

한국어-중국 조선어 비교 분석 및 변환 시스템 개발

윤광진^U 김동일
중국 연변과학기술대학 컴퓨터 전자통신공학부
kijoon@ybust.edu.cn

Comparative Analysis of Korean & Chinese Korean and Converting System Development

Kwang-Jin Yoon^U Dong-Il Kim
School of Computer Science, Electronic and Communication, Yanbian University
of Science & Technology, Jilin Province, China

요 약

한국어는 7,000만 우리 민족이 사용하는 언어로 세계의 많은 언어 가운데에서 사용인구로 볼 때 상위권에 속하며, 동아시아 지역의 사용 분포를 살펴보면 한반도와 중국 동북 3성으로 구분 할 수 있다. 지금까지 한국어와 중국 조선어의 비교 연구가 많이 진행되어 왔으나 객관적이고 과학적으로 어느 정도 차이가 있는지 알려져 있지 않다. 본 연구를 통하여 한국어와 중국 조선어의 대략적인 차이를 분석하고, 컴퓨터를 이용하여 한국어를 중국 조선어로 전환하며 지속적인 연구 분석을 위해 중국 조선어 기초 자료 구축 및 분석 시스템을 개발하고자 한다.

1. 서론

지구상에는 많은 종류의 언어가 있으며, 그 중 한국어는 7,000만 우리 민족이 사용하는 언어로 세계의 많은 언어 가운데 사용인구로 볼 때 상위권에 속한다. 또한 한반도뿐만 아니라 중국, 중앙아시아, 러시아, 일본, 미주, 유럽 등 여러 지역에 흩어져 있는 한민족이 사용하고 있다. 특히 중국의 길림성 연변지역에는 조선족 자치주로서 우리말을 공용어로 사용하고 있다.

하나의 언어가 국경을 사이에 두고 서로 다른 나라에서 쓰이게 될 때 서로 다른 사회 환경과 역사 배경으로 말미암아 많은 변화를 가져온다. 사회의 발전 변화에 가장 민감하게 반영하는 어휘층에는 그와 같은 국제간 변화가 많이 나타날 수 있다.

한반도를 제외하고 한민족이 가장 많이 살고 있는 곳이 중국의 동북 3성이다. 이 곳에서 사용하고 있는 한국어를 한국을 중심으로 비교해 볼 때 어음, 문법에서는 큰 차이가 없으나 어휘 사용에서는 정치, 경제, 문화의 차이로 인하여 어음이나, 문법에 비하여 차이가 많다.

이 논문은 2001년도 한국학술진흥재단의 해외한국학진흥사업 연구비에 의하여 지원되었음(과제번호 00-C-10)

한국과 중국 사이에는 근 반세기 동안 왕래가 없었다. 중국 경내에 거주하고 있는 조선민족과 한국에서 살고 있는 한민족은 같은 민족으로 공동한 언어를 사용하여 왔다. 이데올로기로 남북이 분단된 이후 같은 민족이지만 서로 왕래가 없었기 때문에 언어 사용에서 여러 가지 차이가 나타났다. 이는 언어 발전의 객관 법칙에 맞는 현상이다. 언어의 3요소인 어음, 어휘, 문법 가운데에서 어휘가 사회 발전과 밀접한 관계를 가지기 때문에 어음이나 문법에 비해 가변성이 많다.

한국어와 중국 조선어 사이에 나타난 차이에서도 어휘 사용에서 나타난 차이가 기본적인 것으로 되고 있다. 이로 말미암아 중국 조선민족들이 한국 영화나 TV 및 책을 보고 잘 이해하지 못하고, 한민족이 중국 조선어로 출판된 책을 보고 잘 이해하지 못한다. 이러한 차이는 우리민족 내부의 의사소통에 영향을 주고 있을 뿐 아니라 우리말의 국제 표준화에도 큰 영향을 주고 있다. 이는 우리 민족 자체의 발전은 더 말할 것도 없고, 동북아 지역의 발전에도 불리하다. 때문에 우리말의 통일된 표준 제정이 밀어붙일 수 없는 시대적 과업으로 우리 눈앞에 나타나고 있다.

21세기는 화합의 시대, 공동 발전의 시대로 될 것이다. 이를 위해서는 서로 간의 교류에 지장을 주는 언어

사용에서 나타난 차이를 해소하는 것이 급선무이다. 차이를 해소하려면 먼저 어떤 차이가 있는가를 알아야 하고 그 다음 이러한 차이를 해소시킬 방법을 연구해야 한다. 본 논문은 바로 이러한 목적에서 쓰여졌다. 우리가 언어 사용에서 나타나는 차이를 안다면 서로 간의 이해와 의사 소통이 이루어질 수 있는 방안들은 고안해 낼 수 있을 것이다. 참고적으로 중국에서 사용되고 있는 한글 코드는 중국 표준 부호계(GB 12052-89)에 확정되어 사용하고 있다.

2. 중국 조선어의 변천

중국 조선민족은 지역적 특징을 갖고 있다. 길림성 연변 지역과 흑룡강성의 목단강 지역은 함경도 지역의 주민이 주류를 이루며, 요녕성의 대부분 지역은 평안도 지역 주민, 그리고 흑룡강성의 대부분 지역은 경상도 지역의 주민이 이주해와 정착을 하였다. 이와 같은 지역적 특징으로 각 지역의 어휘는 마치 한반도의 각 지방 방언을 그 데로 사용하는 듯 한 특징들을 갖고 있다.

이 전에는 중국과 한국의 교류가 없는 데다 중국의 조선어는 평양을 기준으로 해야 한다는 전 주은래 총리의 지시에 의하여 방송과 같은 공식적인 언어는 북한 조선어의 언어 규범을 사용했다. 즉 1969년 중국에서 자체로 띄어쓰기 규정을 제정하였으나 북한의 1966년도 "조선말 규범집"에 준하여 만들어 졌다. 그러다가 중국의 문화 대혁명이 일어나 중국과 북한의 관계가 소원해지면서 중국의 조선말 사용은 혼란을 가져왔다. 그 결과 각 지역별 방언이 강하게 사용되었고 대량으로 중국말을 그 데로 옮겨 사용되었다. 문화 혁명이 끝나고 개혁개방이 실시된 후 초기의 북한 조선어를 기준으로 하되 기계적으로 모방할 것이 아니라 중국 조선족 자체의 특징과 수요에 쫓아 개선되어야 한다는 강조에 의하여 1977년 8월에 처음으로 조선어 철자법, 조선어 띄어쓰기, 조선어 표준 발음법, 문장 부호법을 제정하였다. 그러나 이 때에도 철자법, 발음법 등은 북한과 일치시키는 것을 전제로 하였다. 1978년 12월에 와서 정식으로 "조선말 명사, 술어 규범화 원칙"을 제정하였다. 전체적으로 중국에서 조선어 규범화는 철자법과 발음법 면에서 북한과의 일치성을 기하며 어휘 면에서는 차이성을 인정하면서도 공통성을 증가하는 원칙에서 진행되었다. 이 후 한국과 중국의 수교가 이루어진 후 중국 조선족의 언어환경에 새로운 변화가 일어났다. 즉 한국과의 교류가 빈번해지자 중국에서 한국어가 중시를 받게 되어 많은 부분에서 한국어가 사용되고 있다.

3. 한국어와 중국 조선어의 차이

다민족, 다문화, 다언어 국가인 중국에서 거주하는 조선 민족은 중국의 사회환경과 언어 영향에 의하여 한국에서 볼 수 없는 어휘들을 사용하기도 하고 변형된 어휘를 사용하기도 한다. 이와 같은 어휘는 조선족의 일상

생활에서 잘 쓰이는 까닭에 중국 조선말에서 무시 할 수 없는 변화로 되어 있다.

3.1 한국에서 볼 수 없는 어휘

한국에서 볼 수 없는 어휘는 중국에서 인구의 대다수가 쓰고 있는 한어의 영향을 많이 받았으며, 새로운 말을 형성하는 데 있어 수용 방식 면에서 조선어로서의 특성이 나타난다. 이에 크게 다음과 같이 4가지로 구분할 수 있다.

- 한어 어휘를 조선말 한자어 음독법으로 음독한 것
自治區 -> 자치구
- 반 음독 반 의역한 것
五星紅旗 -> 오성홍기
- 의역한 것
傅家玉 -> 대물림 보배
- 음차한 것
秧歌 -> 양걸(춤)

3.2 변형된 어휘

변형된 어휘는 동일한 의미에 대한 단어 자체의 변형과 중국 조선어 자체의 문법 규칙이 생성되면서 한국어와의 차이를 보이는 부분으로 나눌 수 있다. 특히 맞춤법 총칙부분에 보면 다음과 같은 차이가 있음을 알 수 있다.

- 한국어
한글맞춤법: 표준어를 소리대로 적되, 어법에 맞도록 함을 원칙으로 한다.
- 중국 조선어
조선말맞춤법: 단어에서 뜻을 가지는 매개 부분을 언제나 같게 적는 원칙을 기본으로 하면서 일부 경우 소리나는 대로 적거나 관습을 따르는 것을 허용한다.

문법 규칙의 몇 가지 차이를 예를 들면 다음과 같다.

- ① 형태소에서 -르거나, -르걸, -르개, -르세, -르수룩, -르지, -르진대 의 경우 된소리로 발음되더라도 된소리로 적지 않는다.
-르까, -르쏘냐, -르꼬 => -르가, -르소냐, -르꼬
- ② 사잇소리 표기
깃발 -> 기발
깨끗 -> 깨이
- ③ 부사에서 접미사 "이", "히"의 경우
가득이 -> 가득히
그옥이 -> 그옥히

4. 한국어와 중국 조선어의 차이 분석

한국과 중국의 어휘 상이성 연구를 위하여 대량의 말뭉치를 이용하여 분석하고, 분석한 데이터를 이용하여 한국어 중국 조선어 대조 사전을 구축한다. 여기서 사용된 말뭉치는 "21세기 세종계획"국어 기초자료의 현대 국어 균형 말뭉치 1000만 어절을 기본으로 빈도수 분석에 의하여 빈도수가 높은 구어체 20,000어와 문어체 50,000어를 기준으로 분석을 하였다.

4.1 구어체 자료

선택된 구어체 단어 20,000어를 중등 이상의 문화를 소유한 대상자 십 여명을 선정하여 모르거나 이해하기 어려운 단어를 선택해 내게 했다. 이들이 선택해낸 단어 가운데에서 비교적 보편적으로 모르는 단어를 한국어와 중국 조선어의 단어 사용에서 차이가 나는 단어로 인정하였다.

조사 결과 중국 조선어에서 잘 쓰이지 않는 단어는 약 6900여개로 조사한 단어 총수의 34.5%를 차지한다. 6900여 개의 단어는 다음과 같이 분류된다.

내용	갯수	백분율
표기가 다른 단어	900여개	약13%
고유명사	100개 좌우	약1.4%
외래어	1600개 좌우	약23%
일반 어휘	4290여개	약62%

<표1. 구어체 분석 결과>

4.2 문어체 자료

중국에서 중국 조선어 교육을 받은 대학 이상 문화 수준의 지식인 15명을 선택하여 선정된 단어 5만개 가운데에서 표기가 다른 단어와 이해하기 어려운 단어를 조사하여 표시하게 하였다. 그 가운데에서 반수 이상이 표시해 놓은 단어를 한국어와 중국조선어 사이의 어휘 사용에서 차이가 있는 단어로 인정하였다.

조사 결과 도합 18459개 단어가 표기가 다르거나 중국조선어에서는 잘 쓰이지 않는 단어로 집계되었다. 이는 전체 단어 5만 개 가운데에서 근 37%를 차지한다. 여기에서 한국어와 중국 조선어 사이에서 나타나는 단어 사용 차이의 주류는 외래어와 비 상용어 단어이며 비상용 단어의 대다수가 한자어이다. 이를 세분하면 다음과 같다.

내용	갯수	백분율
표기가 다른 단어	2290	약4.58%
비규범어	201	약0.004%
한국문화와 관련된 단어	205	약0.004%
학술용어와 낱은 어휘	2483	약4.97%
비 상용 어휘	6084	약12.17%
사전에 없는 단어	3472	약6.94%
외래어	3724	약7.45%

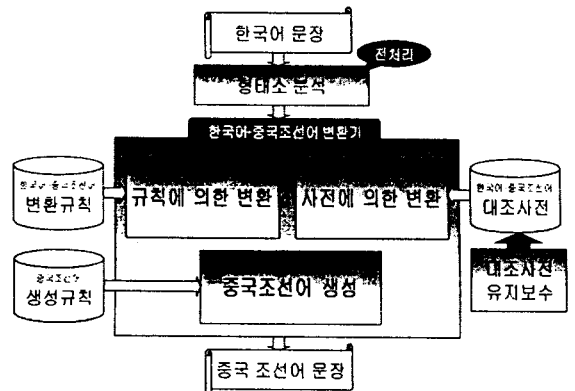
<표2. 문어체 분석 결과>

5. 시스템 개발

중국 조선민족들이 한국 영화나 TV 및 책을 보고 잘 이해하지 못하고, 한국 국민이 중국 조선어로 출판한 책을 보고 잘 이해하지 못한다. 이와 같은 필요에 의해 먼저 중국에 있는 조선민족에게 좀더 정확히 한국의 문서화된 자료를 이해시키기 위하여 한국어를 중국 조선어로 자동 변환하는 시스템을 설계하고 구현했다.

즉 입력되는 한국어 문장을 한국어 형태소 분석기를 통하여 최소 의미 단위인 형태소 단위로 분리하고 분리된 결과로부터 중국 조선어와의 차이를 분석하여 변환하고 최종적으로 중국 조선어로 생성하게 된다.

전체 시스템 구성도는 다음과 같다.



<그림1 전체 시스템 구성도 >

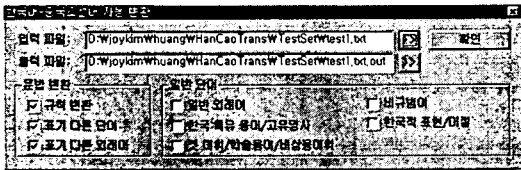
위 그림에서 보듯 변환 방법은 크게 규칙에 의한 변환과 사전에 의한 변환으로 구분된다.

사전에 의한 변환은 한국어와 중국어 대조 사전을 이용하여 사전 검색을 통해 변환을 완성하며, 규칙에 의한 변환은 사전 변환으로 변환시키지 못하는 부분은 주로 한국어와 중국 조선어의 문법 규칙 차이로 말미암아 발생하는 부분으로 일정한 문법 규칙을 이용하여 변환하는 부분이다.

사전에 의한 변환을 위하여 앞에서 조사한 구어체와 문어체의 한국어와 중국 조선어의 단어들을 기반으로 각 분류별로 6개의 전자사전을 구성하였다.

- 표기가 다른 단어 사전
- 표기가 다른 외래어 사전
- 일반 외래어 사전
- 한국 문화용어, 고유명사 사전
- 학술용어, 옛 어휘, 비 상용어휘 사전
- 비 규범어 사전

변환시스템의 구현된 GUI는 다음과 같다.



< 그림 2. 변환기 GUI >

변환 결과는 다음과 같다.

입력 : 태어나서
 변환 : 태어나/VV+아서/EFC
 생성 : 태어나서

입력 : 노동은
 변환 : 로동/NNCV+은/PX
 생성 : 로동은

입력 : 아래동네에서
 변환 : 아래동/NNCG+네/XSNN+에서/PS
 생성 : 아래동네에서

안변역에서 동해 북부선에 연결된 기차를 타고 다시 남쪽으로 달리면 차창 서편은 울창한 삼림에 덮인 태백산맥이, 동편으로는 쪽빛 동해가 무한하게 펼쳐진다. 아버님은 나를 일등 농사꾼으로 키워낼 생각이셨다. 나의 노동은 열 살 무렵부터 시작되었다. 그후 오랜 세월이 흘러 처음으로 사업차 미국에 갈 일이 생겨 적십자 병원에서 신체검사를 받았는데, 의사가 폐병을 앓은 일이 있느냐고 물었다. 나는 물론 그런 병은 앓은 적이 없다고 대답했더니 의사가 폐를 앓은 흔적이 있는 엑스 레이 필름을 보여주었다.

안변역에서 동해 북부선에 연결된 기차를 타고 다시 남쪽으로 달리면 차창 서편은 울창한 삼림에 덮인 태백산맥이, 동편으로는 쪽빛 동해가 무한하게 펼쳐진다. 아버님은 날 일등 농사꾼으로 키워낼 생각이셨다. 내 노동은 열 살 무렵부터 시작되었다. 그후 오랜 세월이 흘러 처음으로 사업차 미국에 갈 일이 생기여 적십자 병원에서 신체검사를 받았는데, 의사가 폐병을 앓은 일이 있느냐고 물었다. 난 물론 그런 병은 앓은 적이 없다고 대답했더니 의사가 폐를 앓은 흔적이 있는 엑스 레이 필름을 보여주었다.

5. 결론

가장 많은 재외 동포가 거주하고 있는 중국에서 중국 동포가 사용하는 중국 조선어는 중국과 한국의 관계 개선으로 많은 부분에서 한국의 어휘를 따르고 있으나 아직 많은 부분의 차이는 그대로 유지되고 있다. 이러한 차이는 우리민족 내부의 의사소통에 영향을 주고 있을 뿐 아니라 우리말의 국제 표준화에도 큰 영향을 주고 있다. 이는 우리 민족 자체의 발전은 더 말할 것도 없고, 동북아 지역의 발전에도 불리하다. 이제는 우리말의 통일된 표준 제정을 시급히 해결해야 할 때가 왔다. 앞으로 좀더 많은 세계 각 지역의 한국어 비교 연구 및 자료 분석으로 기초 자료를 마련한다면 이와 같은 자료를 기본으로 한국어 통일화에 이용 할 수 있으리라 기대된다. 또한 한국어와 중국 조선어의 완벽한 변환을 위하여 사전의 지속적인 Update와 문법적 차이에 대한 변환을 좀더 세분화하여 구현한다면 더욱 좋은 성능의 변환기를 만들 수 있으리라 기대되며 이와 같은 변환기를 인터넷 상의 실시간 변환기로 향상시킨다면 중국에 있는 조선 동포에게 좀더 정확히 모국을 알릴 수 있을 것이다.

6. 참고 문헌

- [1]. 최윤갑, "중국, 조선, 한국에서의 조선어 차이에 대한 연구", 연변인민출판사, 1994
- [2]. 21세기 세종계획 국어 기초자료, 1998
- [3]. 임해창, "한글자동색인을 위한 기초도구 구축 연구", 1995
- [4]. 중앙민족대학 조선연구소, "조선학", 민족 출판사, 1999
- [5]. 장흥권, "조선어, 한어, 일본어 현대어휘와 그 변이에 대한 대비연구", 민족 출판사, 1999
- [6]. 윤광진 외, "대량의 말뭉치를 이용한 한국어-중국조선어 자동 전환 시스템", 세계속의 조선어(한국어)대비연구 국제학술회의, 2000