

# 연관성규칙을 이용한 기계번역프로그램의 유용성 평가와 지속 사용 의향의 관계 분석

## Machine Translation Program's Usefulness and Continuous Use Intention by the Analysis of the Association Rule

안 효 준\* Ahn, Hyo Jun

조 성 빈\*\* Cho, Sungbin

### Abstract

This study attempts to analyze the causal relationship of usefulness and continuous use intention on machine translation program. Usefulness is measured in terms of content equivalence, faithfulness, intonation relevance, paragraph separation, connection smoothness, appropriate word use, logic, context retention, naturalness, and the delivery of meaning and information. Continuous use intention is measured in terms of overall satisfaction, continuous use intention, recommendation intention, substitute program search, and paid program search. The derived association rules indicate that satisfaction and continuous use with the currently using program are related to the positive evaluation on connection smoothness, appropriate word use, logic, and naturalness. On the other hand, when users are satisfied with appropriate word use and the delivery of the intended meaning and information, they tend to perceive less need on searching substitute and even higher level, paid translation programs.

**Keywords:** machine translation program, usefulness, continuous use intention, association rule

---

\* 서강대학교 일반대학원 경영학과

\*\* 교신저자, 서강대학교 경영학과 교수, 이메일: [sungbincho@sogang.ac.kr](mailto:sungbincho@sogang.ac.kr)

투고일 2017.11.17

수정일 2017.12.19

게재일 2017.12.31

## 1. 서 론

지난 수 십 년 간 컴퓨터와 정보통신 기술의 발달과 함께 사람들은 전 세계의 인터넷으로부터 지식과 정보를 습득하며 다양한 언어를 접할 기회가 증가하고 있다. 자국어와 영어 이외의 언어로 표현된 지식과 정보를 자국어와 영어로 번역할 필요성이 증대되었고 이는 컴퓨터의 성능 개선과 인공지능의 발달에 따라 여러 포털 사이트에서 기계번역 프로그램을 제공하고 있으며 모바일을 통한 실시간 번역 프로그램까지 등장하고 있다.

대표적인 번역 프로그램 제공 사이트로 많은 사람들이 인식하고 사용하고 있는 구글 번역 프로그램은 2016년 11월 기준으로 103개의 언어를 번역할 수 있고 일일 번역 단어수는 1,400억 개에 달하고 있다. 초중고 대학 등 학생들은 학업을 위한 자료의 검색과 수집 뿐 아니라 개인적 취미와 관심사에 관련된 방대한 정보를 인터넷을 통하여 접하고 한국어나 영어로 번역하여 사용하고 있으며, 일반 사회인들은 실시간 외국 뉴스, 해외 쇼핑, 동영상 자막 번역, 이메일 송수신 시의 번역 등 다양한 목적으로 기계번역 프로그램을 사용하고 있다. 구글 번역기 뿐 아니라, 네이버 번역기 파파고, 한컴 지니톡 등 다양한 프로그램이 개발되고 있어 사용자들은 점점 선택의 폭이 넓어지고 있다. 2017년 2월 세종대학교에서 개최된 인간과 인공지능의 번역 대회에서는 번역 품질 측면에서는 인간이 나았고 번역 속도 측면에서는 인공지능이 앞선 바 있다. 다양한 언어의 번역이란 측면에서 보면 인간의 능력은 매우 제한적이지만 기계번역기의 능력은 월등히 앞서고 있고 향후에는 더욱 앞설 것으로 전망되고 있다.

본 연구에서는 우리나라에서 가장 수요가 많은 영어 - 한국어 번역을 대상으로 기계번역 프로

그램의 번역문에 대하여 다양한 차원에서 유용성을 평가하고 그에 따른 만족도와 지속 사용 의향 등에 어떠한 영향을 미치고 있는지 조사하고자 하였다. 분석 모형으로는 데이터 마이닝 기법의 하나인 연관성 규칙(association rule induction)을 적용하여 재사용과 추천 등에 긍정적 혹은 부정적 영향을 주는 요인들을 찾아내고 그 효과를 분석하고 관리적 시사점을 도출하고자 시도하였다.

## 2. 선행연구

### 2.1 기계번역 프로그램

기계번역 프로그램은 크게 세 가지 기술에 기반을 두고 있다. 기반 기술에는 규칙 기반 기계번역(rule-based machine translation: RBMT), 통계 기반 기계번역(statistical machine translation: SMT), 그리고 신경망 기계번역(neural machine translation: NMT)이 있다. 첫째, 규칙 기반 기계번역은 문장을 구성하는 문법적 요소와 많은 양의 어휘를 포함하는 사전적 지식에 기반을 두고 있다. 이러한 문법과 사전 및 규칙들은 다양한 언어별 언어 전문가에 의하여 만들어지고 있다(Bennett and Slocum, 1985). 규칙 기반 번역의 가장 큰 장점은 정의된 규칙에 따라 정확도 높은 번역문을 산출해내는 것이라고 할 수 있다. 반면에 단점으로는 새로운 도메인에 대한 적응력이 떨어지며, 인간의 추가적인 노력에 따른 비용이 지속적으로 발생하는 점을 들 수 있다(Lagarda et al., 2009).

둘째, 통계 기반 기계번역은 이중으로 된 언어 말뭉치를 분석하여 구축한 통계적 모델에 기반을 두고 번역을 수행한다(Brown et al., 1993). 기본적으로 계량 모델을 기반으로 하고 있으므로 인적 비용이 적게 든다는 장점이 있다(Chen and Chen,

1996). 하지만, 말뭉치를 형성하려면 비용이 많이 들고, 특히 언어에 따라 어순이 다르기 때문에 이에 따라 번역이 완벽하지 않고 어순이 바뀌어 어색한 경우가 발생한다는 약점(Okpor, 2014)이 있으며, 필연적으로 고성능의 컴퓨터가 필요하다는 제한조건을 가지고 있다.

셋째, 신경망 기반 기계번역은 인공 신경망(artificial neural network)을 훈련용 데이터와 검증용 데이터로 분리하여 훈련시킴으로써 번역의 성능을 극대화시키는 방법이라고 할 수 있다. 대표적인 알고리즘은 recurrent neural network(RNN)를 들 수 있는데, 이는 번역하고자 하는 원문을 encode하고 번역되어야 하는 언어를 decode하는 방식으로 결과물을 산출하는 방식이다(Bahdanau et al., 2014; Wolk and Marasek, 2015). 2016년 9월 소개된 구글의 구글 신경망 기계번역(Google neural machine translate: GNMT)은 wikipedia news의 샘플을 주요 언어들로 번역하여 기존의 구문 기계번역(phrase-based machine translation: PBMT)보다 오류가 60% 정도 감소한다고 발표하였다(Wu et al., 2016). Johnson et al.(2016)에 따르면 구글이 2017년 11월에 소개한 Zero-shot translation은 단일 시스템에서 다중 언어를 번역할 수 있으며 특히 시스템에서 지원되지 않는 언어까지도 번역이 가능하다고 한다.

## 2.2 기계번역 성능 평가

기계번역 프로그램의 성능에 대한 평가는 번역기의 성능과 번역 결과물의 정확도를 기준으로 하며 부족한 부분에 대한 지적과 개선 방향에 대한 제시를 하고 있다(최기선, 2001). 평가는 크게 자동 평가와 수동 평가로 나눌 수 있다.

자동 평가에 관련하여 Banerjee and Lavie(2005)은 상관성, 민감도, 일관성, 신뢰성과 보편성을 자동 평가 척도 속성으로 제시하였다. 이외에도 여러 차원이 평가에 사용되고 있는데, BLEU, NIST, WORD ERROR RATE, METEOR, LEPOR 등이 있다. Papineni et al.(2002)이 제시한 BLEU 지수는 인간의 번역에 근접하는 정도를 측정하였다. NIST 지수는 BLEU metric을 기반으로 하여 n-gram에 가중치를 부여하고 있다(Doddington, 2002). WORD ERROR RATE은 단어 차원에서 삽입 횟수와 삭제 횟수를 계산하고 있는데 어순이 적절히 고려되지 못한다는 단점이 있다. 이러한 약점을 극복하기 위하여 position-independent word error rate가 제시되기도 하였다(Tillmann et al., 1997). METEOR 지수는 unigram의 정밀도와 unigram recall의 가중 평균을 기반으로 하여 BLEU의 결함을 극복하기 위하여 설계되었다(Banerjee and Lavie, 2005). LEPOR 지수는 언어 편향 현상을 감소시키기 위하여 length penalty, N-gram position difference penalty, precision and recall을 고려하여 설계되었으며, Han et al.(2013)은 좀 더 개선된 hLEPOR를 소개하였다.

수동 평가는 적절성(adequacy)과 유창성(fluency)을 주요 평가 기준으로 사용하는데, 사람의 참여가 필수적이기 때문에 시간과 비용이 크게 발생한다. 평가에는 주로, 예를 들어 리커트 5점 척도가 사용된다(Koehn and Monz, 2006). 자동 평가에 비하여 평가자의 주관에 의하여 결과가 크게 영향을 받기 때문에 다수의 전문가 집단의 참여가 요구된다. Multidimensional quality metrics(MQM)은 유창성(fluency), 정확성(Accuracy), 진실성(verity), 디자인(design), 국제화(Internationalization) 차원이 사용되고 있으며 인간에 의한 번역과 기계 번역에 공히 적용되는 척도이다(Lommel et al., 2013). Zaretskaya et al.(2016)은 MQM 방법이 정확하고

세분화된 평가가 가능하지만 평가자가 다양한 오류를 범할 가능성이 있고 평가에 장시간이 걸린다는 단점을 지적하였다. 국내에서 번역 평가와 관련된 연구에서는 다음과 같은 기준이 제시되었다. 이향(2010)은 주관적/직관적 접근, 독자 반응 중심 접근, 언어학적 접근, 기능주의적 접근을 적용하여 번역 결과에 대한 평가를 실시하였다. 김정우(2004)는 학생에 의하여 작성된 번역문을 독해 측면과 표현 측면에서 평가하였고, 김혜림과 강려영(2015)은 한국의 공공기관 웹사이트 번역에 대한 오류 구분에 있어 텍스트 측면, 언어적 측면, 화용론적 측면, 웹 사용성 측면을 반영하였다.

국내에서는 2000년 기계번역과 관련하여 대용량 음성/언어/영상 DB 구축 및 표준화 1차 사업 발표회에서 영-한 기계번역 평가방법론에 관하여 제시하였다. 여기에서 개발자 평가에는 일본의 JEDIA를 제시하고, 일반 이용자 평가에는 ARPA의 유창성 평가 방법을 제시하였다. 박경리 등(2013)은 영한 번역에서 중간 단어를 사용하여 번역의 효율성을 높이려 시도하였으며, 임순정과 한미선(2014)의 연구는 스마트폰 제품 설명서를 영어-한국어와 프랑스어-한국어로 번역된 문장을 번역 단위별로 분석하였다. 김종희(2016)는 여러 포털 사이트에서 제공하는 번역기의 결과물 분석을 통하여 일반적인 설명과 정보 텍스트에 대하여 높지 않은 수준에서 요구되는 번역 수요는 충족될 수 있다고 조사하였다. 최효은과 이지은(2017)은 한국어-영어 특허 기계 번역 시스템인 K2E-PAT를 사용하여 특허 관련 공보 초록문의 번역에 자동 평가와 수동 평가를 실시하였다.

선행 연구를 토대로 파악한 번역 평가의 구성 차원에는 텍스트의 완결성, 목적 충실성, 환경 요인의 반영, 언어 규범 등을 들 수 있다. 텍스트의 완결성 차원에는 글 전체의 자연스러움, 언어적 결속성, 정보 전달의 정확성 등이 포함되며, 목적

충실성 차원에는 원저자의 의도와 독자에 대한 효과가 포함되고, 환경 요인의 반영 차원에는 목표 독자에 대한 이해, 원저자에 대한 이해, 번역물의 용도에 대한 이해 등이 포함된다. 언어 규범 차원에는 원천어의 언어간접 배제와 목표언어 규범 준수 등이 포함된다.

### 3. 연구 모형과 분석 결과

언어번역 프로그램의 유용성 평가와 그에 따른 지속 사용 의향에 대한 연구를 위하여 본 연구에서는 유용성 평가에 10개의 항목, 지속 사용 의향에는 5개의 항목을 선정하였다. 유용성과 지속 사용 의향의 인과 관계를 규명하기 위하여 데이터 마이닝 기법 중의 하나인 연관성 규칙(Association Rule)을 적용하였고, 하나의 차원을 여러 항목으로 측정하지 않았으므로 변수들의 신뢰성과 타당성 검증은 필요로 하지 않았다. <표 1>은 언어번역 프로그램의 평가에 관한 10개 항목을 기술하고 있고, <표 2>는 지속 사용 의향에 관한 5개 항목을 기술하고 있다.

<그림 1>은 연구 모형으로서 유용성 평가와 지속 사용 의향과의 인과관계를 살펴보고자 하였다. 예비 설문을 실시하여 수정 과정을 거친 후 2017년 7월 한 달 간 학생과 직장인을 대상으로 오프라인 설문 160부와 온라인 설문 91부를 취합하였고 충실히 답변된 242부의 설문지를 분석에 사용하였다.

응답자의 구성은 다음과 같다. 성별은 남성 145명(59.9%)과 여성 97명(40.1%)으로 구성되었고, 연령대는 20세 이하 1명(0.4%), 20세~30세 201명(83.1%), 30~40세 39명 (16.1%), 40세~50세 1명(0.4%)으로서 주로 청장년층이 언어번역기를 사용하고 있음을 확인할 수 있었다. 학력은 4년제 대학교 재학 또는 졸업 182명(75.2%), 대학원 이상

<표 1> 언어번역 프로그램의 유용성 평가 항목

|     |            |                                   |
|-----|------------|-----------------------------------|
| A1  | 내용 동등성     | 원문과 번역문의 내용이 동등하다.                |
| A2  | 충실성        | 번역문이 충실하게 쓰여졌다.                   |
| A3  | 억양 적절성     | 번역문의 억양이 적절하다.                    |
| A4  | 단락 구분      | 번역문의 단락 구분이 잘 되었다.                |
| A5  | 연결 매끄러움    | 문장 연결이 매끄럽다.                      |
| A6  | 단어 적재적소 사용 | 번역문의 단어가 적재적소에 사용되었다.             |
| A7  | 논리성        | 번역문은 논리적이다.                       |
| A8  | 맥락 유지      | 번역문은 맥락을 유지하고 있다.                 |
| A9  | 자연스러움      | 번역문은 자연스럽다.                       |
| A10 | 의미와 정보 전달  | 번역문은 원문이 담고 있는 의미와 정보를 충실히 전달하였다. |

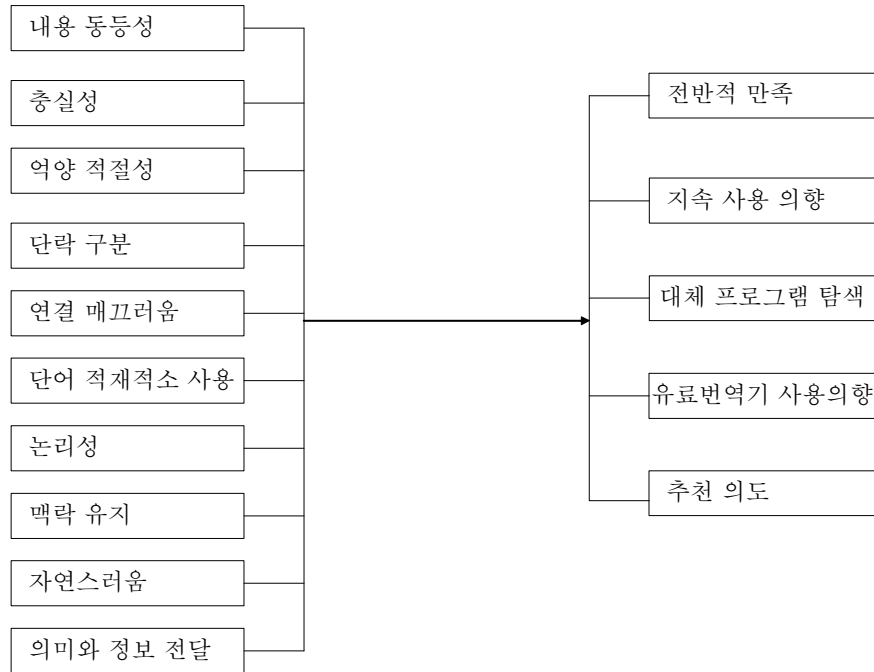
<표 2> 언어번역 프로그램의 지속 사용 의향 평가 항목

|    |            |  |
|----|------------|--|
| B1 | 전반적 만족     | 나는 번역프로그램에 전반적으로 만족한다.                         |
| B2 | 지속 사용 의향   | 나는 기존에 사용하고 있는 번역프로그램을 앞으로도 계속 사용할 의향이 있다.     |
| B3 | 대체 프로그램 탐색 | 나는 좀 더 나은 번역프로그램을 찾아보겠다.                       |
| B4 | 유료번역기 사용의향 | 나는 무료번역기가 아닌 유료번역기를 써서 좀 더 수준 높은 번역을 할 의향이 있다. |
| B5 | 추천 의도      | 나는 주변에 내가 쓰고 있는 번역기를 추천하겠다.                    |

재학 또는 졸업 56명(23.1%), 고등학교 졸업 2명(0.8%), 전문대학교 재학 또는 졸업 2명(0.8%) 순으로 나타나 비교적 고학력자들이 언어번역기를 사용하고 있음을 알 수 있었다. 직업은 학생이 185명(76.4%)으로 다수를 차지했고, 그 외에 전문직 31명(12.8%), 관리직 18명(7.4%), 서비스직 8명(3.3%) 순으로 나타났다. 요약하면, 20대의 대학교 재학 이상의 그룹에서 언어번역기를 가장 많이 쓰는 것으로 보인다.

언어번역기 사용 경험에 관한 설문 응답은 다음과 같다. 사용 빈도에 관한 질문에는 10번 이하가 104명(43.0%), 50번 이상 57명(23.5%), 10번~20번 52명(21.5%), 20번~50번 29명(12.0%)으로 나타나 10번 이하로 아주 적게 쓰거나 50번 이상으로 아주 자주 쓰는 그룹이 주를 이루고 있었다. 사용

목적으로는 학술적 목적이 141명(58.3%)으로 가장 많고, 업무 목적이 59명(24.4%)으로 다음 순이며, 그 외 취미에 17명(7.0%)이 응답하고 기타 목적이 25명(10.3%: 여행 등)이었다. 자신의 영어 수준에 관한 자신의 평가에는 평균 수준이 109명(45.0%), 평균 수준 이상이 68명(28.1%), 매우 우수함이 29명(12.0%), 최고의 수준이 4명(1.7%)으로서 상당 수준 이상의 영어 실력을 가지고 있다고 대다수가 판단하고 있었고, 평균 수준 이하라고 답변한 응답자는 32명(13.2%)에 그쳤다. 가장 많이 사용하고 있는 기계번역 프로그램이 무엇인가에 관한 질문에는 구글 번역기가 1위로서 200명(82.6%), 그 다음은 네이버 27명(11.2%), 파파고 11명(4.6%), 플리토 2명(0.8%)순으로 나타났다.



<그림 1> 연구 모형

기계번역 프로그램을 상당 기간 이상 사용하여 오며 인지하여 왔던 유용성과 지속 사용, 추천 등에 관한 응답을 돕기 위하여, 미국 드라마 News Room season 1의 대본 중 일부를 발췌하여 예시로 제시하고 5점 척도(1 = 매우 부정, 2 = 약간 부정, 3 = 중립, 4 = 약간 긍정, 5 = 매우 긍정)로 유용성과 지속 사용 의향에 관한 답변을 하도록 하였다. 제시문은 다음과 같다.

• 영어 원문

: We stood up for what was right. We fought for moral reasons. We passed and struck down laws for moral reasons. We aged wars on poverty, not poor people. The first step in solving any problem is recognizing there is one.

• 기계번역 프로그램에 의한 번역문

: 우리는 옳은 것을 위해 일어섰다. 우리는 도덕적 이유로 싸웠습니다. 우리는 도덕적 이유로 법을 통과 시켰습니다. 우리는 빈곤층이 아닌 빈곤층에서 전쟁을 겪었습니다. 문제를 해결하는 첫 번째 단계는 문제가 있음을 인식하는 것입니다.

기계언어번역 프로그램의 평가에 관한 분석을 위하여 본 연구는 연관성 규칙을 적용하였다. 각 항목의 5점 척도에 대한 응답에서 규칙을 추출하기 위하여 응답에 긍정 답변을 한 4점(약간 긍정)과 5점(매우 긍정)을 한 그룹으로 묶고, 나머지 부정과 중립 답변을 다른 한 그룹으로 묶은 후, SAS E-Miner의 Association Rule Induction을 이용하여 결과를 도출하였다(조성빈과 하병천, 2010). 연관성 규칙은 데이터 마이닝 기법 중 대표적인 undirected decision model에 속하기 때문에 많은

수의 규칙들이 도출되었고, 본 연구의 취지에 부합하기 위하여 종속변수의 긍정 평가에 영향을 주는 규칙들을 향상도(lift) 기준으로 선별하였다.

전반적인 만족도(B1)에 대한 긍정평가에 영향을 주는 규칙들은 <표 3>에 요약되어 있다. Set size는 도출된 규칙의 적절한 설명이 가능하도록 2와 3으로 제한하였고, 각각의 set size별로 향상도가 높은 상위 5개의 규칙만을 선별하였다. 두 사건 A, B에 대한 확률과 비율을 의미하는 용어에 대한 정의는 다음과 같다.

- expected confidence =  $\Pr(B)$
- confidence =  $\Pr(B|A)$
- support =  $\Pr(A, B)$
- lift = confidence / expected confidence

경우에 비하여 전반적 만족도가 높아질 가능성이 49% 더 높은 것으로 해석할 수 있다. 응답자가 해석의 논리성이 높다고 판단하면 전반적 만족도 높게 평가할 가능성은 45% 증가된다. Set size가 3인 경우, 가장 향상도가 큰 규칙은 “만약 자연스러움과 단어의 적재적소 사용에서 모두 긍정 평가를 할 때 전반적 만족도에서도 긍정 평가를 한다”이며 그 가능성은 유용성 평가에 대한 정보가 없을 경우에 비하여 67% 향상된다고 할 수 있다. 두 번째 향상도가 큰 규칙은 “연결의 매끄러움과 억양의 적절성에서 긍정 평가를 하면 전반적 만족도에 대해서도 긍정 평가를 한다”는 것으로 분석되었다.

지속 사용 의향(B2)에 긍정적 영향을 주는 규칙들은 <표 4>에 요약되어 있다. Set size가 2인 경

<표 3> 전반적 만족도(B1)에 대한 규칙: Ai high --> B1 high

| Rule         | Lift | Expected Confidence | Confidence | Support | Count |
|--------------|------|---------------------|------------|---------|-------|
| A5 --> B1    | 1.49 | 0.44                | 0.66       | 0.09    | 23    |
| A7 --> B1    | 1.45 | 0.44                | 0.64       | 0.11    | 28    |
| A6 --> B1    | 1.37 | 0.44                | 0.60       | 0.10    | 26    |
| A9 --> B1    | 1.33 | 0.44                | 0.59       | 0.07    | 17    |
| A10 --> B1   | 1.28 | 0.44                | 0.56       | 0.14    | 35    |
| A9&A6 --> B1 | 1.67 | 0.44                | 0.73       | 0.04    | 11    |
| A5&A3 --> B1 | 1.62 | 0.44                | 0.71       | 0.04    | 10    |
| A4&A3 --> B1 | 1.58 | 0.44                | 0.70       | 0.06    | 16    |
| A7&A2 --> B1 | 1.57 | 0.44                | 0.69       | 0.07    | 18    |
| A9&A7 --> B1 | 1.56 | 0.44                | 0.69       | 0.04    | 11    |

Set size가 2인 경우, 전반적 만족도(B1)에 대한 긍정적 평가에 영향을 주는 규칙으로 연결의 매끄러움, 논리성, 단어의 적재적소 사용, 자연스러움, 의미와 정보 전달에 대한 긍정적 평가인 것으로 나타났다(향상도 순). 연결의 매끄러움이 높은 경우에 연결의 매끄러움 여부가 가정되지 않는

우, 지속 사용 의향(B2)에 대한 긍정적 평가에 영향을 주는 규칙으로는 연결의 매끄러움, 자연스러움, 단어의 적재적소 사용, 내용 동등성, 논리성으로 조사되었다(향상도 순). 가장 향상도가 큰 요인은 연결의 매끄러움으로서, 연결의 매끄러움에 긍정 평가를 하면 지속 사용 의향에도 긍정 평가를 할 가능성이 51% 증가하는 것을 확인할 수 있다.

다음으로 향상도가 큰 요인은 자연스러움으로서, 응답자가 기계번역문에 대하여 자연스럽다고 평가하면 지속 사용 의향이 46% 증가하는 것으로 해석할 수 있다. Set size가 3인 경우에 향상도가 가장 높게 나타난 규칙은 “만약 자연스러움과 논리성에서 모두 긍정 평가를 할 경우 지속 사용 의향에 대하여 긍정 평가를 한다”이며 그 향상도는 1.54로 나타났다. 두 번째 중요한 규칙으로는 “자연스러움과 단어의 적재적소 사용에서 긍정 평가를 하면 지속 사용 의향에 대하여 긍정 평가를 한다”는 규칙으로 향상도는 1.53으로 나타났다.

대체 프로그램 탐색 의지(B3)에 부정적 영향을 주는 규칙들은 <표 5>에 요약되어 있다. Set size가 2인 경우, 대체 프로그램을 탐색(B3)할 필요성을 덜 느끼게 하는 요인으로는 단어의 적재적소 사용과 의미와 정보의 전달에 대한 긍정 평가로 조사되었다. 즉, 연결의 매끄러움이나 자연스러움 등 완벽한 번역을 위한 이차적인 느낌보다는 일차적인 번역의 요인이라고 할 수 있는 단어의 선택과 의미와 정보 전달에 충실하다면 굳이 또 다른 번역 프로그램을 찾을 필요성을 덜 느끼게 하는 것으로 판단된다.

<표 4> 지속 사용 의향(B2)에 대한 규칙: Ai high --> B2 high

| Rule         | Lift | Expected Confidence | Confidence | Support | Count |
|--------------|------|---------------------|------------|---------|-------|
| A5 --> B2    | 1.51 | 0.57                | 0.86       | 0.12    | 30    |
| A9 --> B2    | 1.46 | 0.57                | 0.83       | 0.10    | 24    |
| A6 --> B2    | 1.39 | 0.57                | 0.79       | 0.14    | 34    |
| A1 --> B2    | 1.33 | 0.57                | 0.76       | 0.22    | 56    |
| A7 --> B2    | 1.28 | 0.57                | 0.73       | 0.13    | 32    |
| A9&A7 --> B2 | 1.54 | 0.57                | 0.88       | 0.06    | 14    |
| A9&A6 --> B2 | 1.53 | 0.57                | 0.87       | 0.05    | 13    |
| A9&A5 --> B2 | 1.51 | 0.57                | 0.86       | 0.05    | 12    |
| A9&A4 --> B2 | 1.51 | 0.57                | 0.86       | 0.05    | 12    |
| A5&A3 --> B2 | 1.51 | 0.57                | 0.86       | 0.05    | 12    |

<표 5> 대체 프로그램 탐색(B3)에 대한 규칙: Ai high --> B3 low

| Rule          | Lift | Expected Confidence | Confidence | Support | Count |
|---------------|------|---------------------|------------|---------|-------|
| A6 --> B3     | 1.05 | 0.49                | 0.51       | 0.09    | 22    |
| A10 --> B3    | 1.02 | 0.49                | 0.50       | 0.12    | 31    |
| A3&A10 --> B3 | 1.23 | 0.49                | 0.60       | 0.05    | 12    |
| A9&A6 --> B3  | 1.23 | 0.49                | 0.60       | 0.04    | 9     |
| A8&A10 --> B3 | 1.22 | 0.49                | 0.59       | 0.09    | 22    |
| A8&A3 --> B3  | 1.19 | 0.49                | 0.58       | 0.04    | 11    |
| A8&A6 --> B3  | 1.15 | 0.49                | 0.56       | 0.06    | 14    |



하지만 이 두 규칙에 대한 향상도는 각각 1.05, 1.02로서 현격한 차이를 나타내지는 않는 것으로 보인다. Set size가 3인 경우에 향상도가 가장 높게 나타난 규칙은 “억양 적절성과 의미와 정보 전달에서 모두 긍정 평가를 할 경우 대체 프로그램 탐색 의지는 낮다”이며 그 향상도는 1.23으로 나타났다. 두 번째 중요한 규칙으로는 “자연스러움과 단어의 적재적소 사용에서 긍정 평가를 하면 대체 프로그램 탐색 의지는 낮다”로서 향상도는 1.23으로 나타났다.

유료 번역기 사용 의향(B4)에 부정적 영향을 주는 규칙들은 <표 6>에 요약되어 있다. Set size가 2인 경우, 유료 번역 프로그램을 탐색(B4)할 필요성을 덜 느끼게 하는 요인으로는 의미와 정보의 전달에 대한 긍정 평가가 가장 중요한 요인으로 나타났다.

라크 구분, 충실성, 내용 동등성이 영향을 주는 것으로 나타났으나 향상도는 1.04 이하로 비교적 영향력은 크지 않은 것으로 보인다. Set size가 3인 경우에는 “충실성과 의미와 정보 전달에서 모두 긍정 평가를 할 경우 유료 번역기를 사용하고자 하는 의향은 낮다”는 규칙이 향상도가 가장 높게 나타났으며 향상도는 1.18로 추정되었다. 다음으로 의미 있는 규칙으로는 “단어의 적재적소 사용과 의미와 정보의 전달에서 모두 긍정 평가를 하면 유료 번역기 탐색 의지는 낮다”로서 향상도는 1.15로 나타났다.

추천 의도(B5)에 긍정적 영향을 주는 규칙들은 <표 7>에 요약되어 있다. Set size가 2인 경우, “연결의 매끄러움이 높으면 추천 의도가 높다”는 규칙은 향상도 1.62로 나타나 set size 2인 규칙 중 가장 높은 향상도를 나타냈다. 기존 프로그램의

<표 6> 유료번역기 사용 의향(B4)에 대한 규칙: Ai high --> B4 low

| Rule          | Lift | Expected Confidence | Confidence | Support | Count |
|---------------|------|---------------------|------------|---------|-------|
| A10 --> B4    | 1.14 | 0.80                | 0.92       | 0.23    | 57    |
| A6 --> B4     | 1.04 | 0.80                | 0.84       | 0.14    | 36    |
| A4 --> B4     | 1.02 | 0.80                | 0.82       | 0.28    | 70    |
| A2 --> B4     | 1.02 | 0.80                | 0.82       | 0.28    | 68    |
| A1 --> B4     | 1.01 | 0.80                | 0.81       | 0.24    | 60    |
| A2&A10 --> B4 | 1.18 | 0.80                | 0.95       | 0.14    | 36    |
| A6&A10 --> B4 | 1.15 | 0.80                | 0.93       | 0.10    | 26    |
| A10&A1 --> B4 | 1.14 | 0.80                | 0.92       | 0.13    | 33    |
| A3&A10 --> B4 | 1.12 | 0.80                | 0.90       | 0.07    | 18    |
| A6&A2 --> B4  | 1.11 | 0.80                | 0.89       | 0.10    | 24    |

기준에 사용하고 있는 기계번역 프로그램이 번역하고자 하는 문장의 의미와 정보를 잘 전달하고 있는 경우 유료의 타 번역프로그램을 찾고자 하는 필요성이 적어진다는 의미로서 향상도는 1.14로 나타났다. 기타 단어의 적재적소 사용, 단

추천에 가장 큰 영향을 주는 요인으로 연결의 매끄러움 여부가 선별되었다. 그 다음 영향력이 큰 요인은 단어의 적재적소 사용, 논리성 등이며 각각 향상도 1.50, 1.46을 나타냈다. Set size가 3인 경우에는 “자연스러움과 단어의 적재적소 사용에서 모두 긍정 평가를 할 경우 기존 번역프로그램을 추천하겠다”는 규칙이 향상도 2.27로서 set size 3인 규칙 중에서 가장 중요한 규칙으로 나타났다. 즉, 자연스러움이나 단어의 적재적소 사용이라는 조건이 없는 경우에 비하여 이 조건이 있을 때 추천할 가능성이 1.27배 향상되는 것을 의미한다. 두 번째 중요한 규칙으로서 논리성과 단어의 적재적소 사용에서 모두 긍정 평가를 하면 추천 의도는 82% 향상하는 것으로 분석되었다.

<표 8>은 각각의 지속 사용 의향 변수에 관하여 생성된 규칙에서 유용성 변수가 포함된 횟수를 요약하고 있다. 전반적 만족(B1), 지속 사용 의향(B2), 추천 의도(B3)에 긍정적 영향을 주는 유용성 변수는 2회 이상 규칙을 생성한 연결의 매끄러움(A5), 단어의 적재적소 사용(A6), 논리성(A7), 자연스러움(A9)에 대한 긍정 평가로 나타났다. 반면, 대체 프로그램 탐색(B3)과 유료번역기 사용 의향(B4)을 억제하는 유용성 변수는 단어의 적재적소 사용(A6)과 의미와 정보의 전달(A10)으로 조사되었다. 현재 사용하고 있는 기계번역 프로그램을 대신할 새로운 프로그램을 찾으려는 수고를 하지 않는 조건은 번역의 가장 기본적인 요소라 할 수 있는 적절한 단어의 선정과 전달하고자 하

<표 7> 추천 의도(B5)에 대한 규칙: Ai high --> B5 high

| Rule          | Lift | Expected Confidence | Confidence | Support | Count |
|---------------|------|---------------------|------------|---------|-------|
| A5 --> B5     | 1.62 | 0.26                | 0.43       | 0.06    | 15    |
| A6 --> B5     | 1.50 | 0.26                | 0.40       | 0.07    | 17    |
| A7 --> B5     | 1.46 | 0.26                | 0.39       | 0.07    | 17    |
| A4 --> B5     | 1.34 | 0.26                | 0.35       | 0.12    | 30    |
| A9 --> B5     | 1.31 | 0.26                | 0.34       | 0.04    | 10    |
| A9&A6 --> B5  | 2.27 | 0.26                | 0.60       | 0.04    | 9     |
| A7&A6 --> B5  | 1.82 | 0.26                | 0.48       | 0.05    | 12    |
| A5&A10 --> B5 | 1.80 | 0.26                | 0.48       | 0.04    | 10    |
| A4&A3 --> B5  | 1.65 | 0.26                | 0.43       | 0.04    | 10    |
| A6&A1 --> B5  | 1.64 | 0.26                | 0.43       | 0.05    | 13    |

<표 8> 지속 사용 의향 별 유용성 변수의 규칙 생성 횟수

|    | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | A10 |
|----|----|----|----|----|----|----|----|----|----|-----|
| B1 |    | 1  | 2  | 1  | 2  | 2  | 3  |    | 3  | 1   |
| B2 | 1  |    | 1  | 1  | 3  | 2  | 2  |    | 5  |     |
| B3 |    |    | 2  |    |    | 3  |    | 3  | 1  | 3   |
| B4 | 2  | 3  | 1  | 1  |    | 3  |    |    |    | 5   |
| B5 | 1  |    | 1  | 2  | 2  | 4  | 2  |    | 2  | 1   |

는 의미나 정보가 제대로 되느냐의 여부에 달린 것으로 해석할 수 있다. 현재 사용 중인 번역 프로그램에 대한 전반적으로 만족하고 타인에게 추천까지 하는 적극적인 행위는 번역이 한 차원 높은 품질을 갖는다고 말 할 수 있는 연결의 매끄러움, 논리성, 자연스러움 등이 수반될 때 일어난다고 종합적으로 해석할 수 있을 것이다.

#### 4. 결론 및 시사점

본 연구에서는 오늘날 다양한 언어의 번역에 쓰임이 증가하고 있는 기계번역 프로그램에 대한 평가와 향후에도 지속적으로 사용하고 추천할 의향이 있는가에 관한 주제에 대하여 인과관계를 파악하고자 하였다. 가장 널리 이용되고 있는 영문의 국문 번역을 예시로 들어 유용성을 10개 변수로 평가하고 지속 사용 의향을 5개의 변수로 평가하였다. 기계번역기를 사용한 경험이 있는 응답자를 대상으로 설문을 실시하고 연관성 규칙을 적용하여 유용성과 지속 사용 의향 간에 의미 있는 규칙들을 탐색하였다.

생성된 수많은 규칙 중에서 해석의 용이성을 위하여 set size가 2와 3인 규칙들만을 대상으로 한정하고 향상도가 높은 일정 수의 규칙만을 선별하여 해석을 시도하였다. 기존에 사용하던 기계번역 프로그램에 대한 전반적 만족 및 추천에 관련된 전반적 만족, 지속 사용 의향, 추천 의도에 영향을 규칙에 빈도 높게 등장하는 유용성 변수는 좀 더 고급스럽고 차원 높은 번역이 되었다고 인지하게 하여 주는 유용성 변수들이 선정되었다. 즉, 문장 연결의 매끄러움, 단어의 적재적소 사용, 문장의 논리성, 자연스러움이 주요한 유용성 변수로 분석되었다. 반면 기존의 기계번역 프로그램에 대한 불만족이 커져 대체 프로그램이나 유료번역기를 찾고자 하는데 영향을 주는 유용성 변수로

는 단어의 적재적소 사용과 의미와 정보의 전달이 선정되었다. 번역의 일차적 목적이며 가장 낮은 차원에서 번역이 성공적이었는지를 판별할 수 있는 유용성 변수들로 보이며, 이러한 목적이 달성되면 대체 프로그램을 찾고자 하는 의도는 낮아지는 것으로 분석되었다.

본 연구에서는 이용자들이 이미 상당 기간 기계번역 프로그램을 사용한 경험이 있기 때문에 여러 가지 번역문 중 가장 연구의 취지에 부합한다고 판단되는 한 가지 예시문을 제시하여 설문에 응답토록 유도하였다. 설문응답의 신뢰도를 높이기 위하여 향후에는 좀 더 긴 지문을 사용하여 기계번역의 장단점이 드러날 수 있도록 하고 유용성과 지속 사용 의향에 영향을 주는 매개변수를 도입할 필요가 있을 것으로 보인다. 또한 사용자의 인지를 한 가지 차원이 아닌 두 가지 차원에서 평가하고 이 두 가지 인지의 차이로 지속 사용 의향을 설명할 수 있는지 시도하여 볼 필요가 있다고 사료된다.

#### 참고문헌

- [1] 김정우, 2004, "번역 평가의 실제," 번역학연구, 5(2), 5-29.
- [2] 김종희, 2016, "포털사이트 (portal site) 번역기의 오역에 관한 연구," 일어일문학, 69, 141-158.
- [3] 김혜림, 강려영, 2015, "한국 공공기관 중문 웹사이트 오류 분석" 번역학연구, 16(3), 55-84.
- [4] 박경리, 조수연, 전종섭, 2013, "웹기반 영한 번역을 위한 중간 언어의 효율성 연구." 통번역학연구, 17(3), 201-229.
- [5] 이향, 2010, "번역품질평가의 현황과 과제" 번역학연구, 11(2), 107-125.
- [6] 임순정, 한미선, 2014, "구글 번역의 사례연구,"

- 번역학연구, 15(1), 177-209.
- [7] 조성빈, 하병천, 2010, “공급체인 파트너와의 신뢰와 물류효율성,” 한국경영공학회지, 15(3), 173-186.
- [8] 최기선, 2001, 대용량 음성(음향)/언어/영상 DB 구축 및 표준화, 한국과학기술원.
- [9] 최효은, 이지은, 2017, “특허 기계번역 결과물의 평가-KIPRIS 의 무료 한영 기계번역을 중심으로,” 통역과 번역, 19(1), 139-178.
- [10] Bahdanau, D., Cho, K., and Bengio, Y.(2015), “Neural machine translation by jointly learning to align and translate,” 3rd International Conference on Learning Representations(ICLR2015), 1-15.
- [11] Banerjee, S. and Lavie, A.(2005), “METEOR: An automatic metric for MT evaluation with improved correlation with human judgments,” Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization, 29, 65-72.
- [12] Bennett, W.S. and Slocum, J.(1985), “The LRC machine translation system,” Computational Linguistics, 11(2-3), 111-121.
- [13] Brown, P.F., Pietra, V.J.D., Pietra, S.A.D., and Mercer, R.L.(1993), “The mathematics of statistical machine translation: Parameter estimation,” Computational linguistics, 19(2), 263-311.
- [14] Chen, K.H. and Chen, H.H.(1996), “A hybrid approach to machine translation system design,” Computational Linguistics and Chinese Language Processing, 1(1), 159-182.
- [15] Doddington, G.(2002), “Automatic evaluation of machine translation quality using n-gram co-occurrence statistics,” Proceedings of the second international conference on Human Language Technology Research. Morgan Kaufmann Publishers Inc., 138-145.
- [16] Han, A.L.F., Wong, D.F., Chao, L.S., He, L., Lu, Y., Xing, J., and Zeng, X.(2013), “Language-independent model for machine translation evaluation with reinforced factors,” Proceedings of the 14th International Conference of Machine Translation Summit. International Association for Machine Translation Press, 215-222.
- [17] Johnson, M., Schuster, M., Le, Q.V., Krikun, M., Wu, Y., Chen, Z., Throat, N., Vigas, F., Wattenberg, M., Corrado, G., Dean, J., and Hughes, M.(2016), “Google’s multilingual neural machine translation system: enabling zero-shot translation,” arXiv preprint arXiv:1611.04558, 1-15.
- [18] Koehn, P. and Monz, C.(2006), “Manual and automatic evaluation of machine translation between european languages,” Proceedings of the Workshop on Statistical Machine Translation. Association for Computational Linguistics, 102-121.
- [19] Lagarda, A.L., Alabau, V., Casacuberta, F., Silva, R., and Diaz-de-Liano, E.(2009), “Statistical post-editing of a rule-based machine translation system,” Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers. Association for Computational Linguistics, 217-220.
- [20] Lommel, A.R., Burchardt, A., and Uszkoreit, H.(2013), “Multidimensional Quality Metrics: A

- Flexible System for Assessing Translation Quality," Proceedings of ASLIB Translating and the Computer, 34, 1-7.
- [21] Okpor, M.D.(2014), "Machine translation approaches: issues and challenges," International Journal of Computer Science Issues(IJCSI), 11(5), 159-165.
- [22] Papineni, K., Roukos, S., Ward, T., and Zhu, W.J.(2002), "BLEU: a method for automatic evaluation of machine translation," Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 311-318.
- [23] Tillmann, C., Vogel, S., Ney, H., Zubiaga, A., and Sawaf, H.(1997), "Accelerated DP based search for statistical translation," European Conference on Speech Communication and Technology, 2667 - 2670.
- [24] Wolk, K., and Marasek, K.(2015), "Neural-based machine translation for medical text domain," Based on European Medicines Agency leaflet texts. Procedia Computer Science, 64, 2-9.
- [25] Wu, Y., Schuster, M., Chen, Z., Le, Q.V., Norouzi, M., Macherey, W., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., and Dean, J.(2016), "Google's neural machine translation system: Bridging the gap between human and machine translation," arXiv preprint arXiv:1609.08144, 1-23.
- [26] Zaretskaya, A., Pastor, G.C., and Seghiri, M.(2016), "A Quality Evaluation Template for Machine Translation," Translation Journal, 19(1).