim Öffnen des Gehäuses können spannungsführende Teile freigelegt werden.

통제독일어: Spannungsführende Teile können beim Öffnen des Gehäuses freigelegt werden.

Lehrndorfer는 독일어에서 S-V-O 어순이 가장 보편적인 어순이라고 하며 주어를 항상 문두에 위치시킬 것을 권한다. 그러나 이 규칙은 기술문서를 작성하는 테크니컬라이터들에 의해 쉽게 받아들여지기 어려운 규칙이다. 왜냐하면 상황에 따라서 문장의 주어가 아닌 다른 성분을 테마화하여 문두에 위치시킬 필요가 있기 때문이다. 그럼에도 불구하고 다음절에서 보게 될 것과 같이 주어문두 규칙은 규칙기반, 통계기반 기계번역시스템에 공통으로 번역수월성의 향상을 가져오게 된다.

### III. 실험

본 논문에서는 Lehrndorfer(1996)가 제안한 통제독일어 규칙에 기반하여 독일 기술 문서에서 무작위로 추출한 문장을 재작성한 후, 두 개의 독-영 기계번역시스템에 통

제독일어로 작성되지 않은 원문과 통제독일어로 재작성된 수정문을 입력한 후, 영어 번역문의 결과를 평가하였다.<sup>13)</sup> 본 실험에서 두 개의 독-영 기계번역시스템을 사용한 이유는 현재 기계번역관련 연구의 방향이 기존의 규칙기반 방법론 Regelbasierter Ansatz로부터 통계기반 방법론 Statistikbasierter Ansatz로 급격히 변하고 있기 때문이다.<sup>14)</sup> 기존의 상용기계번역시스템은 대부분 규칙기반 방법론을 적용하여 개발되었으나, 최근에는 언어학적 지식에 의존하기 보다는 두 언어 병렬말뭉치 bilinguale Korpora에서 통계적인 방법을 통해 계산해낸 번역지식을 이용한 통계기반 방법론이 연구의 대세를 이루고 있다.<sup>15)</sup> 규칙기반 방법론은 원문의 올바른 분석에 더 큰 의존을 하고 있으므로 규칙기반 방법론을 채택한 기계번역시스템에서 통제언어가 좀 더 큰 영향을 발휘할 것이라는 것이 일반적인 예측이다. 본 논문에서는 이러한 특성을 고려하여 통제독일어가 각 기계번역시스템에 얼마만큼의 영향을 미치고 그 차이가 어느 정도인지 알아보려고 하였다. 이 실험에서 사용한 규칙기반 기계번역시스템은 미국 'Systran'사에서 개발한 온라인 독-영 기계번역시스템이고,<sup>16)</sup> 통계기반 기계번역시스템은 'Google'사의 온라인 독-영 기계번역시스템이다.<sup>17)</sup>

영어로 번역된 문장은 3명의 평가자에 의해 평가되었다.<sup>18)</sup> 각 문장은 다음과 같은 기준에 따라 0점에서 최고 4점까지 점수가 부여되었다.

점수	비고
4점	원문이 정보의 손실없이 완벽하게 번역됨
3점	번역상 약간의 어색함은 있으나 정보가 거의 완벽하게 전달됨

13) 본 논문에서는 독일어 기술문서 원문으로 이루어진 평가문장을 테스트셀 1이라 칭하고, 통제독일어로 재작성된 문장을 포함한 평가문장집합을 테스트셀 2라 부르기로 한다

14) 통계기반 기계번역에 대한 점증하는 관심 등 때문에 2007년도에는 기계번역 분야의 최대 학회인 'MT Summit'에서 통계기반 기계번역에 관한 튜토리얼 Tutorial이 따로 개최되기도 하였다

15) 그러나 통계기반 방법론 연구의 가장 최근 경향에 따르면, 형태소 혹은 구문레벨의 언어학적 지식이 기반이 된 통계정보를 사용하는 기계번역 시스템이 가장 좋은 성능을 내는 것으로 보고되고 있다. 이와 관련된 최신 연구는 Hewavitharana et al. (2007) 참조

16) URL: <http://www.systran.com>

17) URL: [http://www.google.co.kr/language\\_tools?hl=ko](http://www.google.co.kr/language_tools?hl=ko)

18) 본 실험에 참여한 3명의 평가자는 각각 모두 5년 이상의 영어와 독일어 기술문서 번역경력을 가지고 있는 전문번역인력이다

2점	의미가 구 Phrase단위로 부분적으로 전달됨
1점	의미가 단어레벨에서만 전달됨
0점	번역실패

표 2 : 번역률 평가 기준표

각 평가자는 위의 표에 나온 지침에 따라 독일어에서 영어로 번역된 문장에 0점부터 4점까지의 점수를 부여하였다. 만약 번역결과가 표상에 제시된 특정 점수를 부여하기에 애매한 경우, 예를 들어, 4점과 3점의 중간수준이라고 판단되는 경우는 0.5점 단위로 점수로 부여하도록 하였다. 이 경우 평가자는 3.5점을 부여하게 된다. 각 문장별로 3명의 평가결과에 대해 그 중 중간점수를 채택하였다.<sup>19)</sup> 번역률은 각 문장별 점수를 합산한 후 만점으로 나누고 백분율로 환산한 값이다. 번역률을 구하는 공식은 다음과 같다.

$$\text{번역률(\%)} = \frac{\text{총점}}{(\text{문장수} * 4\text{점})} * 100$$

실험을 위해 3개의 기술문서 (디지털 텔레비전, 프린터, 모니터)로부터 264개의 문장을 무작위로 추출하였다. 이 중 Lehrndorfer의 통제독일어 작성규칙이 적용되어 채 작성된 문장의 수는 62개였다. 이는 전체 문장 중 23.48%의 적용률에 해당된다. 이는 독일어에서 발간되는 독일어로 작성된 기술문서들이 이미 언어학적 측면에서 상당히 높은 수준으로 작성되고 있음을 간접적으로 보여준다.

실험문장수	264
통제독일어 규칙이 적용된 문장	62
규칙적용률	23.48%

표 3 : 일반 기술문서에 대한 통제독일어규칙 적용률

규칙기반 방법론을 채택하고 있는 ‘Systran’ 독-영 기계번역시스템은 테스트셀 1에 대해 73.17%의 번역률을 보였다. 테스트셀 2에 대해서는 75.58%의 번역률을 나타내었다. 원문과 비교하여 통제독일어로 작성된 문장을 포함한 경우, 2.41%의 번역률 향상효과가 관찰되었다. 그러나 Lehrndorfer의 통제독일어 작성규칙이 적용된 62개의 문장만을 보면, 62개의 원문만(테스트셀 1-1)을 그대로 기계번역한 경우 64.91%의 번

19) 이를 위해 ‘MS Excel’에서 제공하는 ‘MEDIAN’함수를 사용하였다

역률이 조사되었으나, 62개의 문장에 대해 통제독일어로 재작성한 경우(테스트셀 2-1)만을 보면, 74.19%로 번역률이 9% 이상 상승하게 되었다.

	Systran시스템	비고
테스트셀1 (원문)	73.17%	
테스트셀2 (통제독일어)	75.58%	+2.41% 향상

표 4 : 테스트셀 1, 2에 대한 'Systran'시스템의 번역률

	Systran시스템	비고
테스트셀 1-1 (62문장)	64.91%	
테스트셀 2-1 (62문장)	74.19%	+9.28% 향상

표 5 : 테스트셀 1-1, 2-1에 대한 'Systran'시스템의 번역률

통계기반 방법론에 기반한 'Google' 독-영 기계번역시스템은 테스트셀 1에 대해 66.44%를 나타내었다. 테스트셀 2에 대해서는 67.60%의 번역률이 측정되었다.<sup>20)</sup> 원문을 그대로 번역한 결과와 비교하여 통제독일어로 작성된 문장을 포함하고 있는 테스트셀 2를 번역한 결과 약 1.16%의 성능향상이 관찰되었다.

	Google시스템	비고
테스트셀1 (원문)	66.44%	
테스트셀2 (통제독일어)	67.60%	+1.16% 향상

표 6 : 테스트셀 1, 2에 대한 'Google'시스템의 번역률

	Google시스템	비고
테스트셀 1-1 (62문장)	61.69%	
테스트셀 2-1 (62문장)	66.53%	+4.84% 향상

표 7 : 테스트셀 1-1, 2-1에 대한 'Google'시스템의 번역률

'Google' 시스템에 대해서도 통제독일어 작성규칙이 적용된 62개의 문장(테스트셀 1-1, 2-1)만을 살펴보았다. 통제독일어가 적용되기 이전의 상태인 테스트셀 1-1의 경

20) 'Systran' 독-영 시스템과 비교하여 아직 'Google' 독-영 시스템의 성능이 현저히 낮은 것으로 보인다. 그러나 통계기반 기계번역시스템은 방대한 양의 이중언어말뭉치만 있다면 언어모델 Language Model과 번역모델 Translation Model을 비교적 쉽고 빠르게 구축할 수 있으므로, 향후 급속한 번역성능의 향상이 기대된다.

우 'Google' 시스템은 61.69%의 번역률을 나타내었다. 그러나 통제독일어 작성규칙이 적용된 후에는 66.53%로, 4.84%가 향상되었다.

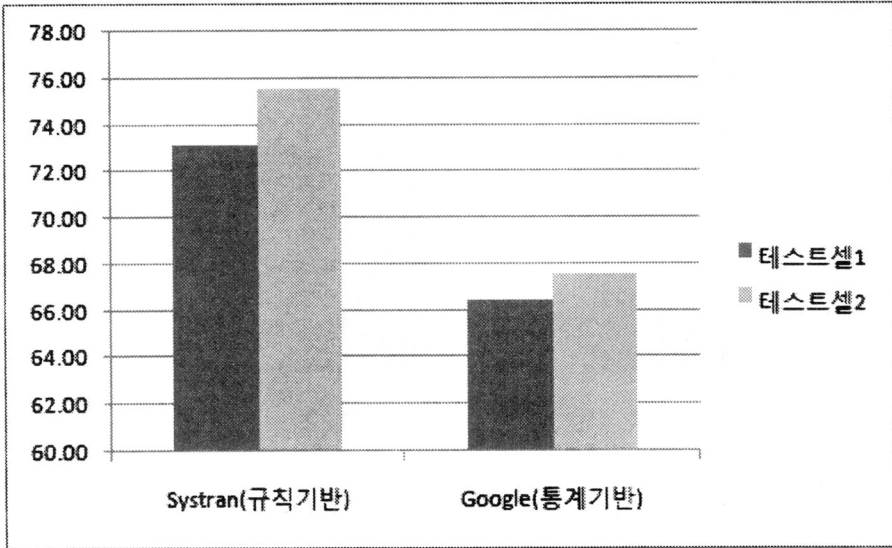


그림 1 : 테스트셀 1,2에 대한 'Systran'시스템과 'Google'시스템 결과비교

이 실험을 통하여 우선 통제독일어는 통계기반 기계번역시스템(4.84% 향상)보다는 규칙기반 기계번역시스템(9.28% 향상)에 더 큰 효과를 가져옴을 알 수 있었다.<sup>21)</sup> 그 원인은 여러 가지로 분석할 수 있겠지만, 원문의 정확한 분석에 많이 의존하게 되는 규칙기반 기계번역시스템의 경우, 원문이 얼마나 문법적으로 올바르고 스타일상으로도 문제가 없이 작성이 되어있는지에 큰 영향을 받기 때문일 것으로 예상된다. 이에 반해 통계기반 기계번역시스템의 경우에는, 원문에 대한 구문단계의 분석을 수행하지 않고, 코퍼스에서 학습된 결과에만 의존하므로, 원문의 통제독일어 작성규칙 위배여부에 크게 좌우되지 않는 것으로 보인다.<sup>22)</sup>

- 21) 본 실험 결과, 독-영 기계번역의 경우 아직은 규칙기반 기계번역시스템이 통계기반 기계번역시스템보다는 좀 더 좋은 번역성능을 보임을 알게 되었다. 그럼에도 불구하고 이러한 번역성능의 차이를 일반화할 수는 없을 것이다. 왜냐하면 통계기반 시스템은 학습데이터, 즉, 학습코퍼스의 도메인이 무엇이냐에 따라 번역성능이 확연히 달라질 수 있기 때문이다.
- 22) 물론 원문이 문법적으로, 혹은 스타일상 잘못된 문장이라면, 학습데이터에도 이와 유사한 문장이 발견될 가능성이 상대적으로 낮으므로 통계기반 기계번역시스템에서도 평균보다 낮은 번역성능을 보일 가능성이 높다. 이러한 예측은 Aikawa et al.(2007)의 연구에서도

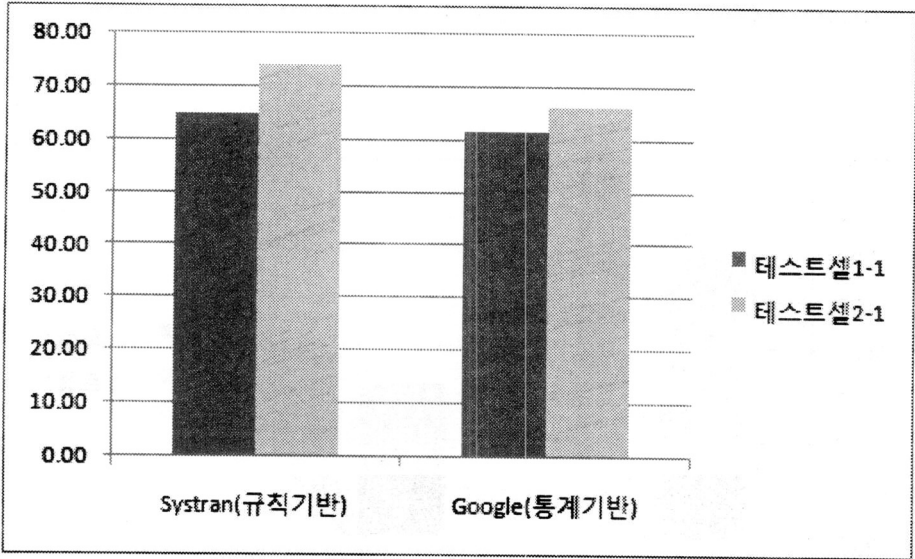


그림 2 : 테스트셀 1-1, 2-1에 대한 'Systran'시스템과 'Google'시스템 결과비교

#### IV. 결과분석

다음은 적용된 통제독일어의 규칙 중, 어떤 규칙이 번역성능의 향상에 도움을 주었는지, 또는 오히려 장애가 되었는지를 살펴보았다. 먼저 'Systran' 시스템의 경우, 통제언어 작성규칙이 적용된 62문장중, 30문장은 통제독일어로 재작성한 후 번역결과가 향상되었고, 2문장은 오히려 번역결과가 개악되었다. 나머지 30문장의 경우는 번역결과, 즉, 점수의 변화가 없었다.

'Google' 시스템의 경우, 24문장이 통제언어 규칙적용을 통해 번역결과가 향상되었고, 10문장은 오히려 번역결과가 개악되었다. 나머지 28문장의 경우는 점수의 차이가 없었다.

	Systran(규칙기반)	Google(통계기반)
개선	30	24
개악	2	10
변화없음	30	28

표 8 : 번역수월성에 영향을 미친 통제독일어 규칙의 수

이 실험결과를 보면, 'Systran'시스템의 경우 통제독일어 규칙으로 재작성된 62문장 중 30문장, 즉, 48.38%의 번역결과가 향상되었고, 다른 30문장의 경우는 번역결과가 원문의 번역결과와 큰 차이가 없었다. 통제독일어의 적용으로 오히려 번역결과가 나빠진 문장은 총 2문장으로 전체의 3.22%에 지나지 않았다. 반면 'Google'시스템의 경우에는 통제독일어 규칙으로 재작성된 62문장 중 24문장, 즉, 38.70%의 번역결과는 향상되었으나, 10문장, 즉, 16.12%의 경우에는 오히려 번역결과가 개악된 것으로 드러났다. 나머지 28문장의 경우에는 번역결과와 큰 차이가 없었다.

두 시스템의 번역결과를 모두 향상시킨 통제독일어 작성규칙과 규칙이 적용된 예문은 다음과 같다.<sup>23)</sup>

관계절의 사용대신 같은 의미의 다른 문장으로 표현

원문: Blitzgerät nicht zerlegen! HOCHSPANNUNG! Im Geräteinneren befinden sich keine Bauteile, die von einem Laien repariert werden können.

Systran: Flash apparatus do not divide! HIGH VOLTAGE! In the equipment inside no construction units rule, those by a layman to repair itself can

Google: Flash device is not decompose! HIGH! Device Inside there are no parts, From a layman can be repaired.

통제독일어: Blitzgerät nicht zerlegen! HOCHSPANNUNG! Im Geräteinneren befinden sich keine Bauteile. Sie können nur von einem Experten repariert werden..<sup>24)</sup>

Systran: Flash apparatus do not divide! HIGH VOLTAGE! In the equipment inside are no construction units. They can be repaired only by an expert.

23) 이 규칙들은 모두 2번 이상 적용되었다.

24) 이 예문의 경우 원문의 관계절을 대명사를 사용하여 두 개의 문장으로 분리하였다. 대명사를 될 수 있는대로 사용하지 않는 것이 기술문서 작성의 원칙이지만, 본 문장에서는 대명사와 선행사간에 다른 명사가 들어가서는 안된다는 대명사 사용규칙에 위배되지 않으므로 문제가 없다.

Google: Not decompose in flash! HIGH! The devices are no interior components. They can only be repaired by an expert.

긴 문장을 두 문장으로 분리<sup>25)</sup>

원문: Wenn Sie nicht speichern möchten, können Sie mit der Taste „MENU“ zurückspringen zum entsprechenden „Menü-Hauptpunkt“ oder mit der Taste „EXIT“ aus dem Menü ganz aussteigen.

Systran: If you did not like to store, you can jump back with the key „MENU“ to the appropriate „menu main point“ or with the key „EXIT“ of the menu to step out completely.

Google: If you do not want to save, you can use the key, " 'MENU" back to the "" main-menu item", "or with the button" 'EXIT' "from the menu entirely off.

통제독일어: Wenn Sie nicht speichern möchten, können Sie mit der Taste „MENU“ zurückspringen zum entsprechenden „Menü-Hauptpunkt“. Oder Sie können mit der Taste „EXIT“ aus dem Menü ganz aussteigen.<sup>26)</sup>

Systran: If you did not like to store, you can jump back with the key „MENU“ to the appropriate „menu main point“. Or you can step out with the key „EXIT“ of the menu completely.

Google: If you do not want to save, you can use the key, " 'MENU" back to the "" main-menu item". Or you can use the key, " 'EXIT'" from the menu entirely off.

위 두 규칙의 경우는 모두 문장길이를 짧게 하면서 문장을 뉘 수 있으면 단순하게 작성한 경우였다.

### 병렬규칙 준수

원문: Die Themen/Kapitel können mit dem roten dreieckigen Cursor über

25) Lehmrdorfer(1996)는 기술문서 텍스트의 문장을 작성할 때 짧게 쓰라는 지침을 내리고는 있지만, 'AECMA'의 예에서와 같이 구체적으로 몇 단어 이하로 작성하라는 지침은 없다. 문장의 길이에 따라 이해도와 번역수월성이 어떻게 변화하는지에 대한 연구는 별도로 진행되어야 한다.

26) 이 문장에 적용된 규칙은 하나의 긴문장을 다수개의 짧은 문장으로 작성하는 규칙으로 볼 수도 있고, 병렬 Koordination구조를 사용할 경우에 병렬구조로 연결되는 성분의 문법 범주 Grammatische Kategorien를 같게 하라는 규칙으로 볼 수도 있다.



die ovale Cursorsteuertaste oder mit den Zifferntasten angesteuert werden.

Systran: The topics/chapters can be addressed with the red triangular cursor over the oval cursor expensive key or with the figure keys

Google: The themes / Chapter can with the red triangular cursor on the oval cursor control button Or with the number keys are activated.

통제독일어: Die Themen/Kapitel können mit dem roten dreieckigen Cursor über die ovale Cursorsteuertaste angesteuert werden. Die Themen/Kapitel können auch mit den Zifferntasten angesteuert werden.

Systran: The topics/chapters can be addressed with the red triangular cursor over the oval cursor expensive key. The topics/chapters can be addressed also with the figure keys.

Google: The topics / Chapter can with the red triangular cursor on the oval cursor control button activated. The topics / chapters can also use the number buttons to be activated.

위의 원문에서 병렬접속사 ‘oder’는 전치사구 ‘mit dem roten dreieckigen Cursor über die ovale Cursorsteuertaste’와 ‘mit den Zifferntasten’을 연결하고 있다. 첫 번째 전치사구는 구조내에 다른 전치사구 ‘über die ovale Cursorsteuertaste’를 포함하고 있는 구조로서 문장을 분석할 때에 중의성을 일으키게 된다. 따라서 이러한 구조는 병렬접속사로 연결하지 못하고, 별개의 문장으로 분리하여야 한다. 이와 같이 문장 분석시의 모호성을 원문 작성단계에서부터 원천적으로 배제할 경우, 규칙기반 기계번역시스템의 경우에는 분석의 정확도가 훨씬 높아지게 되어 정확한 번역결과를 내놓게 될 가능성이 크다. 또한 통계기반 번역시스템의 경우에도, 문장의 길이가 짧아지면서 번역문장 어순의 재배열 단계인 디코딩 decoding 결과가 향상될 가능성이 높아진다.<sup>27)</sup>

주어를 문두로 한 일반적인 어순

원문: In der Regel hängt das von den örtlichen Empfangsverhältnissen ab.

Systran: Usually depends on local receipt conditions.

Google: As a rule, which hangs from the local reception conditions.

통제독일어: Das hängt in der Regel von den örtlichen Empfangsverhältnissen ab.

27) Vogel (2003) 참조.

Systran: That usually depends of local receipt conditions.  
Google: This is usually issued by the local reception conditions.

접속사 생략의 금지

원문: Bleibt diese Funktion abgeschaltet, werden solche Filmformate nicht automatisch angepasst.  
Systran: This function remains switched off, such film formats is not automatically adapted.  
Google: If this feature off, such film formats are not automatically adjusted  
통제독일어: Wenn diese Funktion abgeschaltet bleibt, werden solche Filmformate nicht automatisch angepasst.  
Systran: If this function remains switched off, such film formats is not adapted automatically.  
Google: If this feature remains disabled, such film formats are not automatically adjusted.

이번에는 규칙기반 기계번역시스템인 ‘Systran’의 번역수월성만 향상시킨 규칙들을 살펴보도록 한다. 해당 규칙들은 다음과 같다.

화행에 맞는 문장형태 사용

원문: Ihr Gerät ist für den Betrieb in trockenen Räumen bestimmt.  
Systran: Your equipment is intended for the enterprise in dry areas  
통제독일어: Ihr Gerät in trockenen Räumen benutzen!  
Systran: Their equipment in dry areas use!

분사수식어(과거분사 형용사형) 사용금지

원문: Die Ausschaltzeit schaltet das TV-Gerät zu einen von Ihnen eingegebenen Zeitpunkt automatisch aus.  
Systran: Fall time scolded the TV-equipment to a time entered by you automatically out.  
통제독일어: Wenn Sie die Ausschaltzeit eingeben, wird das TV-Gerät in dem Zeitpunkt automatisch ausgeschaltet.  
Systran: If you enter fall time, the TV-equipment is switched off automatically in the time.

다음의 규칙들은 통계기반 기계번역시스템인 ‘Google’시스템의 번역수월성만을 향상시킨 규칙들이다.

행위 수동금지

원문: In das Batteriefach der Fernbedienung dürfen keine Akkus eingelegt werden.

Google: In the battery compartment of the remote control can not be inserted batteries.

통제독일어: Keine Akkus in das Batteriefach der Fernbedienung einlegen

Google: No batteries in the battery compartment of the remote control

분사수식어(zu부정사) 사용금지

원문: Schalten Sie das zu decodierende Programm ein (z.B. PREMIERE)

Google: Turn to decodierende program (eg PREMIERE).

통제독일어: Um ein Program (z.B. PREMIERE) zu decodieren, schalten Sie das Program ein.

Google: To a program (eg PREMIERE) to decode, shut down the program.

이번에는 두 시스템의 번역수월성을 오히려 하락시킨 규칙에 대해서 알아보기로 한다. 두 시스템의 번역수월성을 공통으로 저하시킨 규칙은 없었다. ‘Systran’ 시스템의 번역수월성을 떨어뜨린 규칙은 한 개가 존재하였는데, 이 규칙은 ‘Google’ 시스템에서는 오히려 번역수월성을 향상시킨 ‘행위수동 금지’ 규칙이었다. 이 규칙이 규칙기반 시스템의 번역결과를 오히려 개악시킨 이유는 좀 더 깊이 분석해봐야 할 것이나, 아마도 목적어를 문두에 위치시키고 동사가 후치하게 되는 명령형 구조에 대해, 규칙기반 기계번역시스템의 분석모듈이 분석의 불안정성을 보이는 것으로 추측된다.<sup>28)</sup>

‘Google’ 시스템의 번역수월성을 하락시킨 규칙 중의 하나는 ‘화행에 맞는 문장형태로 작성’ 규칙이었다. 여기서 재미있는 현상은 이 규칙은 ‘Systran’의 경우에는 오히려 번역수월성을 향상시켰던 규칙이라는 점이다. 번역수월성의 하락원인은 아마도 학습코퍼스에서는 직접명령형이 아닌 간접명령과 유사한 표현이 더 많이 사용되었기 때문이 아닐까 예측해본다. 그러나 이 경우도 역시 통계기반 기계번역시스템의 특성상 번역결과에 대한 원인을 규명해볼 수가 없기 때문에, 확실한 원인을 규명하기는 어렵다.

28) ‘Systran’시스템은 상용시스템이므로, 번역의 중간과정인 분석단계 등을 살펴볼 수 없다.

지금까지 본 실험을 통해 규칙기반 기계번역시스템과 통계기반 기계번역시스템의 번역수월성을 공통으로 향상시킨 규칙, 한 쪽만 향상시킨 규칙, 번역수월성을 오히려 하락시킨 규칙에 대해 알아보았다. 이를 표로 정리해보면 다음과 같다.

	공통	Systran	Google
향상	관계절 사용금지 긴 문장 사용금지 병렬규칙 준수 주어를 문두로 한 일반적인 어순 사 용 접속사 생략금지	화행에 맞는 문장형태 사용 분사수식어(형 용사형 과거분 사) 사용금지	행위 수동금지 분사수식어(zu 부정사) 사용금 지
하락	없음	행위 수동금지	화행에 맞는 문장형태 사용 여러 문장으로 분리

표 9 : 번역수월성에 영향을 미치는 통제독일어 규칙

### V. 맺는말

본 논문에서는 최근 독일학계에서도 활발히 논의되고 있는 통제독일어와 번역수월성의 관계에 대해 다루었다. Lehrndorfer가 1996년에 소개한 통제독일어 개념은 이제 독일에서 기술문서를 작성할 때 언어학적 지침으로 통용되고 있다. Lehrndorfer는 자신의 저서에서 통제독일어가 문서의 가독성을 높일 뿐만 아니라 기계번역 시스템의 번역수월성 향상에도 도움이 된다고 주장하였다. 본 논문에서는 Lehrndorfer가 통제독일어의 개념을 소개한지도 벌써 십여년이 지난 지금, 기계번역의 측면에서 바라본 통제독일어의 번역수월성에 대한 관계를 실험을 통해 고찰하였다. 독-영 기계번역 실험결과 규칙기반 기계번역시스템의 경우 모든 문장이 통제독일어로 작성된 경우, 그렇지 않은 경우와 비교하여 9.28% 번역률이 향상됨을 알 수 있었다. 통계기반 기계번역시스템의 경우도 모든 문장이 통제독일어로 작성된 경우 그렇지 않은 경우와 비교하여 번역률이 4.84% 향상됨을 볼 수 있었다. Lehrndorfer는 통제독일어가 기계번역시스템의 번역수월성을 향상시킬 것으로 예상하였으나, 이는 'METAL'이라는 당시 지멘스에서 개발된 규칙기반 기계번역시스템으로 예측한 것이다.<sup>29)</sup> 그러나 우리

29) Lehrndorfer/Mangold(1997) 참조

는 이번 실험을 통해서 통제독일어는 규칙기반 기계번역시스템 뿐만 아니라 통계기반 기계번역시스템에서도 번역수월성을 향상시킨다는 사실을 확인할 수 있었다. 물론 규칙기반 기계번역시스템의 경우와 비교하여 통계기반 기계번역시스템의 경우, 번역수월성 향상의 정도는 아직 비교적 낮음도 확인할 수 있었다. 이것은 아마도 Lehrndorfer의 통제독일어 규칙이 규칙기반 기계번역시스템만을 고려하여 설계되었기 때문으로 생각된다. 우리는 이번 실험에서 또한 규칙기반, 통계기반 기계번역시스템 모두의 번역수월성을 향상시키는 규칙이 무엇인지를 알게 되었다. 이러한 규칙들은 대부분 특정 문장구조에 관한 것이라기 보다는 전반적인 문장의 길이와 단순성에 관련된 규칙이라고 할 수 있다. 따라서 향후 통제독일어를 좀 더 개선하고, 한국어에도 이러한 개념을 적용하기 위해서는 문장의 길이 및 복잡성이 번역수월성에 미치는 영향에 대한 연구가 뒤따라야 할 것이다.

## 참 고 문 헌

- 박주형/이창우/강명주 (2001): "자동 번역과 CAT의 현황과 전망", 정보과학회지 제 19권, 제10호, S.19-26
- 최명원/최지영 (2008): "통제언어의 관점에서 본 기술문서의 화행", 제1회 통제언어 산학협동 심포지엄 '통제언어 모형 개발의 필요성과 방향' 논문집, S.41-53
- 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 2007
- Göppferich, S. (1995): Textsorten in Naturwissenschaft und Technik. Pragmatische Typologie - Kontrastierung - Translation. Tübingen: Narr.
- Hewavitharana et al. (2007): "Experiments with a Noun-Phrase driven Statistical Machine Translation System", Proceedings of MT-Summit 2007
- Krenn, B. / G. Erbach (1994): Idioms and Support Verb Constructions, in German in Head-Driven Phrase Structure Grammar, Hrsg: Nerbonne et al., S.341-363
- 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>
- Ley, M. (2005): Kontrollierte Textstrukturen. Ein (linguistisches) Informationsmodell für die Technische Kommunikation. Dissertation, Justus-Liebig-Universität Gießen.
- Mitamura, T. / Nyberg, E. H. (1995): "Controlled English for Knowledge-Based MT: Experience with the KANT System", Proceedings of TMI-95.
- Mitamura, T. (1999): "Controlled Language for Multilingual Machine Translation". Proceedings of MT-Summit 1999
- 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
- 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.
- Roh et al. (2003): "For the Proper Treatment of Long Sentences in a Sentence Pattern-based English-Korean MT System", Proceedings of MT-Summit 2003
- Vogel, S. (2003): "SMT decoder dissected: word reordering", Proceedings of 2003 International Conference on Natural Language Processing and Knowledge Engineering

## Zusammenfassung

### Über die Möglichkeit zur Verbesserung der Übersetzbarkeit durch Kontrolliertes Deutsch

Hong, Munpyo

Die vorliegende Arbeit befasst sich mit Kontrolliertem Deutsch, das u.a. von Lehrndorfer in ihrem Buch „Kontrolliertes Deutsch (1996)“ vorgestellt wurde. Kontrolliertes Deutsch ist eine künstlich vordefinierte Subsprache, die dazu dienen soll, die Lesbarkeit, die Verständlichkeit und die Übersetzbarkeit eines Textes zu erhöhen. In dieser Arbeit wurde ein Experiment unternommen, um zu prüfen, ob Kontrolliertes Deutsch in Hinblick auf die maschinelle Übersetzung(MÜ) in welchem Maße zur Verbesserung der Übersetzbarkeit eines Textes führen kann. Für das Experiment wurden zwei unterschiedliche MÜ-Systeme, von Systran und Google, bereitgestellt. Das Systran System ist ein traditionelles regelbasiertes System. Das Google System stützt sich auf den Statistikbasierten Ansatz. In dem Experiment wurde gezeigt, dass im Fall des Systran Systems d.h., einem regelbasierten MÜ-System, das Übersetzungsergebnis mit Hilfe von Kontrolliertes Deutsch im Vergleich zu dem Originelltext um 9.28% verbessert wurde. Im Fall des Google Systems, einem statistikbasierten System, wurde das Übersetzungsergebnis von dem Text, der mit Kontrolliertem Deutsch erfasst wurde, um 4.84% verbessert. Lehrndorfer(1996) behauptete anhand des Beispiels des METAL MÜ Systems, dass Kontrolliertes Deutsch für die Verbesserung des Übersetzungsergebnisses von einem regelbasierten MÜ System einen großen Beitrag leisten kann. Unser Experiment zeigte, dass Kontrolliertes Deutsch auch bei einem statistikbasierten MÜ System die Verbesserung der Übersetzung hervorbringt. Darüber hinaus wurde in der vorliegenden Arbeit eine Liste von den Schreibregeln zusammengestellt, die nicht nur bei einem regelbasierten MÜ System sondern auch bei einem statistikbasierten System zur Verbesserung der Übersetzung führen. Es wurde herausgestellt, dass sich diese Schreibregeln eher auf die Länge und die Simplität eines Satzes als auf die bestimmten grammatischen Strukturen beziehen.

-----  
 핵심어: 통제독일어 Kontrolliertes Deutsch, 기계번역 Maschinelle Übersetzung, 번역  
 수월성 Übersetzbarkeit, 가독성 Lesbarkeit

필자메일: [skkhmp@skku.edu](mailto:skkhmp@skku.edu)

논문투고일: 2008.03.31./ 심사일: 2008.05.10 / 심사완료일: 2008.05.24