

## 기계번역 시스템 측정 장치 연구<sup>1</sup>

이민행, 지광신, 정소우

연세대학교, 연세대학교, 성신여자대학교

**Lee, Minhaeng, Kwangsin Jee, and So-Woo Chung.** 1998. **A Research on Test Suites for Machine Translation Systems.** *Language and Information* 2.2, 185-220. The purpose of this research is to propose a set of basic guidelines for the construction of English test suites, a set of basic guidelines for the construction of Korean test suites to objectively evaluate the performance of machine translation systems. For this end, we constructed 650 English test sentences, 650 Korean test sentences, and developed the statistical methods and tools for the comparative evaluation of the English-Korean machine translation systems. It also evaluates the existing commercial English-Korean machine translation systems. The importance of this research lies in that it will promote an awareness of the importance and need of testing machine translation systems within the Natural Language Community. This research will also make a big contribution to the development of evaluation methods and techniques for appropriate test suites for Korean information processing systems. The results of this research can be used by the natural language community to test the performance and development of their information processing systems or machine translation systems. (**Yonsei University, Yonsei University, Sungshin Women's University**)

---

<sup>1</sup>익명의 심사위원들께 감사드린다. 이 연구는 한국전자통신연구원 부설 시스템공학연구소의 연구비 지원을 받아 1997년 3월부터 11월까지 수행되었다. 과제명은 “기계번역 시스템 평가방안 연구”였다.

## 1. 서론

### 1.1. 연구의 배경 및 필요성

넷스케이프, 익스플로러 등 웹 브라우저의 발달로, 대중들이 인터넷 상의 정보에 손쉽게 접근할 수 있는 시대에 우리가 살고 있다. 말 그대로 정보화의 시대에 들어서 있는 것이다. 그러나 인터넷을 통하여 얻어지는 대부분의 정보가 영어로 되어 있어, 영어에 대한 지식 없이는 그 정보를 소화하기가 힘든 실정이다. 이러한 사실을 고려할 때, 성능 좋은 영한 기계번역 시스템의 유용성이 과거 어느 때보다 크다고 하겠다. 그 동안 자연언어처리 분야에서 기계번역 시스템의 개발 연구에 힘써 온 결과, 현재는 몇 가지 영한 번역기가 상용화되어 있는 상태이다. 하지만, 지금까지 개발된 영한 번역기의 번역정확도나 성능에 대한 평가가 체계적으로 수행된 바도 없거니와 그 유용성도 검증된 바가 없다. 정보의 정확성이 정보 수집 시 고려되어야 할 가장 중요한 점임을 감안할 때, 상용 영한 번역 시스템에 대한 평가와 검증이 꼭 필요한 시점에 우리가 와 있다고 할 수 있다. 오역을 양산하는 영한 번역기를 통한 부정확하거나 잘못된 정보 수집은 연구 분야나 실용 분야에 결정적으로 나쁜 결과를 초래할 수 있기 때문이다.

영한 기계번역 시스템의 성능을 객관적으로 평가하기 위한 방안을 마련하는 것이 본 연구의 목적이다. 국내의 경우, 언어정보처리 기술에 대한 평가나 기계번역 시스템의 평가에 대한 연구와 기술개발이 전혀 이루어지지 않은 상태이다. 국외의 경우, 언어정보처리 기술을 객관적으로 평가하기 위한 연구가 1990년대에 들어서 활발하게 이루어져 왔다. 이와 관련한 대표적인 과제는 1993년부터 1995년까지 수행된 TSNLP라 불리는 영국, 스위스와 독일의 3개국 공동연구과제로서, 그 연구결과를 인터넷을 통해 확인해 볼 수 있다. 그 과제의 여러 성과들 중 하나인 영어, 독일어, 불어의 측정장치 구축은 본 과제의 수행에 있어 방향성을 제시한다. 그러나 이 과제의 경우에도 기계번역 시스템의 평가 방안에 대한 연구를 포괄

하지 않는다.

공동과제 TSNLP의 경우, 영어, 독일어, 불어를 대상으로 있으므로, 여기에서 개발된 기술, 예컨대, 자동 측정장치 생성 기술은 한국어 처리를 위해서는 쓸모가 많지 않다. 곧, 한국어를 처리 대상으로 하는 한국어 정보처리 기술의 평가는 별도의 과정을 통해 이루어져야 한다. 여기에 이 분야에 대한 연구의 필요성이 있다.

일종의 기초연구인 본 연구를 통해 한국어 정보처리 시스템의 평가에 대한 중요성을 정보처리 시스템 개발자들에게 인식시킬 수 있고, 본 연구의 성과를 바탕으로 하여 한국어 정보처리 시스템을 평가하는 기술개발이 국내에서도 지속적으로 이루어질 수 있다는 점에 본 연구의 의의가 있다.

## 1.2. 연구의 목표 및 범위

본 연구의 목표 및 범위는 다음과 같다. 먼저, 영어 측정장치(test suite) 구축을 위한 기본 원칙과 한국어 측정장치 구축을 위한 기본 원칙을 기술한다. 그리고 영 한 기계번역 시스템 평가를 위한 여러 가지 척도를 설정하여, 평가척도에 기초한 평가결과의 계량화 방안 마련하는 것이다. 끝으로, 이렇게 구축한 문장과 평가 척도를 기준으로 현재 상용화되어 있는 영 한 기계번역 시스템에 의한 번역결과를 분석하고 검토하고자 한다.

## 2. 영어와 한국어 측정장치의 구축

### 2.1. 영어 측정장치 구축의 절차와 방법

영어 측정장치의 구축을 위하여, 다음 (1)에 제시되는 바와 같이 다양한 문법구조와 현상들을 선정하고, 각 유형을 대표한다고 여겨지는 문장들을 균형 있게 수집했다.<sup>2</sup>

---

<sup>2</sup>영어 측정장치 구축을 위한 문법구조와 현상들의 선정시에, 언어학적인 중요성과

(1) 평가를 위한 문법구조와 현상

- 가정법
- 강조구문
- 관계절
- 복합명사구
- 복합문
- 부사
- 부정사
- 부정표현
- 분사구문
- 비교구문
- 상관사구
- 수동태
- 수량표현
- 시제
- 완료형
- 전치사구
- 진행형
- 축약형 구문<sup>3</sup>
- 관용표현

평가용 문장의 선택 시에 특별히 제약을 두지 않았다. 따라서 어휘적인 중의성, 통사적 중의성, 의미적 중의성 등이 나타나는 문장들도 포함되고, 평서문 외에 의문문이나 명령문도 포함시켰다. 단순문뿐만 아니라 복합문도 포함되고, 생략된 요소가 들어있는, 문장이해에 있어 난이도가 높은 문장들도 함께 모았다.

측정장치의 구축 시에 기계번역 시스템의 문장 패턴 인식능력을

---

빈도수를 고려한, Balkan *et al.*(1994)의 제안을 따랐다.

<sup>3</sup>축약형(Abbreviations) 구문이란 축약형 표현이 포함된 문장을 지칭한다. 문항 423의 "He still doesn't understand."이 그 예이다.

평가하기 위해 Hornby 영문법에 기반하여 아래와 같은 38 문형을 설정하고 각 한 문장씩을 문항으로 첨가하였다.

(2) 영어 문장패턴 38 가지

- P1. NP/S[+that, that] + V[+val1]
- P2. NP + V[+be] + ADJ[+pred]/NP[+pred]/ADV[+pred]
- P3. NP + V[+be] + Inf[+to]
- P4. NP + V[+val1] + ADV
- P5. NP + V[+val1] + ADJ/NP[+prn/ prn]
- P6. NP + V[+val1] + PrtP[+pres]
- P7. NP + V[+val1] + PP/NP[+prn/ prn]
- P8. NP + V[+val1] + (P+it) + S
- P9. NP + V[+val1] + Inf[+to]
- P10. NP + V[+seem/+appear/+happen/+chance] +  
(to be) + Inf[+to]
- P11. NP + V[+aux] + Inf[ to]
- P12. NP + V[+val2] + NP[+acc, +prn/ prn]
- P13. NP + V[+val2] + NP[+grnd]
- P14. NP + V[+need/+want] + NP[+grnd]
- P15. NP + V[+have/+ought] + Inf[+to]
- P16. NP + V[+val2] + Inf[+to]
- P17. NP + V[+val2] + NP[+pron, +wh]/ADV + Inf[+to]
- P18. NP + V[+val2] + S[+that]
- P19 . NP + V[+val2] + S[ that]
- P20. NP + V[+val2] + NP + S[+that]
- P21 . NP + V[+val2] + NP[+dat] + NP[+acc]
- P22 . NP + V[+val2] + NP[+acc] + PP[+to]
- P23 . NP + V[+val2] + NP[+acc] + PP[+for]
- P24 . NP + V[+val2] + NP[+acc] + PP[ to, for]
- P25 . NP + V[+val2] + NP[+acc] + ADV

- P26 . NP + V[+val2] + NP[+acc] + ADV[+pt]
- P27 . NP + V[+val2] + NP[+acc] + Inf[+to]
- P28 . NP + V[+val2] + NP[+acc] + Inf[ to]
- P29. NP + V[+val2] + NP[+acc] + PP[+as/+like/+for]
- P30. NP + V[+have] + NP[+acc] + Inf[ to]
- P31. NP + V[+val2] + NP[+acc/+gen] + PrtP[+pres]
- P32. NP + V[+val2] + NP[+prn/ prn, +wh] + Inf[+to]
- P33. NP + V[+val2] + NP + S[ that]
- P34. NP + V[+val2] + NP[ prn/+prn/+grnd] + ADJ
- P35. NP + V[+val2] + NP[ prn/+prn] + NP[ prn]
- P36. NP + V[+val2] + NP[+acc] + PrtP[ pres]
- P37. NP + V[+have/+get] + NP[+acc] + PrtP[ pres]
- P38. NP + V[+val2] + NP[ prn/+prn,+acc] + (to be) + ADJ/NP

앞서 논의한 문장패턴의 설정 시에는 다음과 같은 통사범주를 기초로 하였다.

### (3) 영어 통사범주 목록

- 명사구                    NP
- 동사                      V
  - 자동사                V[+val1]
  - 타동사                V[+val2]
- 전치사구                PP
- 형용사                  ADJ
- 부사                     ADV
- 분사                     PrtP
  - 현재분사            PrtP[+pres]
  - 과거분사            PrtP[ pres]
- 동명사                  NP[+grnd]

- 부정사                      Inf

## 2.2. 영어 측정장치의 한국어 번역

영어 측정장치의 각 평가용 문장에는 한국어의 번역을 첨부했다. 이는 기계번역시스템에 의한 번역결과를 전문인에 의한 번역결과와의 비교를 쉽게 하고, 이 비교결과를 계량화하여, 시스템의 평가에 바로 반영하기 위한 것이다. 다음의 예는 영어 평가용 문장과 그 한국어 번역을 보여준다.

(4) a.

162	I have never met a kinder man. 나는 (이보다) 더 친절한 사람을 만난 적이 없다.
-----	--

b.

232	The instructions were hard to follow. 그 지시들은 따르기가 어려웠다.
-----	--

c.

592	We like to have our friends visit us on sundays. 우리는 친구들이 일요일에 찾아주는 것을 좋아한다.
-----	---

## 2.3. 한국어 측정장치 구축의 절차와 방법

한국어 측정장치의 구축 시에 다음 (5)에 제시된 여러 문법현상을 고려하여, 해당 평가용 문장들을 균형 있게 수집하였다.

(5) 한국어 문법구조와 현상

- 평서문
- 의문문
- 명령문/청유문
- 부정문

- 피동문
- 관형문
- 동위접속구문
- 사역문
- 복합명사구
- 주제화
- 비명사구주어
- 이중주어
- 이중목적어
- 특수조사
- 시제
- 수식
- 생략현상
- 중의성
- 불완전명사
- 복수명사

## 2.4. 한국어 측정장치의 영어 번역

한국어 평가용 문장 각각에 대해서 전문인에 의한 영어번역문을 첨부하였다. 다음의 예가 그러한 사실을 보여준다.<sup>4</sup>

(6) a.

14	나는 저녁식사에 초대 받았다.
	I was invited to dinner.

b.

22	그는 푸른 옷을 입고 있는 소녀를 만났다.
	He met the girl in blue.

<sup>4</sup>여기에서 전문인이란 전문적으로 번역을 직업으로 하는 특정인이나 그룹을 지칭하기보다는 “기계”가 아닌, 영어에 정통한 “인간”을 통칭하는 것이다.



c.

587	그 시장에서 어떠한 물건도 산 적이 없습니다.
	I have never bought anything in that market.

위의 예 (6c) 에서와 같이 한국어에 특징적이며, 빈번히 나타나는 불완전명사 '일', '셈' 등이 포함된 문장들을 적지 않게 측정장치 안에 포함시킨 것이 여기에서 제시하는 측정장치의 특징 중의 하나이다.

### 3. 기계번역 평가 방법

#### 3.1. 평가의 필요성

자연언어처리 (NLP)와 기계번역 (MT) 시스템 평가는, 자연언어처리에 관한 연구가 시작되면서부터 이루어졌으나 아직도 평가의 기준과 방법에 있어서 많은 문제점을 갖고 있다. 근본적인 문제점은 사용자, 개발자, 연구가에 따라 각각 평가의 목적이 다르다는 데서 연유한다. 그럼에도 불구하고 시스템에 관한 평가는 시스템 사용자나 개발자 또는 연구자 모두에게 대단히 중요하다. 즉 시스템 평가를 통하여 사용자는 "어떤 시스템이 번역의 질과 경제성에 있어서 얼마나 효과적인가?"를 진단할 수 있으며, 개발자에게는 현재 하고 있는 일들이 시스템개발에 긍정적으로 작용하고 있는가를 알 수 있게 해주고, 연구자에게는 적용되고 있는 이론에 대한 평가와 더불어 문제점 지적을 통하여 연구의 우선 순위와 가치 있는 연구영역을 결정하는데 도움을 주기 때문이다. 앞서 언급한 바와 같이 평가는 목적에 따라 그 접근 방법도 다양하다. 그러나 가장 일반적이고 공통적인 평가의 관심은 생산된 번역물의 질에 대한 평가와 번역과정을 좀 더 효율적이게 해주는 요인들에 대한 평가일 것이다. 따라서 3.2절부터 3.4절까지는 현재 널리 사용되고 있는 기본적인 중요한 평가방법 3가지를 중심으로 그 각각의 장, 단점과 한계를 살펴보고<sup>5</sup> 3.5절에서는 두 개 이상의 기계번역 시스템을 비교 평가하기 위한 새로운 평가방법의 하나로 유형적 비교평가 방법을 제안한다.

---

<sup>5</sup>이하, 평가방법의 유형에 대해서는 주로 Arnold *et al.*(1993)을 참조하여 서술하였음을 밝혀둔다.

### 3.2. 연산평가 (Operational Evaluation)

사용자가 관심을 갖는 것은 시스템 엔진의 질이나 구조가 아니라 시스템이 전체적으로 번역을 얼마나 잘 해내느냐 하는 것이다. 평가에서도 마찬가지로 사용자는 시스템의 성능을 전체적으로 평가하는 데 관심을 갖는다. Van Slype가 1982년에 제안한 이런 종류의 평가로서 "Evaluation de l'efficacit (Evaluation of effectiveness)"가 있는데 이는 일명 "연산평가(Operational Evaluation)" 또는 "경제성 분석(Economic Analysis)"라고 불린다. 이 평가의 기본적인 개념은 MT의 "단어당 비용 (cost per word)"와 "후편집(post editing)"에 드는 비용을 계산해 보는 것이다.

이 평가의 장점은 "전체적"인 관점에서 평가할 수 있다는 것이다. 즉 이러한 평가에서는 시스템 엔진의 성능에 대한 평가로부터 직업 만족도가 번역가의 번역품질에 미치는 영향까지가 평가에 포함된다. 그러나 이 평가의 중요한 단점은 이 평가의 결과가 일반성을 갖지 않는다는 것이다. 즉 다른 사용자나 다른 시스템에 적용 가능한 결론을 끌어내기가 어렵고 개별적인 구성요소에 대한 평가가 어렵다. 그래서 이와 같은 평가를 "블랙박스(blackbox) 평가"라고도 한다. 그 외에도 시간과 비용이 많이 들고, 현실적인 요구를 맞추기가 어렵다는 문제점이 있다. 이러한 문제점은 일차적으로 사용 가능한 텍스트가 적고, 번역가를 훈련시키는데 시간과 노력이 많이 드는데 있다. 더 나아가, 번역의 질을 평가할 수 있는 기준을 정하기가 어렵고, 현재 번역기 어휘사전(system dictionary)에 특수한 어휘가 부족할 경우, 이를 보완하는 데는 시간과 훈련이 필요하다는 점에서도 문제점을 발견할 수 있다.

### 3.3. 선언적 평가 (Declarative Evaluation)

명료성(Intelligibility), 정확성(Accuracy), 문체(Style) 등을 기준으로 하여 입력과 출력을 모두 평가하는 방법이다. 명료성은 번역된 문장이 목표언어(target language)로 잘 되어있는가를 지수로 표현하고, 정확성은 원천언어로 된 텍스트 (source text)의 내용이 얼마만큼 잘 보존되어 있는가를 지수로 나타낸다. 이 평가는 출력의 질을 평가하기 때문에 "질적 평가 (Qualitative Evaluation)"라고도 한다.

이 평가방법의 장점은 비용이 저렴하고, 번역물의 질을 지수로 표시하여 한 눈에 알 수 있게 해주므로, 평가결과가 일반성을 갖는다고 할 수 있다. 특히 다른 기계번역 시스템과의 비교를 가능하게 해준다.

그러나 이 평가 또한 여러 가지 단점을 갖고 있는데, 하나는 번역물의 질에 초점을 맞추으로써 다른 중요한 요인들을 간과할 수 있다는 것이다. 예컨대 연산평가에서 하는 것처럼 상이한 번역기를 사용할 때는 그에 맞는 환경조성을 해 줌으로써 더 좋은 결과를 얻을 수 있는 것과 같은 것이다. 또한 번역물의 질만을 평가대상으로 함으로써 그에 영향을 미칠 수 있는 시스템 구조나 기계의 질 등 다른 요인들을 간과할 수도 있다는 것이다. 따라서 이러한 평가방법에서는 어떤 요인들이 번역물의 질에 영향을 미치는 가를 알 수가 없다. 그 외에도 다음과 같은 일반적이고 중요한 문제점을 갖는다. 예컨대, 어떤 두 개의 시스템을 비교하고자 할 때, 두 시스템에 공정하게 적용될 수 있는 자료를 찾기가 어렵다는 것이다. 각각의 시스템에 맞는 자료를 찾기 위해서는 또한 별도로 비용이 든다는 문제점이 있다. 또한 기본적인 평가 기준이 다소 주관적일 수밖에 없다는 것이다. 따라서 어느 정도 타당한 결과를 얻기 위해서는 상당히 많은 자료와 다수의 평가자들이 참여해야 한다. 그밖에도 평가되는 시스템이 나타내는 가장 높은 지수를 비교하여 어떤 시스템이 가장 좋은 것인가는 보여줄 수 있으나 정확한 평가는 어렵다.

### 3.4. 유형적 평가 (Typological Evaluation)

MT 시스템은 그 구조상 크게 두 가지로 구분해 볼 수 있다. 그 하나는 원천언어(source language)나 목표언어(target language)의 어떤 문법규칙도 갖지 않는 시스템으로서 비문법적인 문장도 번역해 낼 수 있다. 그런데 이런 유형에 속하는 시스템은 문법규칙을 적용하여 정확하게 번역하기보다는, 많은 자료를 분석하여 그것을 토대로 번역에 유추 적용하는 방법을 사용한다. 따라서 비문법적인 문장도 번역해내기 때문에 많은 양을 번역할 수 있다는 장점을 갖는다.

반면에 많은 시스템은 엄격한 언어규칙을 적용한다. 여기서는 비문법적인 문장은 처리를 하지 못하고, 생산되는 문장도 완벽한 목표언어로 번역되거나 아니면 전혀 번역을 못 해낸다. 시스템 개발자들이 관심을 갖는 것은 오류의 유형이다. 그래서 이 테스트의 도구가 되는 측정장치(test suit)는 의도적으로 만들어진 구조를 갖는 문장으로 구성된다. 이 방법은 시스템 개발자에게는 어떤 구조와 구조의 조합이 문제점을 갖는지를 분명하게 알 수 있게 해 줌으로써 기술개발을 용이하게 해주고, 최종사용자에게는 측정장치 평가를 용이하게 해주고 또 똑같은 측정장치를 다른 시스템에도 적용하여 비교할 수 있도록 해준다. 그러나 이것은 다음과 같은 두 가지 큰 문제점을 갖는다.

첫째는, 시스템이 잘 되어 있고 버그가 나지 않을 때만이 다른 유사한 문장에도 잘 적용이 될 수 있다는 점이다. 또한 측정장치 자체가 실제로 사용되는 텍스트가 갖는 문제점, 곧 실생활에서의 비문법적인 언어표현의 사용과 관련한 문제점을 보완해 줄 수 없다는 것이다.

또 하나는, 중요한 모든 구조를 갖춘 측정장치를 만드는 일이 쉬운 일이 아니다. 또한 원래 이 방법은 한가지 언어를 대상으로 하는 자연언어처리시스템을 위해 만들어진 것이므로 기계번역에 사용하려면 두 언어차이가 반영된 문장들이 보장되어야 한다.

지금까지 언급한 중요한 3가지 평가기준 외에도 장기적으로 볼 때 번역물의 질에 영향을 미칠 수 있는 중요한 여러 가지 요인들이 있다. 기계번역 시스템이 적용되면 사람이 하는 일에 변화가 생기게 되는 데 이것 또한 평가에서 중요한 요인이 될 수 있다. 즉 번역가가 기계에 의해 나온 초벌번역을 수정하는 일은 사람에 의한 초벌번역의 경우와 다를 것이고 또 더 많은 오류가 예상된다. 그 외에도 다른 소프트웨어로의 이식가능성, 사용자 인터페이스의 기능과 질, 텍스트를 입력하는데 필요한 조건과 같은 환경적인 조건들과 속도, 언어규칙의 적용범위, 명료성, 단일성 등이 평가에서 고려되어야 한다.

### 3.5. 유형적 비교평가 방법

이상에서 논의한 세 가지 기계번역 평가 방안 중에서 언어학적인 지식을 활용하기에 가장 적합한 평가방법은 측정장치 기반의 유형적인 평가이다. 본 연구는 이 유형적 평가방법을 보완하는 방향에서 유형적 비교평가 방법을 제안하고자 하는 것이다. 유형적 비교평가 방법이란, 측정장치의 각 평가문항에 포함된 전문인에 의한 번역을 기계에 의한 번역결과와 비교하고, 이 결과자체를 여러 가지 평가기준에 의해 일차적으로 개별 평가한 후에, 그 개별평가 결과를 계량화하여 종합적으로 평가하는 방법을 일컫는다. 구체적인 평가방안과 평가절차에 대해서는 다음 장에서 논의한다.

## 4. 상용 영-한 기계 번역시스템 평가

### 4.1. 평가기준 및 방안

이 장에서는 상용 영-한 기계 번역 시스템 트래니 96과 앙코르 2.0을 계량적으로 평가하기 위한 평가기준 및 방안에 대해서 논의한다.<sup>6</sup>

---

<sup>6</sup>여기에서 평가의 대상으로 정한 “트래니(Trannie) 96”과 “앙코르 (Enkor) 2.0”은

장석진(1998)에서는 일반 번역의 필수적인 평가기준으로서 제안된 명제내용, 개체, 극성, 시제, 수식관계, 화행, 담화연결관계 등이 제안되고, 각 기준의 경.중에 따라 가중치가 부여된다. 그런데, 이 기준들은 중간언어적인 표현체에 적용되기 때문에, 이 기준들을 그대로 우리의 목적을 위해 그대로 수용하는 것이 간단치 않다. 따라서 문장의 의미와 관련하여, 중요하게 여겨지는 기준인 명제내용을 우리의 평가방안에도 핵심기준으로 받아들여 가장 높은 가중치를 부여한다. 앞장에서 논의한 바와 같이 본 연구에서는 기계번역 시스템 평가방안으로서, 번역결과를 측정장치에서 제시된 전문가번역과 비교하여 시스템을 평가하는 유형적 비교평가 방법을 택한다. 이 평가방법을 계량화하기 위해서 별도의 평가기준을 필요로 하는데, 본 연구에서는 아래의 (7)에 제시된 바와 같이 크게 절대적인 기준으로서 명제내용, 문법구조, 단어인식, 완전성 등의 기준과 상대적인 기준으로서 번역속도를 제안한다. 이들 기준들은 각각 그 중요도에 따라 상이한 가중치를 가진다.<sup>7</sup> 각 평가기준은 6 등급으로 세분되어 5, 4, 3, 2, 1, 0 등의 평가점수를 부여받는다. 상대적인 기준인 번역속도의 경우, 다른 시스템과의 비교평가가 필요할 경우에만 전체자료의 번역속도에 있어서의 상대적인 차이에 따라 평가점수를 부여한다. 그러나, 이 경우 개별 평가항목에 대해 특정한 평점을 부여하기 보다, 전체평점의 계산 시에 일괄적으로 평점을 부여하는 것으로 한다. 구체적으로 두 시스템 비교 평가일 경우 전체 테스트항목에 대한 번역속도가 한 시스템이 다른 시스템과 비교하

---

언어학자들과 전산언어학자들이 직접 참여하여 만들었다는 점에서 다른 여타의 상용화된 제품들 “워드체인지”, “번역마당 2.0”과 구별될 뿐만 아니라, 번역의 품질에서도 두 제품그룹은 현격한 차이를 보였다. 언어공학연구소가 내놓은 “트래니 96”은 프로그래밍에 능숙한, 서울대 언어학과의 학부와 대학원생들의 10여년에 걸친 공동작업의 결과이고, 한국 IBM이 내놓은 “앙꼬르 2.0”은 서울대 컴퓨터공학과의 자연언어처리 연구실에서 10여년 동안 한국 IBM의 연구비를 지원받아 수행된 공동연구 결과이다.

<sup>7</sup>평가기준간의 상대적인 가중치 부여는 어떠한 객관적인 근거에 바탕을 둔 것이 아니라, 계량적인 평가를 위한 하나의 방편에 지나지 않음을 여기에서 분명히 하고자 한다. 따라서 다른 목적을 위해서는 다른 상대적인 가중치 부여가 가능할 것으로 생각한다.

여, 2배 이상 4배 이하 빠를 경우에 전자에는 평점 5를, 후자에는 평점 4를 부여하고, 4배 이상 6배 이하 빠를 경우에 전자에 평점 5를, 후자에 평점 3을 부여한다. 한 시스템이 다른 시스템과 비교하여, 6배 이상 8배 이하 빠를 경우에 전자에 평점 5를, 후자에 평점 2를 부여한다. 한 시스템이 다른 시스템과 비교하여, 8배 이상 빠를 경우에 전자에 평점 5를, 후자에 평점 1을 부여한다. 세 개 이상의 시스템 비교의 경우에도 가장 빠른 시스템에 평점 5를 부여함으로써, 동일한 기준을 적용할 수 있다. 그러나 단독평가일 경우 기본적으로 주어지는 값인 5가 부여되는 것으로 가정한다.

(7) 평가기준 목록

- 절대적인 기준
  - 명제내용 (40%) 8 단위
  - 문법구조 (20%) 4 단위
  - 단어인식 (10%) 2 단위
  - 완성도 (20%) 4 단위
- 상대적인 기준
  - 번역속도 (10%) 2 단위

평가기준에 대한 이러한 제안에 따르면, 각각의 평가문항은 다음에 제시된 바와 같은 평가표를 갖는다.

(8) 평가표

문항번호	명제내용	문법구조	단어인식	완성도	번역속도

예를 들어 평가문항 18에 대한 영 한 번역시스템 앙꼬르 2.0의 번역결과는 다음의 표 (10)에 제시된 평가점수를 부여받는다.



(9) 평가문항의 예

평가문항 18: The key opened the door.

시스템(앙포르 2.0)의 번역: 열쇠는 문을 열었다.

전문가번역: 열쇠로 문을 열었다.

(10) 평가결과의 예

문항번호	명제내용	문법구조	단어인식	완성도	번역속도
18	5	4	5	4	5

위에서 예로 든 평가문항 18에 대한 번역시스템의 번역결과는 명제 내용과 단어인식에 있어 오류가 없으나, 한국어의 문법구조에 어긋나는 번역이다. 또한 번역의 결과가 자연스럽지 못하다는 점에서 완성도가 떨어지는 번역이다. 그에 따라 각각의 기준에 대해 평가 점수 4를 부여받게 된다. 평가기준 번역속도는 단독 평가이기 때문에 5를 부여받는다. 그러나 다른 시스템과의 비교 평가일 경우, 트래니 96의 번역속도가 앙포르 2.0의 번역속도보다 7.43배 정도 빠르므로 번역속도의 평점이 2가 된다.<sup>8</sup>

(11)

문항번호	명제내용	문법구조	단어인식	완성도	번역속도
18	5	4	5	4	2

종합평가의 경우에는, 각 평가기준별 평균평점과 총평점이 각 시스템에 부여된다. 아래의 표 (12)는 다음절에서 구체적으로 논의할 앙포르 2.0 시스템에 대한 종합평가표인데, 이 시스템은 종합적으로 B 라는 총 평점을 얻는다. 개별 평가기준별로 살펴보면, 명제내용, 문법구조, 단어인식, 완성도 등 모두 평점 B를 부여받는 반면, 번역속도는 평점 D를 부여받는다.

<sup>8</sup>영어 평가문장 650 개를 트래니 96으로 번역시킨 결과, 150 메가 헤르쯔, 64 MB 주메모리 사양을 가진 펜티엄 프로 PC로 약 2분 30초(156 초)가 걸렸다. 반면 동일한 문장 650개를 앙포르 2.0으로 번역시킨 결과, 약 19분 20초 걸렸다.

(12) 종합평가표

표번호	평제내용	문법구조	단어인식	완성도	번역속도	총평점
1	119	121	120	114	50	
2	94	108	96	88	50	
3	106	105	112	100	50	
합계	329	334	328	302	150	
평균	4.25	4.47	4.37	4.03	2	
가중치	8	4	2	4	2	
총합계	34	27.9	8.7	26.2	4	80.7
총평점	B	B	B	B	D	B

여기에서 개별 평가기준별 총평점은 다음의 <총평점체계>는 평균을 기준으로 삼아 부여한 것이다.

(13) 총평점 체계

총평점	평균
A	4.50 5.00
B	3.50 4.49
C	2.50 3.49
D	1.50 2.49
E	0.50 1.49
F	0.49 이하

한편, 종합평점은 총합계 점수를 기준으로 하여 부여되며 A, B, C, D, E, F 등 6 등급으로 되어, 총합계 점수와 종합평점의 관계는 다음의 표에 제시된 바와 같다.

(14) 종합평점 체계

종합평점	총합계
A	90.0 100.0
B	80.0 89.9
C	70.0 79.9
D	60.0 69.9
E	50.0 59.9
F	49.9 이하

## 4.2. 시스템 앙포르 2.0의 종합 평가

이 절에서는 시스템 앙포르 2.0의 번역성능에 대한 종합적인 평가의 결과를 제시한다. 여기서는 전체 평가예문 650개 전체의 번역결과를 대상으로 하지 않고 문항번호 1 50, 문항번호 625 650 등 75개 예문을 대상으로 제한적인 평가를 실시하였다. 그 각각의 평가 결과는 다음의 <부분 평가표 1> 부터 <부분 평가표 3> 과 <종합 평가표>에 정리되어 있다.

(15) 부분 평가표 1

문항번호	명제내용	문법구조	단어인식	완성도	번역속도
1	5	5	5	4	2
2	5	5	4	5	2
3	5	5	4	5	2
4	5	5	5	4	2
5	5	5	5	4	2
6	5	5	5	4	2
7	5	5	5	5	2
8	5	5	5	5	2
9	5	5	5	5	2
10	5	5	5	5	2
11	5	5	5	5	2
12	5	5	5	5	2
13	5	5	5	5	2
14	5	5	5	5	2
15	3	4	5	4	2
16	5	5	5	5	2
17	5	5	5	5	2
18	5	4	5	4	2
19	5	5	5	5	2
20	4	5	4	4	2
21	4	5	4	3	2
22	3	3	4	4	2
23	5	5	5	4	2
24	5	5	5	5	2
25	5	5	5	5	2
합계	119	121	120	114	50

(16) 부분 평가표 2

문항번호	명제내용	문법구조	단어인식	완성도	번역속도
26	4	5	4	3	2
27	4	5	4	3	2
28	5	5	5	4	2
29	5	5	4	4	2
30	4	5	5	4	2
31	5	5	5	5	2
32	5	5	5	5	2
33	4	4	4	4	2
34	4	5	4	4	2
35	0	0	0	0	2
36	2	5	2	2	2
37	5	5	5	4	2
38	2	5	2	2	2
39	5	5	5	5	2
40	5	5	5	5	2
41	5	5	5	5	2
42	5	5	5	5	2
43	5	5	5	5	2
44	3	5	5	3	2
45	4	3	4	3	2
46	3	3	4	3	2
47	3	4	2	3	2
48	3	4	2	3	2
49	0	0	1	0	2
50	4	5	4	4	2
합계	94	108	96	88	50

(17) 부분 평가표 3

문항번호	명제내용	문법구조	단어인식	완성도	번역속도
626	5	5	5	5	2
627	5	5	5	5	2
628	5	5	5	4	2
629	5	5	5	5	2
630	5	5	5	4	2

631	5	5	4	4	2
632	3	3	5	3	2
633	2	2	5	2	2
634	3	2	3	3	2
635	5	5	5	5	2
636	5	5	5	5	2
637	5	5	5	5	2
638	5	5	5	4	2
639	5	5	5	5	2
640	5	5	5	5	2
641	5	5	5	5	2
642	0	0	0	0	2
643	5	5	5	4	2
644	5	5	5	4	2
645	5	5	5	5	2
646	5	5	5	5	2
647	0	0	0	0	2
648	3	3	5	4	2
649	5	5	5	5	2
650	5	5	5	4	2
합계	106	105	112	100	50

(18) 종합평가표

표번호	명제내용	문법구조	단어인식	완성도	번역속도	총평점
1	119	121	120	114	50	
2	94	108	96	88	50	
3	106	105	112	100	50	
합계	329	334	328	302	150	
평균	4.25	4.47	4.37	4.02	2	
가중치	8	4	2	4	2	
총합계	34	17.9	8.7	16.08	4	80.7
총평점	B	B	B	B	D	B

앞절에서 설명한 바와 같이 이 표에 따르면, 이 시스템은 종합적으로 B 라는 총 평점을 얻고 있다. 그리고 개별 평가기준별로는 명제 내용, 문법구조, 단어인식, 완성도 등 모두 평점 B 를 부여받는 반

면, 번역속도는 평점 D를 부여받는다.

### 4.3. 시스템 트레이니 96의 종합 평가

이 절에서는 시스템 트레이니 96의 번역성능에 대한 종합적인 평가의 결과를 제시한다. 여기서는 전체 평가예문 650개 전체의 번역결과를 대상으로 하지 않고 문항번호 1 50, 문항번호 625 650 등 75개 예문을 대상으로 제한적인 평가를 실시하였다. 그 각각의 평가결과는 다음의 <부분 평가표 1> 부터 <부분 평가표 3> 과 <종합평가표>에 정리되어 있다.

(19) 부분 평가표 1

문항번호	명제내용	문법구조	단어인식	완성도	번역속도
1	5	5	5	4	5
2	4	5	4	4	5
3	4	5	4	4	5
4	4	5	5	4	5
5	4	5	5	4	5
6	4	5	5	4	5
7	4	5	4	4	5
8	5	5	5	5	5
9	5	5	5	5	5
10	5	5	5	5	5
11	5	5	5	5	5
12	5	5	5	5	5
13	5	5	5	5	5
14	5	5	5	5	5
15	3	3	5	3	5
16	3	5	4	3	5
17	0	0	0	0	5
18	5	4	5	4	5
19	3	3	5	4	5
20	3	3	5	4	5

21	3	3	5	4	5
22	2	3	4	2	5
23	4	4	5	4	5
24	4	5	5	4	5
25	5	5	5	5	5
합계	99	108	115	100	125

(20) 부분 평가표 2

문항번호	명제내용	문법구조	단어인식	완성도	번역속도
26	4	5	5	4	5
27	4	5	5	4	5
28	2	2	3	2	5
29	0	0	0	0	5
30	3	4	4	4	5
31	3	4	3	4	5
32	2	3	3	2	5
33	5	5	5	4	5
34	1	0	1	1	5
35	1	0	1	1	5
36	5	4	5	3	5
37	5	5	5	5	5
38	0	0	0	0	5
39	5	5	5	5	5
40	5	5	5	5	5
41	4	5	5	4	5
42	5	5	5	5	5
43	5	5	5	5	5
44	0	0	0	0	5
45	5	5	5	2	5
46	2	3	5	2	5
47	3	3	5	3	5
48	3	3	5	3	5
49	3	3	4	2	5
50	3	3	4	3	5
합계	78	82	93	73	125

(21) 부분 평가표 3

문항번호	명제내용	문법구조	단어인식	완성도	번역속도
626	5	5	5	5	5
627	5	5	5	5	5
628	5	3	5	4	5
629	0	0	3	0	5
630	2	5	3	1	5
631	5	5	5	4	5
632	2	5	3	1	5
633	2	5	2	1	5
634	3	3	3	3	5
635	5	5	5	5	5
636	5	5	5	4	5
637	5	5	5	4	5
638	2	2	5	3	5
639	5	5	5	5	5
640	5	5	5	5	5
641	5	5	5	5	5
642	5	5	5	5	5
643	0	0	0	0	5
644	5	5	5	4	5
645	5	5	5	3	5
646	5	5	5	5	5
647	5	5	5	5	5
648	5	5	3	5	5
649	3	3	4	3	5
650	5	5	5	4	5
합계	99	106	106	89	125



(22) 종합평가표

표번호	명제내용	문법구조	단어인식	완성도	번역속도	총평점
1	99	108	115	100	125	
2	78	82	93	73	125	
3	99	106	106	89	125	
합계	276	296	324	262	375	
평균	3.68	3.95	4.29	3.49	5.00	
가중치	8	4	2	4	2	
총합계	29.4	15.8	8.3	14.0	10.0	77.5
총평점	B	B	B	C	A	C

이 표에 따르면, 이 시스템은 종합적으로 C 라는 총 평점을 얻고 있다. 그리고 개별 평가기준별로는 명제내용, 문법구조, 단어인식 등 모두 평점 B를 부여받고 평가기준 완성도는 평점 C를 부여받는 반면, 번역속도에서 평점 A를 부여받는다.

#### 4.4. 번역오류의 유형과 분석

이 절에서는 영 한 기계번역 시스템이 야기하는 번역오류의 유형들을 분석한다. 트래니 96과 앙포르 2.0 모두 전체적으로 다음의 표에 제시된 바와 같이 20% 이상의 번역오류율을 보인다.

(23) 번역오류 발생률

	트래니 96	앙포르 2.0
총평가문항수	650	650
번역오류 수	266	131
오류발생비율	40.9%	20.1%

번역오류의 유형으로는 크게 문법구조 오류와 어휘층위의 오류로 구분했다. 아래의 표에 제시된 바와 같이 문법구조 중에서는 부정사구, 비교구문, 복합문, 관계문, 강조구문, 부사 포함문 등이 번역오류를 가장 많이 유발하는 구조들이었다. 외에 트래니 96의 경우에

문장패턴의 인식을 위해 구성된 문장들 (문장번호 563 600)의 번역에 있어 아주 많은 오류를 발생시킨 반면 (20/38 52.6 %), 앙포르 2.0 의 경우, 비교적 적은 오류를 냈다 (3/38 7.8). 이 결과를 두 시스템의 전체적인 오류 발생률과 비교해 보면, 기계번역 시스템의 구축 시에 문장패턴 인식 성능을 높일 수 있도록 노력하는 것이 시스템의 전체 성능을 향상시키는 방법중의 하나라고 판단된다.

(24) 영어의 문법구조에 따른 번역오류의 양상

	부정사구	비교구문	복합문	관계문	강조구문
트래니 96	24	20	15	6	7
앙포르 2.0	19	15	9	4	5

	상관사구	수동문	진행형	최상급	부사
트래니 96	3	6	4	6	12
앙포르 2.0	6	1	5	4	4

	복합명사구	수량표현	관용표현	문장패턴	언어층위
트래니 96	5	3	12	20	19
앙포르 2.0	4	1	9	3	9

다음의 예들은 실제 번역결과로 얻어진 번역오류의 예들이다. 아래의 예에서 < ◆ ... >는 영어원문을 지칭하고, < # ... >는 한국어 대역문장이며, < T:... >는 트래니 96의 번역결과를 나타내고, < E: ... >는 앙포르 2.0의 번역결과를 나타낸다.

(25) 부정사 구문

- ◆ It was a sensible precaution to take.
- # 그것은 알맞은 예방조처였다.
- T: 잡는 것이 합리적인 예방조치이었다.
- E: 잡는 것은 분별이 있는 조심이었습니다.

(26) 비교구문

◆ I am taller than he is.

# 나는 그보다 키가 크다.

T: 나는 그가 있는 것 보다 더 크다.

E: 그가 있는 것보다, 나는 더 큼니다.

(27) 복합문

◆ You have a lot of free time but I haven't much.

# 당신은 자유로운 시간이 많지만 나는 많지 않습니다.

T: 당신은 가지고 있는 자유로운 시간.

E: 집에서 공부하는 것은 그것이 불가능했다는 것을 발견했습니다.

(28) 관계문

◆ I was invited to a dinner given to welcome the new ambassador.

# 나는 새로 온 대사를 환영하기 위해 마련된 저녁식사에 초대 받았다.

T: 나는 새로운 대사를 환영하는 것을 주어 진 만찬에 초청되었다.

E: 나는 새로운 대사를 환영하도록 주어져서 정찬에 초대되었습니다.

(29) 강조 구문

◆ There was little time for consultation.

# 상담할 시간이 거의 없다.

T: [번역 결과 출력 안함]

E: 상의를 위한 작은 시간은 있었습니다.

(30) 상관사 구문

◆ He thought it was safer to go.

# 그는 가는 것이 더 안전하다고 생각했다.

T: [번역 결과 출력 안 함]

E: 그는 그것이 가기에 더 안전했다고 생각했습니다.

(31) 수동

◆ He was much admired.

# 그는 존경을 많이 받았다.

T: 그는 많이 존경되었다.

E: 그는 많이 감탄 받았습니다.

(32) 진행형

◆ She is coming today.

# 그녀가 오늘 온다.

T: 오늘에 그녀는 오고있다.

E: 그녀는 오늘 오고 있습니다.

(33) 최상급

◆ The youngest of the family was the most successful.

# 그 가족 중에서 가장 나이가 적은 사람이 가장 성공하였다.

T: 가족의 가장 어린 것이 가장 성공적이던 것.

E: 가족 중에서 가장 젊은 것은 성공적인 대부분이었습니다.

(34) 부사가 포함된 구문

◆ He looked calmly at the angry crowd.

# 그는 조용히 성난 군중을 바라보았다.

T: 그는 화난 군중에(서) calmly 보였다.

E: 그는 화나는 군중을 조용히 보았습니다.

◆ They stay up all night sometimes.  
 # 가끔 그들은 밤을 새운다.  
 T: 그들은 위로 밤새도록 때때로 머무른다.  
 E: 그들은 밤새도록 sometimes 머무릅니다.

(35) 복합 명사구

◆ Only a few of our customers have accounts.  
 # 우리 고객 중에 단지 몇 명만이 계좌를 갖고 있다.  
 T: only a few 우리의 손님에 예금계좌를 가지고 있어라.  
 E: 단지 우리들의 단골 손님의 몇몇은 계산을 가지고 있습니다.

(36) 수량 표현

◆ Few towns have such splendid trees.  
 # 그렇게 훌륭한 나무들이 있는 도시는 거의 없다.  
 T: 거의 도시는 그런 훌륭한 나무를 가지고 있지 않다.  
 E: 약간의 마을이 그러한 훌륭한 나무를 가지고 있습니다.

(37) 관용표현

◆ He had a bad car crash.  
 # 그는 심한 자동차 충돌을 겪었다.  
 T: 그는 나쁜 차 충돌을 가지고 있었다.  
 E: 그는 나쁜 차 충돌을 가지고 있었습니다.

한편, 다음의 표 (38)은 두 시스템이 어휘층위에서 생성한 오류의 빈도와 유형을 보여준다.

(38) 어휘 층위의 오류

	어휘 중의성	미등록어휘
트래니 96	46	19
앙포르 2.0	33	7

어휘충위의 오류를 어휘가 가지는 중의성에 기인한 오류와, 어휘가 번역시스템의 사전에 등재되어 있지 않음으로써 야기되는 오류로 구분할 때, 두 경우 모두 번역시스템 트래니 96이 보다 많은 오류를 보이는 것으로 평가되었다. 아래의 예들은 이러한 오류의 구체적인 예들이다.

(39) 어휘중의성

◆ He was fairly relaxed.

# 그는 꽤 긴장이 풀렸다.

T: 그는 공정하게 느슨한었다.

E: 그는 공정하게 늦추어졌습니다.

(40) 미등록어휘

◆ They decided to leave the town secretly.

# 그들은 그 마을을 몰래 떠나기로 결정하였다.

T: 그들은 그 도시가 secretly한 것을 떠나기로 결정했다.

E: 그들은 마을을 비밀히 남겨두기로 결정했습니다. (어휘인식 성공)

## 5. 결론

본 연구에서는 국내에서 아직 본격적으로 시도된 바 없는, 영한 기계번역 시스템의 계량적인 평가방안을 정립하였다. 이를 위해 먼저 평가용 문장 650 개로 구성되어 있는 영어 측정장치를 문법현상을 중심으로 구축하고, 각 문장에 대해 전문인에 의한 한국어 번역을 첨부하였다. 개별 평가문장에 대해, 전문인에 의한 한국어 번역과 기계번역시스템의 번역결과를 비교함으로써 기계번역시스템의 번역성능을 평가하는, 소위 유형적인 번역평가방안의 하나인 유형적 비교평가방법을 제안하였다. 개별 평가문장의 번역의 질을 평가하기 위해 절대적인 평가기준으로 명제내용, 문법구조, 단어인식 및

완성도 등을 설정하고, 상대적인 평가기준으로 번역속도를 설정하였다. 이들 각 평가기준들에 대해서 4:2:1:2:1의 가중치 차이를 두어 평가의 신뢰도를 높였다. 그런데, 이상에서 기술한 바, 본 연구의 중요한 연구결과인 유형적 비교평가방법 자체를 자동화하는 방안이 앞으로 마련되어야 할 것으로 보인다.

이 연구에서는 이 밖에도 한 영 기계번역 시스템 평가에 있어 기초자료가 될 650 문장규모의 한국어 측정장치를 구축하고, 여기에 속한 각 평가문장에 대해서 전문인에 의한 영어번역을 첨부하였다.

## 참고문헌

- 성원경, 장명길 외. 1997. "SERI Test Suites '97", 제9회 한글 및 한국어정보처리 학술대회논문집, 한국정보과학회, 한국인지과학회.
- 장석진. 1998. "번역: 대응과 평가", 언어와 정보 2, 1 41.
- Albisser, D. 1993. "Evaluation of MT Systems at Union Bank of Switzerland", *Machine Translation* 8, 25 27.
- AMTA. 1992. *MT evaluation: basis for future directions*, Washington, D.C. Association for Machine in the Americas.
- Arnold, D., Sadler, L. and Humprheys, R. L. 1993. "Evaluation: An Assessment", *Machine Translation* 8, 1 24.
- Balkan, L., Meijer, S. et al. 1994. "Test Suite Design: Guidelines and Methodology", *D WP2.1a*, University of Essex.
- Balkan, L., Fouvry, F. and Regnier Prost, S. eds. 1996. *TSNLP User Manual*, vol. 1 vol. 3. University of Essex.
- Estival, L., K. Falkedal et al. 1994. "Analysis of Existing Test Suites", *D WP1*. University of Essex.

Jones, K. S., Galliers, J. R. 1996. *Evaluating Natural Language Processing Systems: An Analysis and Review*. Berlin: Springer.

Hornby, 1976. *Guide to Patterns and Usage in English*. 서울 : 범문사.



## [부록] 평가문장 (626-650)과 기계번역 결과<sup>9</sup>

< ◆ ... >는 영어원문, < # ... >는 한국어 대역문장.

< T:... >는 트래니 96의 번역결과,

< E: ... >는 앙포르 2.0의 번역결과

626.

◆ He can see him.

# 그는 그를 볼 수 있다.

T: 그는 그를 볼 수 있다.

E: 그는 그를 볼 수 있습니다.

627.

◆ It is raining.

# 비가 오고 있다.

T: 비오고있다.

E: 비가 오고 있습니다.

628.

◆ It is requested that she should leave.

# 그녀가 떠나도록 요청받았다.

T: 그녀가 떠나야 할 것은 요청된다.

E: 그녀가 떠난 것은 요구됩니다.

629.

◆ Can she not carry it?

# 그녀가 그것을 운반할 수 없을까요?

T: 할 수 있어라 그녀가 그것을 지니지는 않다?

E: 그녀는 그것을 운반할 수 없을까요?

630.

◆ The committee, that she accepts, comes.

# 그녀가 받아들인 위원회가 온다.

T: 위원회 , 그녀가 받아들이다, 온다.

---

<sup>9</sup>한 심사위원의 의견에 따라, 평가의 공정성을 기하기 위한 목적으로 평가용 문장의 일부와 두 시스템에 의한 기계번역의 결과를 부록으로 첨부한다.

E: 그녀가 받아들이는 위원회는 옵니다.

631.

◆ A lot of them are new.

# 그들 중 많은 사람이 새로운 사람이다.

T: 많은 그들은 새롭다.

E: 많은 그들은 새롭습니다.

632.

◆ The committee, who she accepts, come.

# 그녀가 받아들인 위원들이 온다.

T: 위원회, 그녀는 누구를 받아들이다, 온다.

E: 그녀가 와 저서 받아들이는 위원회

633.

◆ The committee, at whom he aims, succeed.

# 그녀가 표적으로 삼고 있는 위원들이 성공한다.

T: 위원회, 의 그는 누구를 맞춘다, 성공한다.

E: 위원회, 그가 누구를 겨누는 지에 성공하시오.

634.

◆ What does he look at?

# 그는 무엇을 보고 있는가?

T: 그가 무엇처럼 보이는가?

E: 그는 무엇에서 볼까요?

635.

◆ Whom does she see?

# 그녀는 누구를 보고 있는가?

T: 그녀가 누구를 보는가?

E: 그녀는 누구를 볼까요?

636.

◆ You succeed.

# 너는 성공한다.

T: 당신은 성공한다.

E: 당신은 성공합니다.

637.

- ◆ She calls her up.
- # 그녀는 그녀에게 전화를 건다.
- T: 그녀는 그녀를 부른다.
- E: 그녀는 그녀에게 전화합니다.

638.

- ◆ He asks him how she works.
- # 그는 그녀가 어떻게 일하는 지를 그에게 묻는다.
- T: 그는 어떻게 그녀가 일하는 그에게 묻는다.
- E: 그는 그에게 그녀가 어떻게 일을 하는 지를 질문합니다.

639.

- ◆ They succeed.
- # 그들은 성공한다.
- T: 그들은 성공한다.
- E: 그들은 성공합니다.

640.

- ◆ She is compared with the manager.
- # 그녀는 매니저와 비교되고 있다.
- T: 그녀 매니저와 비교되었다.
- E: 그녀는 경영자와 비교됩니다.

641.

- ◆ He comes.
- # 그가 온다.
- T: 그는 온다.
- E: 그는 옵니다.

642.

- ◆ She puts it by.
- # 그녀는 그것을 남겨둔다.  
(문맥에 따라) 그녀는 그것을 옆에 놓는다.)
- T: 그녀는 그것을 놓는다.
- E: [번역결과 출력안함]

643.

- ◆ Both he and the manager are new.
- # 그와 과장은 새로운 사람들이다.
- T: [번역결과 출력 안 함]
- E: 그와 경영자는 새롭습니다.

644.

- ◆ Either she or they are new.
- # 그녀나 그들은 새로운 사람들이다.
- T: 그녀나 혹은 그들은 새롭다.
- E: 그녀나 그들은 새롭습니다.

645.

- ◆ She explains to him that it works.
- # 그녀는 그것이 효과적이라고 그에게 설명한다.  
((문맥에 따라) 그녀는 그것이 잘 작동한다고 그에게 설명한다.
- T: 그녀는 그에 그것이 일하는 것을 설명한다.
- E: 그녀는 그에게 그것이 작동한다고 설명합니다.

646.

- ◆ He depends on the manager.
- # 그는 매니저에게 의존한다.
- T: 그는 매니저에 의존한다.
- E: 그는 경영자에 의지합니다.

647.

- ◆ The committee succeed.
- # 위원회는 성공한다.
- T: 위원회는 성공한다.
- E: 위원회는 성공합니다.

648.

- ◆ How ill is he?
- # 그가 얼마나 아픈가?
- T: 어떻게 악이 그이다?
- E: 그는 얼마나 나쁘게 있습니까?

649.

◆ How late does she come?

# 그녀가 얼마나 늦게 오느냐?

T: 그녀가 너무나 늦게 오는가?

E: 그녀는 얼마나 늦게 올까요?

650.

◆ Either she or they are new.

# 그녀나 그들은 새로운 사람들이다.

T: 그녀나 혹은 그들은 새롭다.

E: 그녀나 그들은 새롭습니다.

이민행, 지광신

서울시 서대문구 신촌동 134

연세대학교 문과대 독어독문학과

120-749

E-mail: {leemh, kwangsin}@bubble.yonsei.ac.kr

정소우

서울시 성북구 동선동 3가 249-1

성신여자대학교 인문과학대 영어영문학과

136-742

E-mail: swchung@cc.sungshin.ac.kr

계재일자 : 1998년 9월 15일

접수일자 : 1998년 11월 30일