

Pre-editing oder Post-editing?

- 통제언어와 포스트에디팅을 통한 독-한 기계번역시스템의 실무적용
가능성에 대하여 -

홍문표(성균관대)

1. 서론

번역은 제2의 창작이라는 말도 있지만, 최근 산업분야에서의 번역개념은 제품의 기획, 설계, 생산, 판매, 애프터서비스 등과 같은 일련의 산업활동 프로세스와 동등한 지위를 지니는 하나의 독립된 단계로 인식되고 있다. 번역작업은 더 이상 영세한 번역업체나 프리랜서 번역사에 의해서만 수행되지 않고, 글로벌 현지화 전문업체 Localization Vendor 등에 의해 전문적이고 대규모로 수행되는 추세이다. 이러한 이유 등으로 번역 혹은 현지화¹⁾ 전문업체들은 번역프로세스를 하나의 작업공정으로 인식하고, 비용절감을 위해 작업공정을 합리화하고 최적화하기 위한 노력을 기울이고 있다. (vgl. Garcia(2009), Pfeiffer(2003))

번역공정을 최적화하고 비용을 절감하기 위한 하나의 방안으로 널리 사용되는 것이 기계번역 Maschinelle Übersetzung 시스템 및 컴퓨터보조 번역도구 Computer-aided Translation System(CAT)이다. 이 중 컴퓨터보조 번역도구는 이미 번역메모리 Translation Memory시스템을 중심으로 널리 사용되고 있다. 그러나 기계번역시스템은 성능의 미흡함으로 인해 아직도 일반 사용자는 물론 많은 전문 번역가들에 의해 외면받고 있는 실정이다.²⁾ 그럼에도 불구하고 급증하고 있는 번역물의 양과 번역 대상 언어쌍의 증가는 기계번역시스템의

1) ‘현지화 Lokalisierung’ 개념은 엄밀히 말하여 ‘번역’의 개념을 포함하고 있다. 현지화 작업은 문서의 현지화 뿐만 아니라 도량형의 현지화, 제품 디자인이나 기능 등을 판매하고자하는 현지인들의 기호에 맞게 변경하는 작업도 포함한다. 예를 들어 가전제품을 중국에 판매할 때 중국인들의 기호에 맞게 제품에 붉은색을 많이 사용하는 것도 일종의 현지화 작업이라고 볼 수 있다.

2) 그러나 기계번역시스템은 국내에서도 이미 완벽한 번역을 위한 용도로는 아니지만 대략의 정보파악을 위한 용도로는 널리 사용되고 있다. 한국특허청에서 외국 특허심사관들을 위해 제공하고 있는 특허자동번역시스템이 그 좋은 예이다. (vgl. Hong et al. (2005))

활용을 불가피하게 하고 있다. 번역메모리 시스템이 번역작업을 위해 실질적인 도움을 줄 수는 있지만, 수많은 언어쌍과 전문분야에 대해 번역메모리를 구축하는데 많은 시간과 비용이 들기 때문에 실제 번역작업에 활용할 수 없는 경우가 많이 발생한다. 이러한 문제를 기계번역시스템은 어느 정도 해결할 수 있는 장점이 있다.

기계번역시스템을 번역실무에 적용하기 위해서는 시스템의 성능이 현재보다 월등히 높아져야 한다. 그러나 현재의 번역성능을 거의 완벽한 수준으로 향상시키는 데에는 상당한 시간과 노력이 필요할 것으로 예상된다. 따라서 이러한 딜레마를 해결하기 위해 생각할 수 있는 현실적인 해결방안은 통제언어를 통한 프리에디팅 pre-editing 방안과 기계번역결과물에 대한 포스트에디팅 post-editing 방안이다.³⁾

통제언어를 통해 번역수월성이 높아진 텍스트를 기계번역시스템을 이용하여 번역하게 되면 통제언어를 적용하지 않은 텍스트보다 번역률이 높아짐은 기존의 연구들을 통해 확인된 바 있다 (vgl. Aikawa et al.(2007), Bernth/ Gdaniec(2002), Fiederer/O'Brien(2009), 홍문표(2008a,b)). 특히 홍문표(2008b)의 연구에서는 통제언어 규칙을 적용할 경우, 사용하는 기계번역시스템이 시스템 시스템과 같은 규칙기반의 시스템이건 구글번역기와 같은 통계기반 시스템이건 상관없이 모든 경우 번역결과를 향상시킴을 보였다.

기계번역시스템의 활용을 최적화하기 위한 또 하나의 방안은 포스트에디팅을 도입하는 것이다. 포스트에디팅이란 기계번역시스템이 출력한 번역물을 번역가가 직접 수정하는 후처리작업을 말한다. 기계번역시스템의 번역결과를 포스트에디팅 과정을 거쳐 개선하는 방법에 관한 연구는 최근 기계번역 관련 연구의 큰 흐름을 형성하고 있다. (vgl. Simard et al. (2007), Schwenk et al. (2009), Lagarda et al. (2009))

본 연구에서는 통제언어에 기반한 프리에디팅 방안과 포스트에디팅 방안의 장단점을 각각 분석해보고, 이를 기반으로 독·한 기계번역시스템을 실무에 적

3) 'pre-editing'의 한국어 표현으로 '전처리'라는 용어를 사용할 수도 있겠으나, 자연언어처리분야의 'pre-processing'의 한국어 표현으로 사용되기도 하여 혼동의 우려가 있다. 이에 따라 본 논문에서는 'pre-editing' 및 'post-editing'에 대해 각각 '프리에디팅' 및 '포스트에디팅'의 용어를 사용하도록 한다.

용할 수 있는 방안에 대해 논의하도록 한다. 이를 위해 2장에서는 기계번역시스템의 연구개발현황에 대해 간단히 다루고, 3장에서는 통제언어를 통한 프리에디팅과 번역수월성의 상관관계에 대한 기존의 연구논의들을 살펴본다. 4장에서는 포스트에디팅과 기계번역의 상관관계에 대한 기존의 연구논의들을 살펴본다. 5장에서는 3장과 4장에서 제안한 방안의 타당성을 검토하기 위해 실험을 수행한다. 6장에서는 본 연구의 내용을 정리하고 향후 연구 방향을 소개함으로써 본 논문을 마무리한다.

2. 기계번역과 현지화

1950년대 구 소련과 미국의 냉전시대부터 군사적인 목적으로 시작된 기계번역에 대한 연구는 1960년대의 침체기, 1970~1980년대의 기반연구시기, 인터넷이 본격적으로 보급되기 시작한 1990년대의 부활기를 거쳐, 2000년대 들어서 본격적인 중흥기를 맞고 있다.⁴⁾ 특히 1990년대 후반부터 본격적으로 시작된 통계기반 번역방법론에 대한 연구는 기존 기계번역 방법론의 한계를 극복할 수 있을 것 같은 낙관론을 관련 연구계 및 산업계에 널리 퍼뜨리고 있다.

통계기반 방법론을 적용한 가장 대표적인 시스템은 구글 Google 번역시스템이다. 통계기반 방법론은 이중언어 말뭉치 이외에는 별다른 언어자원 linguistic resource을 사용하지 않기 때문에, 이중언어 말뭉치에 대한 접근이 상대적으로 용이한 구글사의 경우 기계번역시스템을 구축하는 데 큰 장점이 있다. 이러한 장점 때문에 구글 번역기의 경우 현재 50개의 언어쌍에 대한 기계번역 서비스를 실시하고 있다. 이 중 영어와 독일어, 영어와 스페인어, 영어와 프랑스어는 가장 번역률이 높은 언어쌍으로 알려져있다. 이들 언어쌍의 경우 번역률이 80~85%를 상회하여, 실제 번역업무에도 활용될 수 있는 수준이다.

기계번역과는 별도로 개발되어 이미 현지화 업계를 비롯한 번역관련분야에서 널리 사용되고 있는 시스템이 번역메모리 Translation Memory시스템이다. 번역메모리 시스템은 컴퓨터가 번역의 핵심부분을 수행하는 기계번역 시스템

4) vgl. 최승권 외 (2005)

과는 달리, 원문과 번역문 쌍으로 이루어진 번역메모리라고 불리는 데이터베이스에서 번역하고자 하는 문장과 일치하거나 가장 유사한 문장을 퍼지매칭 Fuzzy Matching 등의 방법을 통해 번역자에게 제시하여주는 기능을 한다. 번역자는 시스템이 제안한 번역문을 그대로 사용하거나 참조하여 번역을 수행하게 된다.

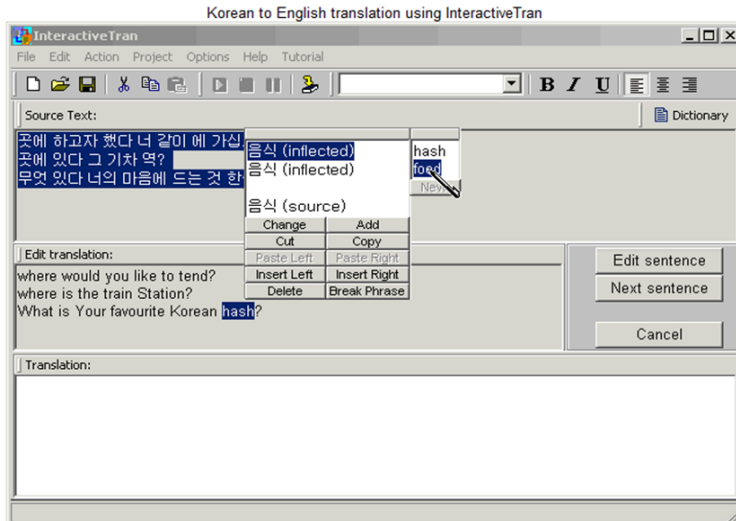
최근에는 기계번역과 컴퓨터보조 번역도구가 별개의 도구로 활용되지 않고, 서로 보완역할을 하며 함께 사용되고 있다. Garcia(2009)에서 밝힌 바와 같이 최근의 번역업무는 번역대상 텍스트를 번역메모리를 통하여 1차 매칭을 시도하고 매칭되지 않은 문장들에 대해서만 기계번역시스템을 활용하여 자동번역을 한 후 이 번역물들에 대해 전문번역가가 포스트에디팅을 수행하는 프로세스로 진행되고 있다.

3. 통제언어에 의한 프리에디팅

3.1. 상호작용 기계번역 Interaktive Maschinelle Übersetzung

기계번역의 성능을 최적화하기 위한 여러 가지 방안 중 하나는 기계번역시스템과 사용자간의 상호작용 Interaktion을 통해 기계번역과정에 사용자가 개입하는 것이다. 이는 주로 사용자, 즉 번역자가 출발언어 Quellsprache와 목표언어 Zielsprache를 모두 잘 이해하고 있어야 적용할 수 있는 방법이다. 이 방법에서 사용자는 기계번역의 중간 단계, 예를 들어 형태소분석이나 구조분석, 변환 등의 번역 중간단계에 개입하여 분석이나 변환 상의 오류를 직접 수정함으로써 기계번역시스템이 자체적으로 출력할 수 있는 번역결과보다 더 향상된 번역결과를 출력할 수 있게 한다(vgl. Kim et al.(2007)).

아래의 <그림 1>은 미국의 ‘Interactive Tran’사에서 개발한 한-영 상호작용 기계번역시스템이다. 이 시스템은 분석단계에서 사용자가 개입하는 것이 아니라 변환단계에서, 좀 더 구체적으로는 대역어 선택 단계에서 한국어 단어에 대해 알맞은 영어단어를 직접 선택하게 하여, 기계번역과정에서 매우 어려운 단계인 대역어 선택의 문제를 사용자가 원하면 직접 해결할 수 있도록 했다.



<그림 1> 상호작용 기계번역시스템의 한 예

이 방법을 적용하기 위해서는 기계번역시스템의 번역과정에 개입할 수 있는 인터페이스가 마련되어 있어야 한다. 그러나 현존하는 대부분의 상용 기계번역시스템은 이러한 인터페이스가 없으며, 사용자는 번역결과물에 대해 스스로 포스트에디팅 할 수밖에 없는 상황이다. 이러한 이유에는 여러 가지 기술적인 요인이 있겠지만, 형태소분석과 구조분석 등과 같은 번역의 중간단계는 일반인들이 이해하기 어려운 형태와 구조로 되어있어, 직접 개입하여 수정하는 작업이 쉽지 않은 일이기 때문이다. 물론 최근 국내의 모기업에서 운영하고 있는 “노띠(notii.co.kr)” 학술논문 자동번역사이트에서는 사용자들이 형태소분석과 구조분석 등의 중간단계에 직접 개입하여 기계번역결과를 개선할 수 있는 인터페이스를 제공하고 있다. 그렇지만 이러한 방법을 통한 기계번역 성능의 개선은 여전히 쉽지 않아 보인다.

3.2. 통제언어와 번역수월성

3.2.1. Bernth/Gdaniec(2002)의 ‘Mtranslatability’

통제언어 *Kontrollierte Sprache*는 기계번역의 성능을 최적화할 수 있는 일종의 프리에디팅 과정으로 이해할 수 있다. 통제언어란 자연언어의 어휘부와 통사부를 인위적으로 통제함으로써 자연언어가 가지고 있는 모호성을 최대한 없애고 이를 통해 문서의 가독성 및 번역수월성을 향상시키기 위해 개발된 일종의 하위언어 *Subsprache*이다. 영어의 경우 ‘AECMA Simplified English’, 독일어의 경우 ‘Kontrolliertes Deutsch’가 대표적인 통제언어의 예이다.⁵⁾

통제언어를 통해 기계번역의 성능을 떨어뜨릴 수 있는 요인들, 예를 들어, 형태나 구조적 모호성을 최대한 배제한 원문을 작성하게 되면 기계번역시스템의 적용시 최적의 번역결과를 기대할 수 있다. Bernth/Gdaniec(2002)의 연구에서는 이와 같이 기계번역시스템을 통한 번역수월성을 ‘Mtranslatability’라는 용어를 통해 나타낸다.

Bernth/Gdaniec(2002)은 영-독 기계번역시 ‘Mtranslatability’를 저해시키는 요인을 찾아내고 이를 개선할 경우 번역률이 향상될 수 있음을 보였다. 이들도 ‘Mtranslatability’를 향상시키는 하나의 방법으로 사용자의 번역단계의 개입을 통한 상호작용 기계번역을 언급하고 있으나, 원문의 작성단계에서부터 기계번역의 수월성을 고려한 원문작성이 기계번역결과를 보다 더 향상시킬 수 있다는 주장을 하고 있다. 이들이 영-독 기계번역에서 번역수월성을 향상시킬 수 있다고 본 통제언어 쓰기 규칙은 크게 ‘비문사용금지’, ‘모호한 구조의 사용금지’, ‘스타일 규칙’, ‘문장기호규칙’, ‘스펠규칙’ 등으로 나눌 수 있다.

현존하는 상용 기계번역 시스템들은 대부분 규칙기반이나 통계기반 방법론에 따라 개발되었다. 따라서 특히 규칙기반 시스템의 경우, 입력문의 분석과정에서 오류가 발생하면 번역결과에 치명적인 영향을 미치게 된다. 통계기반 방법론은 문장의 분석과정을 생략하고 통계알고리즘에 의해 계산된 번역결과를 출력하기 때문에 규칙기반 시스템에 비해 상대적으로 분석오류에 의한 번역결과에 영향도는 적지만, 비문법적으로 작성된 원문은 번역모델 Translation Model을 적용하기 쉽지 않거나, 번역모델이 적용되더라도 잘못 적용될 가능

5) 통제영어와 통제독일어에 대해서는 각각 AECMA(1995), Lehnrdorfer(1996) 참조. . 통제언어 개념에 대한 국내의 주요 연구논문으로는 권민재 외(2008), 류수린 외(2008), 류수린/정동규(2008), Hong/Kim(2008)등이 있으며, 최명원(2009)에서는 통제언어와 포스트에디팅 개념에 대해 다루고 있다.

성이 높아 마찬가지로 번역이 실패할 가능성이 매우 높다. 이들의 연구에서 제안하는 ‘Mtranslatability’를 높이기 위한 첫 번째 규칙은 다음과 같다.⁶⁾

규칙 1: 비문법적 구조의 사용을 피하라

예문 1: *Woven of combed cotton, you will love our sweater’s soft feel.

MÜ 17): Wenn Sie von gekämmter Baumwolle gewoben werden, werden Sie das weiche Gefühl unseres Pullovers lieben

규칙1의 예문에서 영어 원문은 ‘woven’으로 시작하는 종속절의 주어와 주절의 주어가 일치하지 않는 비문이다. 이 문장을 규칙기반 기계번역시스템을 통해 번역하면 위의 결과와 같이 ‘gewoben werden’의 주어를 시스템은 ‘Sie’로 추측하여, 주절의 주어와 일치시킨다. 이 번역결과는 사실 기계번역시스템의 관점에서 보면 완벽하게 번역된 문장이다. 왜냐하면 영어원문의 종속절에서 주어가 생략되었기 때문에 주절의 주어와 일치될 것으로 판단하여 주절의 주어를 독일어의 종속문장에서 접속사와 함께 복원했기 때문이다.

영어원문을 쓰기규칙 1을 적용하여 재작성하고 재작성된 문장을 기계번역한 결과는 다음과 같다.

예문 1-1: Woven of combed cotton, this sweater will delight you with its soft feel

MÜ 1-1: Wenn dieser Pullover von gekämmter Baumwolle gewoben wird, wird er Sie mit seinem weichen Gefühl erfreuen

규칙기반 기계번역시스템은 구조분석 과정에서 영어원문의 종속절에서 주어가 없음을 파악하고, 주절의 주어와 일치시켰다. 독일어의 생성단계에서는 종속절의 접속사 및 주어를 생성하고 주절의 주어는 문체상의 이유로 대명사 ‘er’를 사용하여 나타냈다.

6) ‘Mtranslatability’ 향상 쓰기규칙을 설명하기 위해 제시한 예문은 모두 Bernth/Gdaniec(2002)의 연구에서 제시한 예문을 인용한 것임을 밝혀둠

7) MÜ는 기계번역결과를 나타냄

Bernth/Gdaniec(2002)이 제안한 ‘모호한 구조의 사용을 피하라’는 규칙은 크게 다시 등위구문과 관련된 규칙, 영어의 ‘-ing’형태의 단어와 관련된 규칙, 명사수식어와 관련된 규칙, 대명사와 관련된 규칙, 기타 규칙 등으로 나눌 수 있다.

기계번역시스템이 잘못된 번역을 출력하는 여러 가지 이유 중 가장 큰 이유로 들 수 있는 것이 모호한 문장 구조에 대한 분석실패일 것이다. 단어나 문장의 구조가 두 개 이상으로 분석될 수 있을 때 번역시스템은 하나의 구조만을 선택하게 된다. 이 선택과정에서 올바른 선택을 위한 여러 가지 알고리즘들이 개발되어 있으나 여전히 많은 오류를 범하고 있다. 따라서 아예 원문 작성단계에서 이러한 모호한 구조를 피하여 작성하는 것은 기계번역시스템의 번역성능을 높이는 데 큰 도움이 될 수 있다.

등위구문은 대표적으로 구조적 모호성을 발생시키는 구조이다. 구조적으로 유사한 영-독 기계번역에서도 등위구문은 모호성 해소의 실패시 오역을 출력하게 된다. 따라서 Bernth/Gdaniec(2002)은 다음과 같은 등위구문 쓰기규칙을 제안한다.

규칙 2: 등위구조의 모호성을 피하기 위해 접속사의 왼쪽에 위치한 성분의 맨 마지막 단어나 접속사의 오른쪽에 위치한 성분의 맨 첫 단어를 반복하라

예문 2: The application can use the window to establish a dialog with the user and format text responses

MÜ 2: Die Anwendung kann das Window benutzen, um einen Dialog mit einem Benutzer und Formattextantworten herzustellen

위의 영어 원문에서 ‘to’ 이하의 구조를 보면 접속사 ‘and’가 연결하는 성분이 무엇인지 원문의 내용을 정확하게 이해하지 못하면 인간 번역사조차도 알아내기 어렵다. 따라서 원문의 작성자는 연결되는 성분을 명확하게 하기 위해 아래와 같이 규칙 2를 적용하여 예문 2를 재작성할 필요가 있다.

예문 2-1: The application can use the window to establish a dialog with

the user and with the format text responses

MÜ 2: Die Anwendung kann das Window benutzen, um einen Dialog mit einem Benutzer und mit den Formattextantworten herzustellen

예문 2-1과 같이 재작성된 문장은 문장구조분석시 접속사 ‘and’에 의해 연결되는 성분이 무엇인지 계산하는 단계에서 그 후보 구조의 수를 훨씬 줄이게 됨으로 올바른 구조분석의 가능성이 높아지게 하고, 따라서 올바른 번역의 가능성이 더 높아지게 한다.

그 외 영어에서 ‘-ing’ 형태의 동사는 명사로 처리할 수도 있고 동사로 처리할 수도 있기 때문에 구조적 모호성을 초래한다. Bernth/Gdaniec(2002)의 연구에서는 ‘-ing’형태의 동사 사용에 대한 구체적 지침을 3개의 별도의 규칙으로 명시화하고 있다.

그 밖에 접속사가 생략된 절이 선행하는 명사를 수식하는 경우나 형용사구가 명사구를 뒤에서 수식할 때도 구조적 모호성이 발생할 수 있다. 이러한 구체적인 현상에 대해 Bernth/Gdaniec(2002)은 명시적인 규칙을 제시한다.

스타일과 관련된 쓰기규칙들도 ‘Mtranslatability’를 높일 수 있을 것으로 보인다. 예를 들어 문장의 길이와 관련된 규칙은 기계번역의 성능에 큰 영향을 미칠 수 있다. 이들의 연구에서 문장길이와 관련되어 주목할 만한 점은 기존의 통제언어 규칙들이 주로 지나치게 긴 문장을 제한하는데 중점을 둔 반면, 이들의 연구에서는 매우 긴 문장 뿐만 아니라 매우 짧은 문장의 사용도 금지하고 있다는 점이다. 문장이 두세 단어의 매우 짧은 수의 단어로만 구성되어 있다면 이러한 문장의 분석시 기계번역시스템은 입력문을 하나의 문장으로 분석할 것인지, 아니면 복합명사로 분석할 것인지 등의 판단을 하게 된다.

문장의 길이에 따른 기계번역시스템의 성능평가를 수행해보면 텍스트 레지스터에 따라 약간씩의 차이는 있지만 (예를 들면 신문기사의 경우 평균 12~15단어 기준) 문장의 길이가 5~6단어의 경우에 가장 번역결과가 좋고 문장이 점점 길어질수록 번역률이 점차 낮아지는 것이 일반적이다. 그러나 문장의 길이가 무조건 짧다고 해서 기계번역 결과가 좋은 것은 아니다. 오히려 2~4단어로 구성된 매우 짧은 문장의 경우 분석에 실패하여 잘못된 번역을 출력하게 되는 경우가 많다. 물론 이러한 매우 짧은 문장들은 문서 가독성이라

는 인지심리학적인 측면에서는 큰 문제를 일으키거나 인지적 부담을 가중시킬 것으로 예측되지는 않는다. 그러나 기계번역의 측면에서는 매우 중요한 측면이 될 수 있다. 이렇듯 Bernth/Gdaniec(2002)의 연구는 기존의 통제언어 연구와는 달리 가독성 측면을 완전히 배제하고 번역수월성의 측면만을 강조하고 있음을 알 수 있다.

이들의 연구에서는 통제언어 쓰기규칙이 ‘Mtranslatability’를 어느 정도 향상시키는지를 검증하기 위해 영어텍스트를 독일어, 불어, 스페인어로 각각 기계번역하는 실험을 수행하였다. 실험대상 문장은 69문장이었으며, 이 중 44문장이 통제언어 규칙에 의해 재작성되었다. 이 실험의 결과 전체 텍스트의 번역률은 평균 4-15% 향상되었으며, 통제언어로 재작성된 문장들만을 고려하였을 경우 25-36% 정도 번역률이 향상되었다.

3.2.2. 홍문표(2008b)의 연구

홍문표(2008b)의 연구에서는 Lehrndorfer(1996)가 제안한 통제독일어 규칙 중 번역수월성을 높일 수 있는 규칙을 번역률 평가를 통해 찾아내었다. 특히 규칙기반 기계번역시스템인 시스트란과 통계기반 기계번역시스템인 구글 번역기를 활용해 독-영 기계번역에서 두 시스템의 번역수월성을 공통으로 향상시키는 규칙을 다음과 같이 제시하였다.

- 관계절 사용금지
- 긴 문장 사용금지
- 병렬규칙 준수 (등위구문 작성관련)
- 주어를 문두로 한 일반적인 어순 준수
- 접속사 생략 금지

그 외에 규칙기반 기계번역시스템에서만 성능향상 효과를 보였던 규칙은 다음과 같았다.

- 화행에 맞는 문장형태 사용

- 분사 수식어 사용금지

통계기반 기계번역시스템의 성능만을 향상시킨 규칙은 다음과 같았다.

- 행위수동금지
- 분사수식어 사용금지

통계기반 기계번역시스템의 성능만을 향상시킨 규칙은 규칙기반 기계번역시스템의 성능에 무관하다고 할 수는 없다. 왜냐하면 이 실험에서 사용한 규칙기반 기계번역시스템이 이미 성능측면에서 안정화 단계에 들어가 있어서, 어려운 언어현상을 처리할 수 있도록 훈련되어있기 때문이다. 따라서 ‘행위수동금지’, ‘분사수식어 사용금지’ 등과 같은 규칙들을 적용하여 재작성된 문장은 조금 성능이 낮은 다른 규칙기반 기계번역시스템으로 번역하게 되면 번역률이 올라갈 수 있다. 이 연구에서 독일어 기술문서에 대해 통계독일어 쓰기 규칙을 적용할 경우 기존문서대비 규칙기반 시스템에서는 9.28%, 통계기반 시스템에서는 4.84%의 번역률 향상을 확인할 수 있었다.

Bernth/Gdaniec(2002), 홍문표(2008b)의 연구를 종합해 볼 때 언어쌍에 따라 차이는 있지만 적게는 약 5%에서 많게는 약 36% 정도의 번역률 향상을 기대해볼 수 있다.

4. 기계번역 포스트에디팅

포스트에디팅 Postediting이란 기계번역 결과를 수작업이나 포스트에디팅 도구를 활용하여 개선하는 작업을 의미한다. 기계번역시스템의 번역결과물은 일반적으로 최종결과물로 활용하기에는 많은 부족함이 있으므로, 포스트에디팅 작업을 통해 번역의 완성도를 높인다.

기계번역시스템이 번역실무에 투입되면서 전문번역가들이 기계번역결과를 포스트에디팅하는 일이 많아지고 있다. 많은 전문번역가들은 기계번역결과를 포스트에디팅하는 것보다는 차라리 처음부터 원문을 직접 번역하는 것을 선

호하는 경우도 있다. 이러한 이유로 최근에는 포스트에디팅 작업을 자동으로 수행하는 연구도 진행되고 있다(vgl. Simard et al.(2007), Dugast et al.(2007), Ilarazza et al.(2008)).

기계번역결과에 대한 자동포스트에디팅이 가능한 이유는 기계번역의 오류는 반복적이라는 점이다. 자동포스트에디팅은 통계기반 기계번역의 원리와 거의 유사하다. 학습코퍼스가 존재하여, 이를 통한 기계학습 등이 이루어지고 기계번역결과물이 자동포스트에디팅 모듈에 입력되면 통계기반 기계번역의 원리대로 포스트에디팅이 수행된다. 본 논문의 연구목적은 자동포스트에디팅에 관한 것이 아니므로 본 논문에서는 더 이상 다루지 않는다.

기계번역 포스트에디팅은 그 범위를 정하는 것이 쉽지 않다. 포스트에디팅이란 어느 정도의 번역수정 작업을 의미하는 것인지에 대한 사전 정의가 이루어져야 전문번역가에 의한 포스트에디팅이 가능하다. 포스트에디팅을 수행하는 전문번역가는 기계번역결과를 전면 수정할 수도 있을 것이다. 이러한 경우 우리는 이 작업을 포스트에디팅이라 부를 수 있는가 하는 질문을 할 수 있다. 전문번역가가 포스트에디팅 단계에서 이러한 작업을 수행하게 된다면, 기계번역시스템의 사용을 통해 얻으려고 하는 시간과 비용의 절감효과를 얻을 수 없게 될 것이다. 따라서 포스트에디팅 단계에서는 번역가에게 작업의 범위를 명확하게 하는 것이 필요할 것이다. Allen(2005)은 포스트에디팅 작업을 그 범위에 따라 최소 포스트에디팅 minimum PE, 부분 포스트에디팅 partial PE, 빠른 포스트에디팅 quick PE, 최대 포스트에디팅 maximum PE, 전면 포스트에디팅 full PE으로 나눈바 있다.

포스트에디팅 단계에서 또 하나 명확하게 할 것이 있다. 기계번역과 포스트에디팅을 통해 얻으려고 하는 번역의 품질을 어느 정도로 기대하는가라는 점이다. 기계번역과 포스트에디팅을 통해 전문번역가가 수행한 것 같은 완벽한 품질의 번역을 기대한다면 이는 현재 기술의 수준으로는 어려울 것이다. 그러나 전문번역가가 한 것과 똑같은 수준은 아니더라도 원문의 정보가 완벽하게 전달되고 문체의 측면에서도 어느 정도 납득할만한 수준을 목표로 한다면 기계번역과 포스트에디팅을 결합한 번역프로세스가 하나의 답이 될 수 있을 것이다.

Guzman(2007)에서 제시된 영-스페인어 기계번역 포스트에디팅의 일반적인 범위를 살펴본다면, 독-한 번역과 같은 다른 번역쌍의 경우에도 어느 정도 참

고가 될 수 있을 것이다.

- 문법오류: 예를 들어 성과 수 불일치, 잘못된 어순
- 스펠과 문장기호 오류, 대소문자 표기 오류
- 오역: 중요한 전문용어의 오역, 통사적인 문제는 없지만 원문의 의미가 잘못 전달된 경우

Guzman(2007)이 제시하는 포스트에디팅의 대상범위는 번역문장의 기본틀을 깨뜨리지 않는 범위에서 주로 명사구를 수정하는 것으로 볼 수 있다. 즉, 포스트에디팅의 범위를 번역문의 술어-논항 구조를 크게 변경하지 않는 것으로 제한한다.

기계번역과 포스트에디팅의 활용은 최근 전문번역가들의 업무를 변화시키고 있는 것으로 드러났다. 실제로 기계번역의 성능이 가장 좋은 언어쌍 중의 하나인 영어-스페인어 번역의 경우, 인간번역가들의 역할이 전문번역에서 기계번역 결과의 포스트에디팅으로 변화하고 있다. Garcia(2009)는 미래의 번역 업무는 출발언어 *Quellsprache* 와 목표언어 *Zielsprache*에 모두 해박한 전문가보다는 목표언어의 모국어화자이면서 번역텍스트의 전문분야에 대한 지식을 갖추고 있으며 기계번역시스템을 잘 활용할 수 있는 사람에 의해 수행될 것으로 예측하고 있다.

5. 통제언어와 포스트에디팅을 활용한 실험

5.1. 실험배경

Allen(2005)은 전문번역가의 경우 하루 8시간 기준 최대 2,400단어의 번역을 수행할 수 있다고 한다. 그러나 이는 잠시도 쉬지 않고 번역을 수행하였을 경우의 이상적인 수치이고 실제로는 이것보다는 좀 적은 수의 단어를 번역할 수 있을 것이다. 기계번역과 포스트에디팅을 활용할 경우, 인간번역가에 버금가는 번역품질을 유지하면서 최소 2,400단어 이상의 번역을 하루에 처리할 수

있다면 기계번역시스템을 활용하는 것이 의미가 있을 것이다.

본 연구에서는 독-한 기계번역시스템을 번역업무에 활용할 경우 어느 정도의 성능을 기대할 수 있는지를 알아보기 위한 실험을 실시하였다. 이 실험의 목적은 다음과 같다:

- 현재 독-한 기계번역시스템의 수준은 어느 정도인가?
- ‘Mtranslatability’를 고려한 통제언어 쓰기규칙을 독일어 원문에 적용할 경우 번역품질은 어떤 변화가 있나?
- 통제언어로 작성된 문장을 기계번역한 후 포스트에디팅을 적용하면 어느 정도의 번역결과 개선 및 시간의 절약을 얻을 수 있나?

이 실험에서 사용한 독-한 기계번역시스템은 구글번역기이다.⁸⁾ 구글번역기는 통계기반의 기계번역시스템으로서 현재 한국어, 독일어, 영어를 포함하여 총 50개의 언어를 양방향으로 번역할 수 있다. 이렇듯 수많은 언어쌍을 처리할 수 있는 이유는 이중언어말뭉치와 같은 언어자원만 갖춰지면 번역모델과 언어모델을 자동으로 학습할 수 있는 알고리즘이 개발되어 있기 때문이다. 따라서 어떤 언어라도 새로 추가될 경우, 그 언어로 된 코퍼스가 존재하면 번역 엔진을 개발할 수 있다.

현재 구글번역기는 영어와 스페인어, 영어와 프랑스어, 영어와 독일어간의 번역성능이 가장 높은 것으로 알려져있다. 독일어와 한국어는 상대적으로 두 언어로 구성된 이중언어말뭉치의 규모가 작기 때문에 위 언어쌍들과 같은 높은 번역률을 기대하기는 어렵다. 그럼에도 불구하고 코퍼스의 규모가 점점 커질수록 번역률도 높아지기 때문에, 불과 1년전과 비교해볼 경우 현재의 그 성능은 월등히 향상되어 있음을 알 수 있다.

현재 독일어를 출발언어로 하여 한국어로 번역되는 문서의 전문분야 중 가장 많은 수요가 있는 분야는 기계분야의 문서라고 한다.⁹⁾ 따라서 본 실험에서도 기계분야의 독일어 문서를 채택하여 이를 구글번역기를 통해 한국어로 번

⁸⁾ http://www.google.co.kr/language_tools?hl=ko

⁹⁾ 국내 최대규모의 현지화 전문업체 중 하나인 (주)솔트룩스의 담당임원과의 사신(personal communication)

역하고(번역 1로 부름), 독일어 문서를 통제언어 규칙을 적용하여 다시 구글 번역기로 번역하고(번역 2로 부름), 번역 2에 대해 독일어 지식은 전문번역가 수준은 아니지만 ZMP 수준의 한국어 모국어 화자가 포스트에디팅한 결과(번역 3으로 부름)를 비교 분석하였다.

본 실험에서는 Garcia(2009)의 예측과 유사하게 포스트에디팅을 독일어와 한국어에 모두 완벽한 전문가가 수행하지 않고, 목적언어인 한국어는 완벽하지만 출발언어인 독일어에 대해서는 상대적으로 부족한 사람을 통해 수행하였다. 이를 통해 상대적으로 독일어에 매우 능통한 인력이 부족한 한국에서 독-한 기술번역이 기계번역시스템을 활용할 경우 경제적으로 수행될 수 있는 지를 보고자 하였다.

5.2. 프리에디팅 및 포스트에디팅

본 실험에서 사용된 독일어 기술문서는 독일의 자동서터 Rolladen 조립설명서였다. 독일어 매뉴얼 원본은 총 57문장, 744개의 단어로 구성되어있다. 이를 문장당으로 환산하면 평균 13.05개의 단어로 한 문장이 구성되어있다. 우리는 이 문장을 ‘Mtranslatibility’를 높이기 위한 쓰기규칙을 활용해 재작성하였다. 쓰기규칙은 Bernth/Gdaniec(2002)의 규칙과 홍문표(2008b)에서 제시된 규칙들을 활용하였다.

원문에 쓰기규칙을 적용하여 재작성한 결과 총 59문장으로 늘어났으며, 전체 단어수는 747개였다. 문장당 평균 단어수는 12.66단어로 약간 줄어들었다. 주로 적용된 통제언어 규칙은 ‘긴문장 사용금지 규칙’, ‘복잡한 구조의 사용금지’, ‘등위구문규칙’ 등이었다. 57문장 중 통제언어 규칙이 적용된 문장은 총 18문장이었다. ‘긴문장 사용금지 규칙’을 적용하여 하나의 문장에 두 개 이상의 명제가 불필요하게 전달된 문장은 아래의 예와 같이 두 개의 문장으로 분리하였다.

(1) Es verfügt über eine Vielzahl sinnvoller Programmiermöglichkeiten und wird nach höchsten Qualitätsanforderungen hergestellt.

(1-1) Es verfügt über viele sinnvolle Programmiermöglichkeiten. Es wird

nach höchsten Qualitätsanforderungen hergestellt.

복잡한 구조의 사용을 금지하는 규칙은 불필요하게 사용된 명사구를 절의 형태로 재작성하게 하여 구조적 모호성을 배제하였다.

(2) Im Falle eines Stromausfalls bleiben die zuvor programmierten Zeiten bis zu 6 Stunden gespeichert.

(2-1) Wenn der Strom ausfällt, bleiben die vorprogrammierten Zeiten bis zu 6 Stunden gespeichert.

복잡한 구조로 분석된 자주 등장하는 문장구조는 ‘sein-zu’형태의 구조와 접속사가 생략된 부문장 구조였다. 이 구조들은 각각 (3-1)과 (4-1)의 형태와 같이 재작성 되었다.

(3) Die Steuerung ist sofort auf Beschädigungen zu prüfen

(3-1) Prüfen Sie die Steuerung sofort auf Beschädigungen.

(4) Wird die Auf-Taste bzw. Ab-Taste gedrückt, fährt der Rolladen hoch bzw. runter.

(4-1) Wenn Sie die Auf-Taste bzw. Ab-Taste drücken, fährt der Rolladen hoch bzw. runter.

본 연구의 실험에서는 위와 같이 통제언어를 통해 입력문을 재작성한 후 구글 독-한 기계번역시스템으로 초벌번역을 수행하였다. 초벌번역 결과에 대해 포스트에디팅을 수행할 수 있는 경우에는 포스트에디팅을 적용하였다. 포스트에디팅의 범위는 앞 장에서 밝힌 바와 같이 독-한 번역에서도 번역결과의 술어-논항 구조를 완전히 뒤바꾸는 작업은 번역사의 입장에서 아예 직접 처음부터 번역하는 편이 낫다라고 판단할 수 있으므로, 조사나 동사어미를 문장이 자연스럽도록 첨가하거나 변경, 삭제하는 작업 및 잘못된 용어선택의 수정으로 한정하였다. 독일어 원문을 통제언어로 재작성 후 이를 기계번역한 후 초

별번역에 대한 포스트에디팅의 예는 다음과 같다.

(5) Sollte ein Transportschaden vorliegen, so ist unverzüglich der Lieferant zu informieren. (독일어 원문)

(5-1) Wenn ein Transportschaden vorliegt, informieren Sie bitte sofort dem Lieferant. (통제언어로 재작성)

(5-2) 거기에 전송 손상, 즉시 공급 업체에게 알려 주시기 바랍니다. (초별 기계번역)

(5-3) 운반시의 손상을 즉시 공급 업체에게 통보하십시오. (포스트에디팅결과)

(5-3)에서는 (5-2)의 초별번역결과에 대해 ‘전송손상’이라는 어색한 용어를 ‘운반시의 손상’으로 수정하였으며, 이 명사구에 ‘을’이라는 목적격 조사를 첨가하였다.

포스트에디팅 결과 전체 59문장 중 15문장은 포스트에디팅을 하는 것이 의미가 없을 정도로 잘못 번역이 되어있었으나, 나머지 44문장은 비교적 간단한 포스트에디팅만을 통해 거의 완벽한 번역으로 완성할 수 있었다. 포스트에디팅을 적용한 44문장은 대부분 어색한 대역어를 자연스러운 대역어로 바꾸는 작업과 자연스러운 문장의 스타일을 위하여 ‘이’, ‘가’와 같은 격조사와 ‘은’, ‘는’ 과 같은 보조사를 교체하는 에디팅 작업이 수행되었다.

5.3. 실험결과 및 분석

원문 총 57문장을 구글 독-한 기계번역시스템을 통해 번역한 후, 2명의 평가자가 0점(번역실패)에서 4점(완전한 번역)의 스케일로 각 문장에 점수를 부여하였다. 최종 점수는 2인의 평균점으로 계산하였다. 그 결과 원문을 그대로 기계번역한 번역1의 경우 40.62%의 번역률이 계산되었다. 이는 보통 80%를 상회하는 영-독, 영-스페인어 등과 비교해볼 때 매우 낮은 수치라고 볼 수 있다. 일반적으로 기계번역시스템의 번역률을 얘기할 때 50%미만의 번역률은 실제 사용자가 체감하기에는 훨씬 더 낮게 느껴지므로, 약 40% 정도의 번역률을 보이는 시스템을 실제 업무에 투입한다는 것은 큰 의미가 없을 것으로

보였다.

이 원문을 통제언어로 재작성한 후 기계번역한 결과인 번역 2는 59.98%의 번역률이 측정되었다. 번역 1에 비해 약 20%의 번역률 향상이 이루어졌으나, 이것도 역시 실무에 투입하기에는 약간 부족해 보인다. 번역 2에 대해 ZMP 수준의 독일어 실력을 가진 대학원생이 포스트에디팅을 수행한 문장과 포스트에디팅을 할 수 없을 정도로 번역에 실패한 15문장을 모두 합쳐 번역률을 평가한 결과 82.91%의 번역률이 측정되었다. 번역에 실패한 15문장을 번역가가 직접 번역하게 되면 사실상 전체 문서의 번역률은 100%에 가깝기 때문에 현재의 독-한 기계번역시스템의 포스트에디팅에 걸리는 시간을 포함한 전체 번역속도가 시스템의 실무투입 가능여부를 결정하는 중요한 요인이 될 수 있다.

<표 1> 독-한 기계번역을 활용한 기술문서 번역률

기계번역	통제언어+기계번역	통제언어+기계번역+포스트에디팅
40.62%	59.98%	82.91%

한 문장 당 포스트에디팅을 수행하는데에는 평균 약 1.5분의 시간이 소요되었다. 비록 본 연구에서 직접 실험을 수행하지는 못했지만, 전체 문서 747 단어를 수작업으로 번역할 경우 Allen(2005)의 계산대로라면 747×12 초/단어, 즉 8,964초가 걸릴 것으로 예상된다. 기계번역과 포스트에디팅의 경우 572단어는 기계번역과 포스트에디팅을 활용하여 번역하고, 나머지 175단어는 인간 번역사가 수동으로 번역을 하게 된다. 이 경우 총 $572 \text{단어} \times 6.9 \text{초/단어} + 175 \text{단어} \times 12 \text{초/단어}$, 즉 6,047초가 소요될 것으로 예측된다.¹⁰⁾ 이는 통제언어와 기계번역, 포스트에디팅을 활용할 경우, 기존의 수작업 대비 약 32%의 시간이 단축될 수 있음을 의미한다.

10) 59문장이 기계번역되는 시간은 채 1초도 되지 않으므로 여기서는 무시하도록 한다.

<표 2> 수작업대비 기계번역과 포스트에디팅의 예상 시간절감 효과

수작업 번역	통제언어+기계번역+포스트에디팅	시간절감
8,964초 (약 150분)	6,047초 (약 100분)	약 32% 절감

이 예측은 Lehrndorfer/Mangold(1997)에서 밝힌 기계번역시스템을 번역업무에 투입했을 때 예상되는 시간절감 효과인 30%에 거의 일치하는 수치라고 볼 수 있다. 즉, 현재 수준의 독-한 기계번역시스템을 번역업무에 활용하더라도 통제언어를 통한 프리에디팅과 포스트에디팅을 적절히 활용할 경우, 기존의 수작업 번역대비 번역품질은 어느 정도 유지하면서 충분히 경제성 있는 결과를 얻을 수 있다는 결론을 얻을 수 있었다.

6. 결론

본 연구에서는 통제언어를 통한 프리에디팅과 포스트에디팅을 기계번역에 접목할 경우 기존의 수작업을 통한 번역 및 단순기계번역과 비교하여 어느 정도의 성능 향상 및 시간 절감의 효과를 가져올 수 있는지를 다뤘다. 통제언어는 문서의 가독성과 번역수월성을 향상시키기 위해 고안되었지만, 최근에 들어서는 문서를 다국어로 번역해야할 일이 많아지면서 번역수월성의 측면이 더욱 부각되기 시작하였다.

포스트에디팅은 기계번역의 결과를 수정하는 작업으로서, 그 범위를 어떻게 정하느냐에 따라 간단한 단어 교체 정도의 단순작업에서 문장의 구조 전체를 바꾸는 대규모의 포스트에디팅 작업으로 나눌 수 있다. 본 연구에서는 기계번역을 활용하여 번역시간을 단축하는 효과를 최대한으로 얻기 위해 포스트에디팅의 범위를 단어교체 및 한국어 동사의 어미나 조사 등의 교체 작업으로 한정하였다.

이 연구에서는 구글 독-한 기계번역시스템을 대상으로 기계번역시스템의 번역실무투입 가능여부에 대한 실험을 실시하였다. 현재 수준의 독-한 기계번역시스템은 그 번역률이 약 40%에 지나지 않으므로, 이 시스템을 번역업무에

그대로 투입하는 것은 큰 의미가 없을 것으로 분석되었다. 기계번역에 앞서 통제언어를 통해 프리에디팅을 하게 되면 번역률은 약 60%로 상승함을 보았다. 그러나 60%의 번역률도 번역업무에 투입하기에는 많이 부족한 성능이다. 최종적으로 이 번역결과에 대해 앞서 언급한 수준의 포스트에디팅을 수행한 경우 번역률이 약 83%정도에 다다랐다. 번역에 실패한 나머지 문장에 대해서만 인간번역가가 수동번역을 수행할 경우, 전체 번역의 품질은 모든 문장을 인간번역가가 번역한 결과와 비교하여 크게 뒤떨어지는 않으면서 번역에 소요된 시간은 약 30%정도 단축할 수 있음을 볼 수 있었다.

향후 우리는 이와 관련된 연구에서 여러 가지 도구를 사용하여 포스트에디팅에 필요한 시간과 노력을 좀 더 과학적으로 측정할 수 있는 방안에 대해 연구할 것이다. 이를 위해 고려할 수 있는 방안은 안구추적장치 아이트랙커 Eye-Tracker를 활용하거나 트랜스로그 Translog와 같은 에디팅 추적도구 등을 활용하는 방안일 것이다.

참고문헌

- 권민재/남유선/홍우평 (2008): 기술 커뮤니케이션과 통제언어 - 통제언어의 개발 사례 및 특성을 중심으로, 『독어학』 17, 45-68
- 류수린/임병화/정동규 (2008): 통제언어 모형개발의 필요성과 방향 - 기술문서에서 나타난 한국어 표현을 중심으로, 『독어학』 17, 69-98.
- 류수린/정동규 (2009): 인지언어학적 명제분석과 기술문서의 가독성, 『독어학』 20, 61-86.
- 최명원 (2009): 통제언어, 기계번역 그리고 Post-editing, 『독일언어문학』 46, 113-135.
- 최승권/홍문표/박상규 (2005): 다국어 자동번역기술, 『전자통신동향분석』 20.
- 홍문표 (2008a): 번역수월성의 관점에서 본 통제독일어, 통제영어, 통제한국어의 비교연구, 『독일문학』 107, 286-308.
- 홍문표 (2008b): 통제독일어가 번역수월성 향상에 미치는 영향에 대한 연구, 『독일언어문학』 40, 21-43.
- 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*.
- Allen, J. (2005): What is Post-editing?, Translation Automation, Newsletter published by *CROSS Language, Issue 4*.
- Bernth, A./C. Gdaniec (2002): Mtranslatability, *Machine Translation 16*.
- Dugast et al. (2007): Statistical Postediting on Systran's Rule-based Translation System, *ACL 2007*.
- Fiederer, R./ S. O'Brien (2009): Quality and Machine Translation: A realistic objective?, *Journal of Specialised Translation Issue 11*.
- Garcia, I. (2009): Beyond TM: Computers and the professional Translator, *The Journal of Specialised Translation, Issue 12*.
- Guzman, R. (2007): Manual MT Postediting, *Translation Journal, Vol.11, No.4*.
- Hong, M. / C. Kim (2008): Impact of Controlled Language on Korean-English MT, *Proceedings of PACLIC 22*.
- Ilarazza et al. (2008): Statistical Post-editing. A Valuable Method in Domain Adaptation of RBMT Systems for Less-Resourced Language, *MATMT 2008*.
- Kim, Y. / M. Hong, / S. Park (2007): CL-guided Korean English Paper Machine Translation, *Proceedings of CICLING 2007*.
- 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*.
- O'Brien, S. (2005): Methodologies for Measuring the Correlation between the Post-Editing Effort and Machine Translatability, *Machine Translation 19*.
- Simard et al. (2007): Rule-based Translation with Statistical Phrase-based Post-editing, *ACL 2007*.
- Pfeiffer, W. (2003): *Technical Communication - A Practical Approach*, Pearson Prentice Hall.

Zusammenfassung

Pre-editing oder Post-editing?

- Über die Einsatzbarkeit eines Deutsch-Koreanisch MÜ Systems mithilfe der kontrollierten Sprache und des Post-editings

Hong, Mun-Pyo (Sungkyunkwan Univ.)

Die vorliegende Arbeit befasst sich mit der Frage, ob ein Koreanisch-Deutsch MÜ System mithilfe der kontrollierten Sprache und des Post-editings für die Koreanisch-Deutsch Übersetzung der technischen Texte eingesetzt werden kann. Kontrollierte Sprache ist ursprünglich für die Lesbarkeit und Übersetzbarkeit eines Textes konzipiert. Aber der Aspekt der Übersetzbarkeit der kontrollierten Sprache ist neulich in den Vordergrund gebracht, da die Menge der zu übersetzenden Texte drastisch zunimmt.

Unter dem Begriff Post-editing versteht man einen Prozeß, in dem man das Resultat eines MÜ Systems manuell oder halb-automatisch korrigiert. In dem Post-editing kann eine simple Korrektur, wie z.B. der Wechsel eines Zielwortes, gemacht werden. Manchmal kann die Veränderung der gesamten Satzstruktur vorgenommen werden. In der vorliegenden Arbeit wurde aber nur eine einfache Korrektur, wie der Wechsel des koreanischen Zielwortes oder der Wechsel der koreanischen Postpositionen gemacht, um den Vorteil, der durch die Einführung eines MÜ Systems gebracht wird, zu gewährleisten.

In dem Experiment wurde das Google Deutsch-Koreanisch MÜ System verwendet. Die linguistische Qualität der maschinell übersetzten Texte ist zur Zeit weit entfernt von unserer Erwartung (ungefähr 40%). Wenn die Eingabe mithilfe der kontrollierten Sprache neu erfasst wird, verbessert sich die Qualität der maschinellen Übersetzung zirka um 20%. Jedoch ist die Qualität der Übersetzung noch nicht gut genug, um das MÜ System für die alltägliche Arbeit einsetzen zu können.

Wir konnten schließlich feststellen, dass das Postediting der maschinell übersetzten Texte den Einsatz eines MÜ Systems sinnvoll machen kann. Durch das Postediting konnte nicht nur die Qualität der Übersetzung drastisch verbessert werden, sondern auch die Zeit für den Übersetzungsprozeß um 30% verkürzt werden.

[검색어] 통제언어, 기계번역, 번역수월성, 포스트에디팅

Kontrollierte Sprache, Maschinelle Übersetzung, Übersetzbarkeit, Postediting

홍문표 110-745
서울시 종로구 명륜동 3가 53번지
성균관대학교 문과대학 독어독문학과
skkhmp@skku.edu

논문 접수일: 2010. 4. 23

논문 심사일: 2010. 5. 18

게재 확정일: 2010. 5. 28