JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 19, No. 1, pp. 75-83, Jan. 2018

어휘의미분석 기반 다국어 어휘대역 서비스

류 법 모
부산외국어대학교 동남아창의융합학부 언어처리창의융합전공

Multilingual Word Translation Service based on Word Semantic Analysis

Pum-Mo Ryu
Department of ICT & Language Processing, School of Southeast Asian Studies, Busan University of Foreign Studies

[요 약]

다문화 가정 구성원은 언어 차이 때문에 자녀 교육에서 어려움을 겪고 있다. 이와 같은 어려움을 해결하기 위해서는 실생활에 필요한 한국어 용어들을 간편하고 신속하게 접근할 수 있는 스마트 번역 서비스를 이들에게 제공할 필요가 있다. 그러나 현재의 자동 번역 기술은 영어, 중국어, 일본어 등의 주요 국가 언어 중심으로 개발 되고 있으며, 자녀의 교육, 공공기관과의 소통 등 특수 목적의 용어들은 번역하기에는 한계가 있다. 본 연구에서는 초급 수준의 한국어를 이해하고 있는 다문화가정 구성원을 위한 실시간 자동 어휘대역어 서비스를 제안한다. 어휘대역어 서비스는 한국어 문장에 표현된 각 단어들의 의미를 자동으로 분석하여 다국어 대역어를 제공한다. 이를 위하여 한국어 의미분석 연구, 다국어 번역지식 구축 연구, 언어교육 연구의 융합연구를 수행하였다. 어휘대역서비스를 베트남, 일본 출신의 결혼이주여성을 대상으로 평가하여 의미있는 평가결과를 얻었다.

[Abstract]

Multicultural family members have difficulty in educating their children due to language differences. In order to solve these difficulties, it is necessary to provide smart translation services that enable them easily and quickly access real-life vocabularies. However, the current automatic translation technology is being developed in dominant languages such as English, Chinese, and Japanese. There are also limitations to translating special-purpose terms such as documents of schools and instructions of public institutions. In this study, we propose a real-time automatic word translation service for multicultural family members who understand beginner level Korean. The service automatically analyzes the semantics of each word in the Korean sentences and provides a word-by-word translation. This study includes semantic analysis research for Korean language, building multilingual translation knowledge, and fusion study of language education. We evaluated the word translation service for migrant women from Vietnam and Japan and obtained meaningful evaluation results.

색인어 : 다문화가정, 대역지식, 어휘의미분석, 어휘 대역 서비스

Key word : Multicultural family, Translation knowledge, Word semantics analysis, Word translation service

http://dx.doi.org/10.9728/dcs.2018.19.1.75

Received 14 November 2017; **Revised** 23 January 2018
Accepted 29 January 2018

***Corresponding Author; Pum-Mo Ryu**

Tel: +82-51-509-6782
E-mail: pmryu@bufs.ac.kr

Ⅰ. 서 론

2015년 행정안전부 외국인 주민 현황에 따르면, 결혼이민 및 혼인귀화자의 구성은 중국인 25.47%, 한국계 중국인 24.50%, 베트남인 24.05%, 필리핀인 6.97%, 일본인 5.34%, 캄보디아인 2.66%, 동구 유럽 출신 2.21%의 구성을 보이고 있으며, 동북아 및 동남아 결혼 이민 및 귀화자가 증가하고 있는 추세이다[**]. 결혼이민 및 혼인 귀화자 중에 여성의 비율이 80%이상을 차지하고, 다문화 가정 자녀들도 점차 증가하는 것으로 나타나고 있으며, 이에 따라 다문화 가정에서 나타나는 사회적인 문제들도 증가하고 있다. 여성결혼 이민 및 귀화자들은 한국어 활용 능력이 낮은 상태에서 자녀교육을 지도하고 있기 때문에, 가정, 자녀, 학교, 공공기관 간의 소통의 문제가 발생하며, 더불어 사회 및 문화, 가치관 및 습관의 차이로 다문화 가정과 사회와의 소통에 문제도 발생하고 있다. 이와 같은 상황에서 국가 차원에서 이들에게 초급 수준의 한국어 교육을 제공하고 있으나, 학교 가정통신문, 뉴스, 관광지 안내문 등의 고급 정보 등을 이해하기에 한계가 있다. 따라서 다문화 가정에서 언어 및 사회문화적 차이로 겪고 있는 지역 주민과의 소통, 생활, 자녀 교육 등의 문제를 해결하기 위해서는 기존의 기초 수준의 한국어 교육 수준을 넘어서 실생활에 필요한 용어들을 간편하고 신속하게 접근할 수 있는 기재를 제공할 필요가 있다.

현재 자동 번역 기술은 영어, 중국어, 일본어 등의 주요 국가 언어 중심으로 개발 되고 있으며, 자녀의 교육, 공공기관과의 소통을 위한 수준의 용어들은 번역 수준의 한계가 있고, 동남아 언어와 같은 특수어에 대한 자동 번역 서비스가 아직 구체적으로 제공되지 않고 있는 문제가 있다[1, 2].

본 연구에서는 초급 수준의 한국어를 이해할 수 있는 다문화 가정의 결혼이주 여성, 한류 관광객 등을 위한 실시간 자동 어휘대역 서비스 시스템을 제안하고자 하며, 특히 동남아 국가언어의 개발에 집중하고자 한다. 그림 1은 다국어 어휘대역서비스의 예시이다. 입력된 한국어 문장을 대상으로 베트남어, 태국어, 중국어, 일본어 어휘 대역결과를 보여주고 있다.

본 논문은 다음과 같이 구성되어 있다. 2장에서 선행연구를 살펴보고, 3장에서 제안하는 다국어 어휘대역 서비스를 설명하며, 4장에서 실험 및 평가결과를 제시한다. 마지막으로 5장에서 결론을 맺는다.

Ⅱ. 선행연구

선행연구에서는 외국인 대상 한국어 교육의 문제점, 자동번역 시스템의 현황 그리고 한국어 의미분석 및 대역어 생성 연구에 대해서 설명하고 개선방안을 제시한다.

2-1 외국인 대상 한국어 교육

1990년대 이후 다문화 가정의 문제가 사회적 이슈로 떠오르면서 정부와 사회의 지원이 증가하고 있다. 그러나 결혼이민자 및 귀화자, 외국인 노동자들을 위한 교육 및 사회적 환경이 쌍방적이기 보다는 일방적인 교육 서비스 제공에 한정되어 있는 문제가 있다. 관련 연구로 결혼 이주여성과 그 자녀의 소통문제[3], 한국어 교육[4], 한국어 교육 정책[5], 다문화 가정 아동대상 한국어 교육[6], 다문화가정 자녀 한국어교육 및 학교 교육

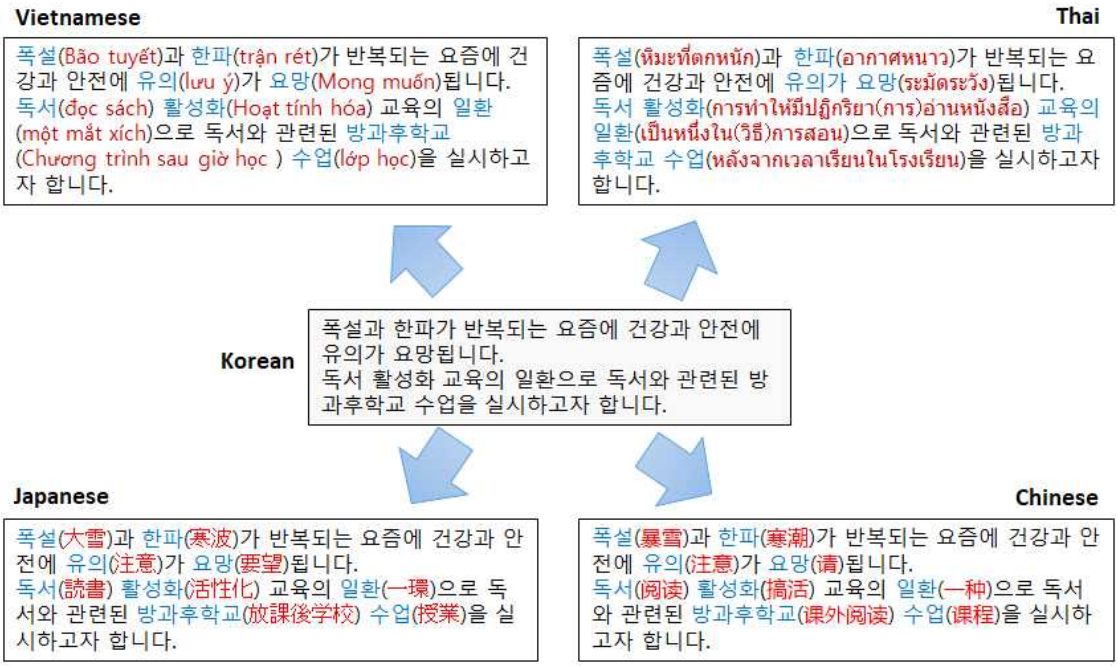

그림 1. 어휘대역서비스 예시
Fig. 1. An example of word Translation service

환경에 대한 교사 인식[7], 다문화가정 자녀의 학교생활에 관한 연구[8] 등이 있으며, 주로 다문화가정 자녀의 학교생활, 교사의 인식, 한국어 교육 등에 국한되어 있으며, 다문화가정의 주부와 학교 및 공공기관 간의 소통에 대한 연구는 부재한 실정이다. 일부 연구에서는 알림장, 가정통신문 등은 일반적인 상황에서 접하기 어려운 것들이며, 이를 바탕으로 교사에게 자녀 교육에 대한 상담을 해야 하는 복합적인 상황들의 경우 여성결혼이민자의 부담은 더욱 커지고 있음을 지적하고 있다[9, 10]. 따라서, 광고나 간판, 공고문, 성적표, 가정통신문 등 여성결혼이민자들이 쉽게 접할 수 있는 읽기 자료를 이용한 교육이 필요하고, 가정통신문 어휘를 한국어교재 기초어휘로 지정할 필요가 있다. 다양한 상황에 대한 한국어 교육 확대가 필요하나, 현실적으로 어려운 문제가 있기 때문에 사회적 약자를 위하여 ICT 기술과 어학을 융합한 우리말 이해 보조도구의 연구가 필요하다.

2-2 자동번역

문장단위 자동번역 기술은 활발하게 연구되고 있으나, 사람 수준의 번역이 어렵기 때문에 투자대비 활용성이 낮은 단점이 있다. 영한 자동번역 시스템의 경우 전문가의 수동평가 정확률이 81.9%에 머물고 있으며 [11], 최신 Google 신경망 한영 번역 시스템의 경우 BLEU값이 0.25수준이다 [12]. 비교적 번역이 쉬운 일한, 한일 자동번역 시스템의 경우 BLEU 값이 0.31 ~ 0.32 수준이다 [13, 14]. BLEU 값은 기계번역 시스템을 자동평가하기 위한 기준의 하나로 0에서 1사이의 값을 가진다. 일반적으로 BLEU 값이 0.5인 경우 전문가에 의한 수동평가에서 정확도 80% 정도를 나타낸다. 자동번역 시스템의 평가 결과는 평가셋의 구성 및 적용 도메인에 따라서 크게 차이를 보이기 때문에 실제 사용자에 의한 평가가 필요하다. 또한 한국어-동남아언어 자동번역 시스템은 구글에서 서비스하고 있으나, 아직 개발 초기단계이기 때문에 번역의 정확도가 낮은 단점이 있다. 따라서 자연어 문장에 대한 사람수준의 번역시스템 개발은 기술적 한계, 예산과 시간의 문제 때문에 어려움이 있기 때문에, 초급 한국어 능력자를 대상으로 어휘수준의 한국어-모국어 대역 시스템 개발이 필요하다.

2-3 한국어 의미분석 및 대역어 생성

한국어는 여러 형태소가 결합하여 하나의 어절을 이루는 교착어이다. 즉 하나의 어절이 여러 개의 형태소로 구성되어 있다. 한국어는 형태소 분석이 매우 어려운 언어에 속하나, 꾸준한 연구로 인해 현재 약 98%의 정확률로 형태소 복원과 품사 구분이 가능하다 [15,16]. 그러나 기계번역을 위해서는 형태소/품사 분석 후에 어휘의미 중의성 해소 과정이 필수적이다. 의미 중의성을 가진 어휘 중 어원이 다른 어휘는 '동형이의어'로, 동일 어원이 세분화된 의미를 가지면 '다의어'로 구분한다. 대부분의 단어는 자동번역을 위해서 동형이의어만 구분해도 되지만, 표 1에서 '일'(명사)과 같은 단어는 다의어 수준의 중의성 해소 기술을 필요하다.

표 1. 단어 "일"에 대한 동형이의어, 다의어 예시
Table 1. Example of homograph and polysemy for a word "일"

Word	Homograph	Polysemy	Definition
일	01	01_01	A physical or brain activity to achieve something for a certain time in a certain place. or the target of the activity (work)
		01_07	a situation or circumstance. (situation)
	05	05_01	A word representing a quantity one. (one)
		05_02	A word representing that the order is first. (first)

기존의 한국어 의미분석기는 동형이의어 분별정확률 약 96.5%, 다의어 분별정확률 약 66% 수준의 성능을 보이고 있으며, 형태소 복원, 품사 태깅, 동형이의어 분별, 다의어 분별 기능을 갖추고 있다 [15]. 의미중의성 해소는 코퍼스학습 방식이 가장 정확률이 높으나, 코퍼스에 나타나지 않은 어휘의 의미중의성 분별 정확률은 매우 낮은 단점이 있다. 따라서, 학습데이터 부족 문제를 해결하기 위하여 어휘의미망과 같은 추가적인 언어자원을 활용하는 연구도 진행되고 있다 [17]. 어휘의미망은 단어들의 상하관계와 용언의 필수논항의 의미제약 정보를 포함하고 있다.

Ⅲ. 다국어 어휘대역 서비스

3-1 시스템 구성도

시스템 구성은 그림 2와 같이 "다국어 어휘 대역 서비스"는 "한국어 어휘 의미분석 모듈", " 다국어 번역지식", "시스템 테스트 및 평가셋"으로 구성되어 있다. "한국어 어휘 의미분석 모듈"은 입력 문장을 형태소 단위로 구분하고, 각 형태소의 의미를 "단어 의미분석 기준 사전"에서 선택하여 해당 의미기호를 할당한다. "다국어 번역지식"은 의미분석된 형태소의 외국어 대역어를 저장한다. "다국어 번역지식"은 "단어 의미분석 기준 사전"에 정의되어 있는 의미 단위별로 대역어가 할당되어야 하기 때문에 두 사전 사이에 "단어 의미 맵핑" 단계를 거친다. 기존의 한국어 어휘 의미분석 모듈과 다국어 번역 지식은 특정 도메인을 가정하지 않고 있다. 본 연구는 다문화가정의 결혼이주여성의 자녀들의 교육 및 학교생활에 도움을 주고자하는 목적이 있기 때문에 학교 가정통신문을 수집 및 분석하여 시스템을 평가한다.

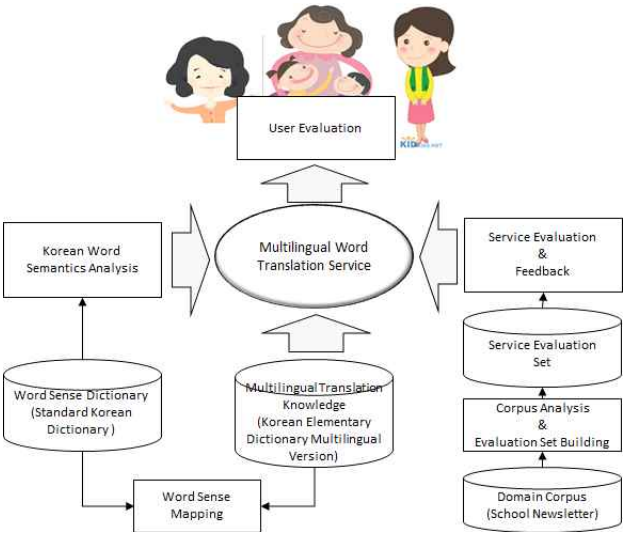

그림 2. 다국어 어휘 대역 서비스 시스템 구성도
Fig. 2. System architecture of multilingual word translation service

3-2 한국어 어휘의미분석

표준국어대사전[1]에 등재된 약 51만 어휘 중 125,600 어휘(약 25%)가 동형이의어이다. 이러한 동형이의어 중 용언의 경우는 형태소 분석 단계에서 어휘의미 분별된다면 구문 분석 시 많은 중의성을 해소할 수 있으며 정확한 의존관계를 분석할 수 있다. 예를 들어 '차다'의 경우 자동사(차다__01: "일정한 공간에 가득하게 되다."), 타동사(차다__02: "발로 내어 지르다.", 차다__03: "물건을 몸의 한 부분에 달아매다."), 형용사(차다__04: "몸에 닿은 물체나 대기의 온 도가 낮다.")와 같이 다른 품사를 가진다. 또한, 서술성 명사는 용언 성격을 내포하고 있어 구문분석 전에 서술성 명사가 분별된다면 정확한 의존관계를 분석할 수 있다. 예를 들어, "아이의 아버지는 중국을 상대로 무역한다"의 문장에서 명사 '상대'가 '상대__04: "서로 마주 대함." ' 로 분별되면 어절 "중국을"과 "상대로"가 의존관계를 맺을 수 있다. 한국어 어휘의미분석을 위해서는 어휘의미망, 의미주석 말뭉치, 의미역 부착 격특사전, 의존관계 말뭉치 등의 어휘의미 자원이 필요하다. 적용 방법으로는 의미주석 말뭉치를 기반으로 Hidden Markov Model(HMM)을 이용한 기계학습 방법, 어휘의미망을 이용한 어휘 간 의미제약 방법이 있다. 어휘의미망 기반 방법에서는 WordNet을 기반으로 하는 Korlex[18]와 표준국어대사전을 기반으로 구축된 UWordMap[17]이 있다. 이 방법은 명사의 상위어와 용언의 하위범주화 정보를 이용하며, 말뭉치기반의 방법에 비해서 앞으로 재현율 문제에서 유리한 측면이 많지만, 재현에 성공하더라도 정확률은 다소 낮은 것으로 확인되었다. 이런 문제를 해결하기 위해서 기존의 말뭉치기반 기계학습방법과 어휘의미망을 같이 사용하는 방법이 연구되었고, 정확률이 조금 향상되는 것을 확인하였다. 최근에는 워드임베딩(Word Embedding)을 이용한 딥러닝 기반 방법이 연구되고 있다 [17].

그림 3은 한국어 어휘의미분석의 예시를 보여준다. 한국어 문장을 입력으로 받아서, 형태소 단위로 분리하고, 각 형태소에 품사 태그를 부착한다. 예시에서 "교장선생님에게" 어절은 세 개의 형태소 "교장", "선생님", "에게"로 분할하고, 각각의 형태소에 대해서 NNG(보통명사), NNG(보통명사), JKB(부사격조사) 태그를 부착하였다. 나머지 예시에서 제시된 형태소 품사는 각각 NNP(고유명사), JX(보조사), VV(동사), EP(과거시제 선어말어미), EF(종결어미), SF(문장종결기호) 를 나타낸다. 다음 단계에서 각각의 형태소에 의미 분석 정보를 부착한 예를 보여준다. 동일한 어휘라도 문맥에 따라 다른 의미를 가질 수 있기 때문에 미리 정의한 의미분석 체계에서 해당 의미에 해당하는 기호를 선택하여 부착한다. 예를 들어 "사장/NNG"는 그림 3의 단어의미분석 기준사전에서 "15: 회사의 책임자"에 해당하기 때문에 의미번호 "15"를 할당한다. 마찬가지로 "교장/NNG"는 "03: 각급 학교의 으뜸 직위"에 해당하기 때문에 의미번호 "03"을 할당하였고, "사과하/VV"는 "02: 잘못을 인정하고 용서를 빌다"에 해당하기 때문에 의미번호 "02"를 할당하였다. 본 연구에서는 [17]에서 개발한 한국어 의미분석기를 적용하였다. 이 의미분석기는 표준국어대사전에서 정의한 표제어의 의미구분을 의미 분류의 기준으로 삼고 있다.

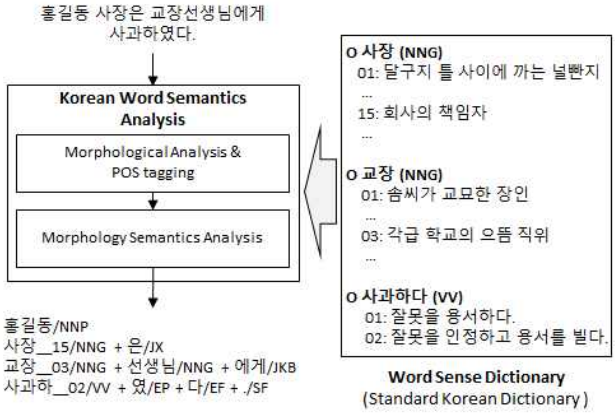

그림 3. 한국어 어휘의미분석 예시
Fig. 3. An example of Korean word semantic analysis

3-3 다국어 어휘 번역 지식 구축

한국어 어휘의미분석 시스템은 표준국어대사전의 다의어 수준의 어휘 의미분석 결과를 제공한다. 본 연구에서는 국립국어원에서 구축한 한국어기초사전 다국어 버전[2]을 번역지식으로 이용한다. 이 사전은 5만 여개의 한국어 기초 어휘에 10개

1) http://stdweb2.korean.go.kr

2) https://krdict.korean.go.kr

국어(영어, 일본어, 프랑스어, 스페인어, 아랍어, 몽골어, 베트남어, 태국어, 인도네시아어, 러시아어) 대역어를 수록하고 있다. 한국어 어휘의미분석 결과에 대응하는 외국어 대역어를 제공하기 위하여, 표준국어대사전의 다의어 수준 어휘와 한국어 기초사전 어휘 사이의 정렬이 필요하다. 본 연구에서는 표준국어대사전과 한국어 기초사전 어휘 사이의 뜻풀이 유사도를 이용하여 자동으로 맵핑하고, 나머지는 언어처리 및 어학을 전공한 전문가들이 수작업으로 정렬하였다. 그림 4는 "폭설"에 대한 두 사전의 의미 구분과 각 의미별 뜻풀이를 보여준다. 표준국어대사전의 두 번째 의미와 기초한국어사전의 첫 번째 의미가 뜻풀이가 일치하기 때문에 자동으로 정렬된 예시를 보여준다. 이 때, 문장에서 "폭설"이 표준국어대사전의 두 번째 의미에 해당하는 경우, 기초다국어사전의 연결된 엔트리에서 해당하는 외국어 대역어를 추출하여 할당하게 된다. 즉 표준국어대사전 "폭설_02"에 대해서 영어 "heavy snow; high snowfall"을 대응시킬 수 있고, 다른 언어의 대역어도 기초한국어사전 다국어 버전의 정보를 이용하여 제시할 수 있다.

그림 4. 다국어 어휘 번역 지식 구축 예
Fig. 4. An example of multilingual word translation knowledge

3-4 어휘 의미단위 대역 문장 생성

그림 5는 한국어 문장의 각 형태소에 목적언어(영어)의 대역어를 할당한 어휘대역문장 생성 예시를 보여준다. "사장_15"에 "president"를, "교장_03"에 "principal"을, "사과하_02"에 "apologize"를 할당하였다. 이를 위해서는 형태소 의미별로 대역어를 할당한 다국어 번역 지식 사전을 이용한다. 그림 5에서 "표준국어대사전"의 단어의 의미별로 "한국어기초사전 다국어 버전"의 영어 대역어가 할당되어 있는 것을 볼 수 있다. 대역어 사전에 대역어가 명시되어 있지 않은 경우, 원언어의 형태소에 대역어를 할당하지 않는다. 마지막으로 "대역 문장 생성" 단계를 통해서 한국어 문장의 각 형태소에 목적언어(영어)의 대역어가 할당되어 있는 문장의 예시를 보여준다. "사장"은 "president", "교장"은 "principal", "선생님"은 "teacher"이 대역어로 병기되었다. 이 결과는 한국어 문장의 구조를 이해하지만, 단어의 의미에 익숙하지 않은 사람들이 한국어 문장의 의미를 쉽게 이해할 수 있는 장점이 있다.

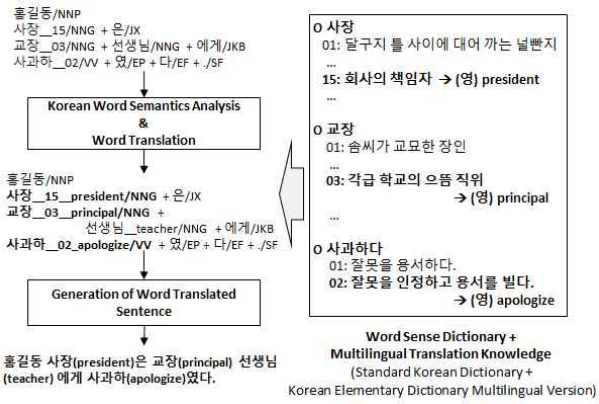

그림 5. 어휘대역 문장 생성 예
Fig. 5. An example of word translated sentence

3-5 가정통신문 도메인 어휘 확장

초기 어휘대역서비스는 특정 도메인을 가정하지 않고 개발하였다. 즉, 한국어 어휘의미분석 모듈, 다국어 어휘번역지식은 표준국어대사전과 한국어기초사전 다국어버전을 기준으로 개발 및 구축되었다. 그러나 가정통신문과 같은 특정 도메인의 문장에 대한 어휘 대역 서비스를 위해서는 해당 도메인만 나타나거나 고빈도로 나타나는 어휘에 대한 어휘의미분석 성능 개선과 대역지식확장이 필요하다. 따라서 본 연구에서는 가정통신문 도메인에 다국어 어휘대역 서비스를 적용하기 위하여 한국건강가정진흥원[3]에서 제공하는 "다문화가족을 위한 초등학교 용어 풀이집"에 포함된 어휘와 가정통신문 코퍼스를 구축한 뒤 고빈도 어휘를 추출하여 각각 한국어 어휘의미분석 시스템의 성능을 개선하였고, 다국어 어휘 번역 지식을 확장하였다. 기존의 표준국어대사전과 한국어기초사전에 포함되지 않은 용어와, 가정통신문 특성 상 복합명사를 분리하지 않고 자체적인 대역어가 필요한 어휘를 추가하였다. 표 2는 가정통신문 도메인 어휘를 1) 가정통신문 전문 용어, 2) 가정통신문 신조어, 3) 가정통신문 고빈도 질병 이름의 예시를 보여준다.

표 2. 가정통신문 도메인 어휘 예
Table 2. An example of words in school information domain

Terminology in School newsletters	교육감, 공개수업, 과학의날, 교육장배, 교육장배, 민방공, 생기부, 생활기록부, 신문고, 영재학급, 자율휴업일, 종업식, 학예회, 경시대회 등
Neologism in School newsletters	아람단, 돌봄교실, 안심알리미, 알리미, 재량휴업일, 아이돌보미, 열린공부방, 스쿨뱅킹, 또래샘, 예비중, 방과후학교, 선플 등
Disease names	인플루엔자, 일본뇌염, 백일해, 디프테리아, 척추측만증, 파상풍, 수족구병, 자궁경부암 등

3) http://www.kihf.or.kr

가정통신문 코퍼스 수집을 위하여 부산, 울산, 경남 지역에서 국제결혼 자녀들의 현황을 분석 후, 국제결혼 가정 및 자녀가 다수 분포하는 지역의 가정통신문을 수집하였다. 표 3은 지역별 국제결혼가정 자녀 현황 및 가정통신문 수집 건수를 보여준다. 국제결혼 자녀들이 많이 거주하는 지역의 가정통신문을 기반으로 어휘를 분석하고 서비스를 개선하면 다문화가정의 언어사용 특징을 잘 반영할 수 있다고 가정하였다.

표 3. 부산, 울산, 경남 지역 국제결혼가정 자녀 현황(일, 중, 동남아) 및 각 지역 별 가정통신문 수집 건수
Table 3. Statistics of multicultural families(Japan, China and Southeast Asian countries) in Busan, Ulsan, Gyeongsangnam-do districts, and number of collected school newsletters in each area

	Busan	Ulsan	Gueongsangnam-do
Total number of multicultural students	2,503	1,392	6,132
Dense area and number of multicultural students	Sasang-gu: 327	Ulju-gun: 350	Gimhae-si: 868 Changwon-si: 1,236
Number of school newsletter (total 1,081)	197	324	560

Ⅳ. 실험 및 평가

4-1 어휘대역서비스 웹서비스

어휘대역서비스 웹서비스(http://converge.bufs.ac.kr)를 개발하여 국내에 거주하는 결혼이주여성을 대상으로 서비스를 평가하였다. 그림 6은 어휘대역 웹서비스 화면을 보여주고 있으며 영어, 일본, 프랑스, 스페인, 아랍, 몽골, 베트남, 태국, 인도네시아, 러시아, 중국 11개 언어에 대한 어휘 대역서비스를 제공하고 있다.

그림 6. 어휘대역 웹서비스
Fig. 6. Web demo of word translation service

4-2 어휘대역서비스 사용자 평가

결혼이주 여성을 대상 어휘 대역 서비스의 성능을 평가하였다. 일본, 베트남 국가별 5명 (전체 10명) 결혼이주 여성을 대상으로 가정통신문 한국어 문장 100개에 대한 어휘대역 결과를 각각 제시하고 1~5점 부여하도록 하였다. 또한 성능 비교를 위하여 구글번역기의 결과도 제시하고 마찬가지로 1~5점을 부여하도록 하였다. 대역 및 번역 결과가 정확한 경우 5점부터 이해하지 못하는 경우 1점까지 점수를 순차적으로 부여하도록 요청하였다. 표 4는 한국어 문장에 대한 베트남어와 일본어 서비스 결과 및 구글 번역기 결과를 제시하고, 평가자가 점수를 부여한 예시이다.

표 5는 결혼이주 여성 대상 성능 평가의 결과를 제시하고 있다. 어휘대역 서비스의 경우 일본어, 베트남어 모두 4점 이상을 받아서 초급 한국어 이해 능력이 있는 사용자에게 유용함을 보였다. 한편 구글 번역기의 경우 문장 전체를 번역하여 제시하기 때문에 어휘 번역 오류, 문장 구조 번역 오류가 중복되어 사용자가 상대적으로 결과를 이해하기 어렵기 때문에 상대적으로 낮은 평가를 받았다. 또한 구글번역기의 결과에서 일본어의 경우 3.8점으로 상대적으로 베트남어의 2.4점보다 높은 결과를 보여준다. 즉 한국어, 일본어 등 주요 언어 사이의 번역 성능이 베트남어와 같은 동남아시아 언어 번역보다 성능이 우수함을 볼 수 있다. 이 결과는 한국어 대상 동남아시아 언어로 자동번역 시스템은 아직 성능이 낮기 때문에 현 상황에서는 어휘수준의 번역이 필요하다는 본 연구의 기본 가정이 옳음을 보여준다.

표 4. 어휘대역 서비스 사용자 평가
Table 4. User evaluation of word translation service

Input (Korean)	Service output (Vietnamese, Japanese)	Score
1,4학년 건강검진 병원 선정을 위한 선호도 조사	**[Word translation service: Vietnamese]** 1 , 4 학년(niênhọc, nămhọc) 건강(sựkhỏemạnh, sứckhỏe) 검진(việckhámbệnh) 병원(bệnhviện) 선정(sựtuyểnchọn) 을 위하(vì, để, cho) ㄴ 선호도(độưathích, mứcđộyêuthích,) 조사(sựđiềutra)	4
	[Google translation: Vietnamese] Sởthíchđểlựachọn 1,4 cấpbệnhviệnkiểmtrasứckhỏe	2
	[Word translation service: Vapanese] 1 , 4 학년(がくねん【学年】) 건강(けんこう【健康】) 검진(けんしん【検診】) 병원(びょういん【病院】) 선정(せんてい【選定】) 을 위하(ためだ【為だ】) ㄴ 선호도(せんこうど【選好度】) 조사(ちょうさ【調査】。とりしらべ【取り調べ・取調べ】)	4
	[Google translation: Japanese] 1,4年生健康診断病院選定のための評価の調査	4

표 5. 어휘대역 서비스 평가 결과 (평균점수)
Table 5. Evaluation result of word translation service (Avg. score)

	Word translation service	Google translation
Korean -> Japanese	4.1	3.8
Korean -> Vietnamese	4.6	2.4

4-3 어휘대역서비스 오류 분석

표 6은 어휘대역 서비스의 오류 유형을 제시하고 있다. 어휘대역서비스는 표준국어대사전의 한국어 어휘를 중심으로 개발되었기 때문에 최신 어휘 또는 고유명사를 하나의 단위로 인식하지 못하는 오류가 발생한다. 첫 번째 예에서 “아람단”을 하나의 고유명사 단위로 인식하지 못하고, “아람”, “단” 으로 분리하고, “단”에 대한 대역어만 제시하고 있다. 두 번째는 “인사”에 대한 의미분석 오류의 예시이다. 문장에서 “마주 대하거나 헤어질 때에 예를 표함. 또는 그런 말이나 행동”의 의미로 사용되었으나, 의미분석 과정에서 “사람의 일. 또는 사람으로서 해야 할 일”의 의미로 해석되어, 영어 대역어 “human resources affairs; personnel affairs”를 제시하고 있다. 세 번째는 대역어 오류의 예로서, “구체”에 대한 영어 대역어가 존재하지 않아서 제시하지 못하고 있다. 표준국어대사전 어휘는 51만여 개이지만, 기초한국어사전은 5만여 개의 어휘와 각 어휘별 대역어를 포함하고 있기 때문에 대역어를 제시하지 못하는 경우도 발생한다. 따라서 “가정통신문” 도메인의 용어를 대상으로 대역어 추가 작업을 추진할 계획이다.

표 6. 어휘대역 서비스 오류 예시
Table 6. Examples of errors in word translation service

Error types	Examples
Proper noun recognition error	<Input> 5월 중 ***아람단*** 활동 신청서 - 5 월(month || A unit used for counting months..) - 중 [A bound noun used to refer to something out of many.] - ***아람 단***(-dan || A suffix used to mean a group.) - 활동(activity; movement) - 신청서(application form)
Word semantics analysis error	<Input> 자녀들에게 바른 ***인사***습관을 갖게 합니다. - 자녀(child) 들(-deul || A suffix used to mean plural.) 에게 - 바르(straight; upright) ㄴ - ***인사***(human resources affairs; personnel affairs) - 습관(habit) 을 - 갖(have; hold) 게
Translation error	<Input> ***구체***일정 : 참가신청서 참조 - ***구체*** 일정(program; schedule) : - 참가(participation) 신청서(application form) - 참조(reference)

Ⅴ. 결 론

본 연구에서는 기초적인 한국어 능력을 가진 다문화 가정 구성원 및 외국 관광객을 위한 어휘대역서비스를 제안하고 가능성을 보였다. 향후 오류 검증, 시스템 성능 개선, 대역지식 확장 등을 포함하는 성능 개선을 위한 프로세스를 확립하여 비전문가도 쉽게 서비스를 운영할 수 있도록 할 예정이고, 스마트폰 앱으로 출시하여 보급할 계획이다. 또한 개체명 인식 기능 추가하여 인명, 지명, 조직명 등 번역이 필요없는 의미 단위를 인식하여 서비스를 개선할 예정이다.

감사의 글

이 논문은 2016년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NNRF-2016S1A5B6913773)

참고문헌

[1] papago (NHN translation service). Available: https://papago.naver.com/
[2] Google translation. Available: https://translate.google.co.kr
[3] S. J. Eo, "A study on Korean Language Teaching Method through Cooperative Learning for Multi-cultural Family - Focusing on International Marriage Migrant Women and their children," *International Society of Korean Language and Literature*, Vol. 5, pp.133-174, Apr. 2011
[4] H. R. Cho, "The Reality and Issues of Korean Language Education for Immigrants -Focused on Implementing Korean Language Education," Journal of Korean Language Education, Vol. 24, No. 1, pp. 237-268, 2013
[5] Y. K. Choi, *“Korean Language Education Policy in Multicultural Society,” Journal of Multiculture & Peace*, Vol. 5, No. 1, pp. 1-24, 2011
[6] D. J. Son, D. W. Chung, "Korean Language Education Learning Activities for Children of Multi-cultural Families," The Education of Korean Language, Vol.149, pp.279-305, 2015
[7] S. G. Park, J. E. Choi, "A Comparative Study of School Teachers’ Recognition upon Korean Education for the Students with Multicultural Family Background," *Studies of Korean & Chinese Humanities*, Vol. 45, pp. 81-109, 2014
[8] H. S. Cheon, G. S. Park, "A Study on the School Life of Children of Multicultural Families," *Journal of Contemporary Society and Multiculture*, Vol. 2, No. 2, pp. 416-444, Dec. 2012
[9] J. E. Noh, J. W. Park, “A Study on Performance-based

Korean Language Education with a consideration of parents role : For married female immigrants who have school age children,", *Journal of Multi-Cultural Contents Studies*, Vol. 15, pp.147-188, Oct. 2013

[10] S. M. Jung, "Communicational Approaching method to involve the children's education as the mothers of women immigrants for wedding," *Journal of Multi-Cultural Contents Studies* , pp. 115-134, April, 2010

[11] S. K. Choi, K. Y. Lee, Y. H. Roh, O. W. Kwon and Y. G. Kim, "Customization Method for Commercialization of a Pattern-based English-Korean Machine Translation System," *Journal of KIISE* :*Software and Applications*, Vol. 39, No. 4, pp. 253-260, April 2012

[12] M Johnson et. al., "Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation", *Transactions of the Association of Computational Linguistics*, Volume 5, Issue 1 , 2017

[13] H. D. Na, J. Li, J. H. Lee, "Integrating Bilingual Dictionary in Statistical Machine Translation between Korean and Japanese," *in Proceedings of KCC 2012*, Vol.39, No.1B, pp.288-290, Jun. 2012

[14] C. K. Lee, J. S. Kim, H. G. Lee and J. S. Lee, "English-to-Japanese Machine Translation using Neural Networks", *Communications of the Korean Institute of Information Scientists and Engineers*, Vol. 33, No. 10, 48-52, Oct. 2015.

[15] J. C. Shin, C. Y Ock, "A Stage Transition Model for Korean Part-of-Speech and Homograph Tagging", *Journal of KIISE :Software and Applications*, Vol. 39, No. 11, pp. 889-901, Nov. 2012

[16] H. Y. Lee, J. S. Lee, B. D. Kang, S. W. Yang, "Functional Expansion of Morphological Analyzer Based on Longest Phrase Matching For Efficient Korean Parsing", *Journal of Digital Contents Society*, Vol. 17, No. 3, pp.203-210, 2016.6

[17] J. C. Shin, C. Y Ock, "Improvement of Korean Homograph Disambiguation using Korean Lexical Semantic Network (UWordMap)", *Journal of KIISE*, Vol. 43, No. 1, pp.71-79, Jan. 2016

[18] A. S. Yoon, "Korean WordNet, KorLex 2.0 - A Language Resource for Semantic Processing and Knowledge Engineering", *HAN-GEUL*, Vol. 295, pp. 163-201, 2012.3.

류법모(Pum-Mo Ryu)

1995년 : 경북대학교 컴퓨터공학과(학사)
1997년 : POSTECH 컴퓨터공학과(공학석사-자연언어처리)
2009년 : 한국과학기술원 전산학과(공학박사-자연언어처리)

1996년~1999년 : 한국전자통신연구원 언어처리연구실 연구원
1999년~2004년 : ㈜K4M SW기술연구소 연구원
2009년~2015년 : 한국전자통신연구원 지식마이닝연구실 책임연구원
2015년~현 재 : 부산외국어대학교 동남아창의융합학부 언어처리창의융합전공 부교수
※관심분야 : 정보검색, 자연어처리, 온톨로지, 질의응답기술 등