

남북 언어 동질성 회복을 위한

제3차 남북 국제 학술 회의 논문집

-민족어 유산의 수집 정리와 고유어 체계의 발전-

담당 연구원: 이승재(어문규범연구부)

민족어 유산의 수집 정리와 고유어 체계의 발전 연구회에 관한 논문
◆ 중국 베이징 · 21세기 호텔 ◆ 2004. 6.22. ~ 2004. 6.25.



국립국어연구원

남북 언어 동질성 회복을 위한

제3차 남북 국제 학술 회의 논문집

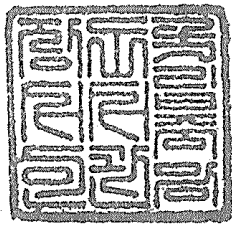
-민족어 유산의 수집 정리와 고유어 체계의 발전-

담당 연구원: 이승재(어문규범연구부)

주 최 : 국제 고려학회 아시아분회 및 평양지부
조선 사회과학원 언어학연구소
한국 국립국어연구원



국립국어연구원



3374310

머 리 말

국립국어연구원은 2001년 12월과 2003년 11월에 중국 베이징시에서 중국의 학자들과 북쪽의 사회과학원 언어학 연구소의 학자들을 만나 국어 순화를 비롯한 말뭉치 구축, 방언 조사 연구에 관한 발표와 토론을 벌인 적이 있습니다. 이것이 인연이 되어 2004년에도 국제 고려학회 아시아분회의 도움으로 중국 베이징시에서 말뭉치 구축과 방언 조사 연구에 관한 발표와 토론을 벌였습니다. 예전에는 무엇을 토론할까 하는 문제로 고민을 했었지만 이제는 어떻게 토론할까 하는 문제를 놓고 고민을 했습니다. 그 이전의 발표와 토론이 원론적인 수준의 것이었다면 이번 회의의 발표와 토론은 좀 더 각론에 근접한 수준의 것이었다고 할 수 있습니다.

이번 학술 회의는 2001년과 2003년 학술 회의의 맥을 잇는 제3차 회의이기 때문에 남북 언어 교류의 디딤돌 역할을 충분히 했다고 생각합니다. 이번 학술 회의에서 토론의 주요 주제로 삼았던 방언은 우리 민족 언어의 뿌리를 이루고 있는 요소이기 때문에 이를 바탕으로 하여 남북 언어의 교류가 자연스럽게 이루어질 수 있을 것이라고 생각합니다.

이번 학술 회의에서는 방언 조사 연구와 말뭉치의 구축과 활용에 관한 2개의 분과가 구성되어 발표와 토론이 진행되었는데 이 논문집은 이번 학술 회의에서 발표된 논문들을 발표 순서에 따라 정리한 것입니다. 이번 학술 회의가 무사히 끝날 수 있도록 귀중한 논문을 보내 주시고 귀중한 시간을 내어 회의에 참석해 주신 여러 선생님들께 다시 한 번 감사하다는 말씀을 드립니다. 모쪼록 이 논문집이 남북 학술 교류의 중요한 기초 자료로 활용되기를 바랍니다.

2004년 9월 1일

국립국어연구원 원장 

‘제3차 남북 국제 학술 회의 논문집’을 내면서

2003년도에 제2차 남북 국제 학술 회의에 참석한 후 2004년 6월에 다시 제3차 남북 국제 학술 회의에 참석할 수 있었습니다. 이번 학술 회의에서는 그동안 논의되었던 사항들을 다시 한 번 확인하고 말뭉치와 방언을 중심으로 우리 민족의 문화 유산을 효과적으로 구축할 수 있는 방안을 진지하게 고민하였습니다.

제1차 회의 때에는 서로의 만남을 기뻐하며 우리 민족의 문화 유산에 대한 토론을 하였는데 이번 회의에서는 지속적으로 다시 만날 수 있다는 굳은 믿음 아래 우리 민족의 문화 유산 정리 사업에 대한 실천적 방안을 깊이 있게 논의하였습니다. 발표가 진행되는 동안 많은 참석자들은 서로의 발표 내용이 일맥상통하고 있음을 느꼈고 발표 내용을 구체적으로 실천에 옮기는 것도 그리 어려운 일이 아니라는 것을 믿게 되었습니다.

이 논문집에는 남과 북 그리고 중국의 글이 섞여 있습니다. 이 논문집에서는 남쪽의 글은 남쪽의 언어 규범대로, 북쪽과 중국의 글은 북쪽의 언어 규범대로 표기를 하였고 교정 과정에서도 남과 북의 언어 규범에 따라 별도로 교정을 하고 일관성이 결여된 듯이 보이는 부분만 수정을 하였습니다. 이 논문집이 나올 수 있도록 애써 주신 남과 북, 그리고 중국의 여러 선생님들께 감사하다는 말씀을 드립니다.

2004년 9월 1일

담당 연구원 이 승 재

2004년 국제 학술 회의 참가자 명단

지 역	소 속	직 위(직 급)	이 름
중 국	국제 고려학회 상임 고문	명예 교수	최 응 구
	베이징대학교 조선문화연구소	소 장	이 선 한
	연변대학교	교 수	전 학 석
	베이징 언어대학	교 수	최 순 희
	베이징대학교 조선문화연구소	연구원	량성, 김지혜, 김아희
북 측	사회과학원 언어학 연구소	소 장	문 영 호
	사회과학원 대외사업처	처 장	곽 선 욱
	사회과학원 언어학 연구소	실 장	권 종 성
	사회과학원 언어학 연구소	실 장	리 기 만
	사회과학원 언어학 연구소	실 장	홍 석 희
	사회과학원 언어학 연구소	연구사	김 성 근
	사회과학원 언어학 연구소	연구사	백 운 혁
	사회과학원 언어학 연구소	연구사	오 철 만
	사회과학원 언어학 연구소	연구사	김 철 호
	사회과학원 과학지도국	부 원	김 명 성
남 측	국립국어연구원	원 장	남 기 심
	국립국어연구원 어문규범연구부	부 장	김 하 수
	국립국어연구원 어문규범연구부	학예연구관	이 승 재
	국립국어연구원 어문자료연구부	학예연구관	박 민 규
	서울대학교 언어학과	교 수	권 재 일
	전주대학교 국어교육학과	교 수	소 강 춘
	문화관광부 국어정책과	과 장	유 병 한
	경북대학교 국어국문학과	교 수	이 상 규
	서울대학교 국어국문학과	교 수	최 명 욱
	연세대학교 국어국문학과	교 수	홍 윤 표

2004년 국제 학술 회의 일정표

(1) 6월 22일(화)

시 간	내 용
18:00~18:30	상견례(대표 인사 및 참석자 소개)
	▶사회: 최순희(언어대학 교수) ▶장소: 2층 明苑餐廳
18:30~21:00	환 영 만찬(2층 明苑餐廳)

(2) 6월 23일(수)

시 간	구 분	내 용		
		이 름	소 속	제 목
09:00~09:30	학술 대회 개회식			
	▶사회: 이선한(베이징대학교 교수)			
	인사말	최응구	베이징대학교 명예 교수, 국제 고려학회 상임 고문	
		문영호	사회과학원 언어학연구소 소장	
		남기심	국립국어연구원 원장	
09:30~10:10	기조 연설			
	▶사회: 최순희(언어대학 교수)			
09:30~10:10	기조 연설	문영호	사회과학원 언어학연구소 소장	민족어 유산의 발굴 정리
		남기심	국립국어연구원 원장	문화적 자산으로서의 우리말 어휘
		전학석	연변대학교 교수	중국 지역의 조선어 자료 수집
10:10~10:20	♠ 휴 식			

시 간	구 분	내 용		
		이 름	소 속	제 목
10:20~11:50	제1분과 민족 방언의 공동 연구			
	▶사회: 권재일 권중성			
10:20~10:50	발표	김하수	국립국어연구원 어문규범연구부장	방언과 사회 발달
		리기만	사회과학원 언어학연구소 실장	황해도방언에서 토의 쓰임에 대하여
	▷질의응답			
	▶사회: 홍운표 홍석희			
10:50~11:50	발표	최명옥	서울대학교 교수	지역어 조사의 원칙과 조사 방법
		김성근	사회과학원 언어학 연구소 연구원	민족어유산으로서의 조선어방언의 조사원칙과 방법
	▷질의응답			
12:00~13:00	점 심 식 사(1층 咖啡廳(커피숍))			
	▶사회: 이상규 리기만			
13:30~14:40	발표	오철만	사회과학원 언어학 연구소 연구원	방언조사 어휘선정에서 제기되는 문제
		이상규	경북대학교 교수	지역어 조사 항목의 설정 원칙과 방향
		홍석희	사회과학원 언어학 연구소 연구원	동북방언의 어음론적특성에 대하여
		<휴식>	14:40~15:00	
		백운혁	사회과학원 언어학 연구소 연구원	서북방언의 지리언어학적 특성과 방 언 조사에서 제기되는 문제에 대하여
15:00~15:40		홍운표	연세대학교 교수	민족어 방언 검색 시스템 개발
	▷질의응답			
15:40~16:00	♠ 휴 식			

시 간	구 분	내 용		
		이 름	소 속	제 목
16:00~17:50	제2분과 : 말뭉치의 구축과 활용			
	▶사회: 김성근 김하수			
16:00~16:30	발표	권종성	사회과학원 언어학연구소 실장	균형 코퍼스에 대한 새로운 이해와 그 제작에서 나서는 몇가지 문제
		소강춘	전주대학교 교수	민족어 말뭉치 지도의 작성
	▷질의응답			
16:30~17:00	♠ 휴 식			
17:00~17:50	발표	권재일	서울대학교 교수	말뭉치 자료 정보 부착과 남북 문법 통합 방안
		김철호	사회과학원 언어학 연구소 연구원	코퍼스를 리용한 다의성처리방법에 대하여
		이승재	국립국어연구원 학예연구관	말뭉치 활용 도구의 개발
	▷질의응답			
18:00~20:00	분과별 만찬 및 방언 문제 협의(류경식당)			

(3) 6월 24일(목)

시 간	구 분	내		용
		이 름	소 속	제 목
09:30~11:50	분과별 종합 토론 1			
12:00~13:00	점 심 식 사(1층 咖啡廳(커피숍))			
13:00~15:30	분과별 종합 토론 2			
	▶ 방언 분과: 발표자			
	▶ 말뭉치 분과: 발표자			
15:30~15:50	♠ 휴 식			
15:50~17:30	전체 종합 토론			
	▶ 사회: 김하수 김성근			
	▷ 질의응답			
18:00~20:00	저녁 식사(1층 咖啡廳(커피숍))			

(3) 6월 25일(금)

시 간	구 분	내		용
		이 름	소 속	제 목
09:00~10:00	회의 내용과 협의 사항 정리			
10:00~17:30	참관(韓村河)			
17:30~18:00	회의 성과 정리, 발표 및 폐회 선언			
18:00~20:00	송 별 만 찬(평양 모란관)			

(4) 6월 26일(토)

- 참가자 귀국

차 례

머 리 말	i
-------------	---

〈기조 연설〉

민족어유산을 발굴정리하는것은 언어분야에서 민족성을 고수하기 위한 중요한 방도 / 문영호(사회과학원 언어학연구소 소장)	3
문화적 자산으로서의 우리말 어휘 / 남기심(국립국어연구원장)	7
중국에서의 조선어 자료 수집에 대하여 / 전학석(연변대)	10

〈제1분과 민족 방언의 공동 연구〉

방언과 사회 발전 / 김하수(국립국어연구원)	27
황해도방언에서 토의 쓰임에 대하여 / 리기만(사회과학원)	32
지역어의 조사 원칙과 조사 방법 / 최명옥(서울대학교)	38
민족어유산으로서의 조선어방언의 조사원칙과 방법 / 김성근(사회과학원)	57
방언조사어휘선정에서 제기되는 문제 / 오철만(사회과학원)	63
방언 조사 항목 선정의 원칙과 방향 / 이상규(경북대학교)	71
동북방언의 어음론적특성에 대하여 / 홍석희(사회과학원)	93
서북방언(평안남도방언)의 지리언어학적특성과 방언조사에서 제기되는 문제에 대하여 / 백운혁(사회과학원)	100
민족어 방언 검색 시스템 개발 / 홍운표(연세대학교)	107

〈제2분과 말뭉치의 구축과 활용〉

균형코퍼스에 대한 새로운 이해와 그의 제작에서 나서는 몇가지 문제 / 권종성(사회과학원)	133
민족어 말뭉치 지도 작성 / 소강춘(전주대학교)	138
말뭉치 자료 정보 부착과 남북 문법의 통합 방안 / 권재일(서울대학교)	155
코퍼스를 리용한 다의성처리방법에 대하여 / 김철호(사회과학원)	171
말뭉치 활용 도구의 개발 / 이승재(국립국어연구원)	176
2004년 학술 회의 관련 사진	193

기조연설

빈 면

민족어유산을 발굴정리하는것은 언어분야에서 민족성을 고수하기 위한 중요한 방도

문 영 호(사회과학원 언어학연구소 소장)

현시기 언어분야에서 민족성을 고수하는 문제는 민족어의 고유성과 우수성을 지켜 나가는 근본원칙으로 될뿐아니라 민족의 자주성과 존엄을 높이고 북남사이의 언어적 차이를 줄이는 기본방도로 된다.

우리 말과 글의 민족성을 특징짓는 가장 중요한 언어수단은 우리 겨레가 오랜 옛날 부터 창조발전시켜온 민족어유산에 집중적으로 반영되어있다.

일반적으로 민족어유산이라고 하면 오랜 역사적기간에 창조 계승하여온 인민대중의 언어문화적재부를 가리킬수 있으나 우리는 여기에서 주로 묻혀있는 입말원천을 중심에 놓고 론하게 된다. 다시말하여 서사어로는 정리되지 않아 문헌에서도 잘 쓰이지 않거나 국어사전에도 등록되지 않은 입말형태의 좋은 고유어원천을 좁은 의미에서 민족어유산이라고 하기로 한다.

그러한 민족어유산가운데는 방언, 민속어휘, 동식물이명과 속칭, 현장용어, 사회계층별생활어, 속어 등 여러 분야의 각이한 지역과 계층에 분포되어있는 입말고유어들이다.

인민들속에서 쓰거나 묻혀있는 입말형태의 고유어원천은 그 량과 표현성에서 대단히 풍부하고 귀중한 언어유산을 이룬다.

위대한 령도자 김정일동지께서는 다음과 같이 지적하시였다.

《민족어는 원래 입말로부터 생겨났으며 입말에 기초하여 글말이 발전하였다.…우리 인민이 오랜 역사를 통하여 창조하고 다듬어온 입말은 우리 문학의 언어를 기를지

게 하는 무진장한 원천이다.》(《주체문학론》, 216페이지)

입말을 기본으로 하는 민족어유산은 지역과 사회분야, 계층별로 볼 때 그 폭과 깊이가 대단히 방대한 고유어원천을 이루고있으며 그 표현성과 문화성에 있어서 매우 섬세하고 복잡한 양상을 이루고 있다.

실례로 동식물이명, 속칭만 보아도 시대별, 지역별, 용도별에 따라 수많은 인민적언어수단이 창조되어 쓰이고있지만 그것들은 다 동식물학명이나 전문화된 표준술어에 눌리워 생활속에 묻히여있다.

식물 특히 풀, 꽃, 나무 등의 경우에만 보아도 지역과 계층에따라 달라 쓰는 이명들이 많을뿐아니라 학술적이름외에 용도(약초, 식용, 관상용)와 생태학적특이성에 따라 부르는 속칭들이 대단히 많다. 이러한 수많은 입말어휘들의 대부분은 인민대중의 창조적지혜와 생활정서가 숨배여있는 귀중한 언어유산이다. 물론 지금까지 나온 《식물지》들과 식물학전문도서들에서 식물이명, 속칭들을 수집정리해놓은것들이 있지만 그것은 극히 학술적측면에서 대표적인것들에 지나지 않을뿐 실지 인민들속에서 감정정서를 담아 쓰는것들은 수집되지도 않았고 생활적으로 쓰이는 고유한 의미색채도 반영할수 없었다.

흔히 《질경이》를 《길장구》라고 하는 방언이명이 대표적으로 등록되고있지만 이 식물을 들나물(식용)로 쓸 때 인민들속에서 쓰는 《뱀조개》, 《갯조개》들은 숨겨져있으며 《고사리》를 방언에서 《끼사리》로 쓰는 것은 알려져있어도 《밝은쟁이》, 《붉은쟁이》 들은 수집되지 않고 있다. 현실깊이 들어가면 《뱀조개》, 《밝은쟁이》라는 말을 써야 들나물로서의 제맛이 나지 《질경이》, 《고사리》로서는 그러한 정서를 느끼지 못한다.

동식물이명, 속칭만 하여도 바다가와 산골깊이 뒤편 뒤편 감칠맛있고 향취나는 말들이 묻혀있고 그것들이 세대교체로 점점 소극화, 사멸화되어감을 알수 있다.

이러한 관점에서 우리 인민이 창조한 귀중한 언어유산이 더 소실되기전에 수집정리하기 위한 《민족어유산의 발굴정리》라는 하나의 큰 언어자료조사사업은 다음과 같은 분야의 어휘발굴조사를 진행하여야 한다.

1. 방언조사: 지금까지 조사발표된 자료를 기초로 하면서도 출발점에서부터 그것을 무시하고 전국 9개도의 방언을 전면 재조사정리하는것이다. 여기에서 제기되는 기본문제는 조사방법이다.
2. 민속어휘: 이미 자취를 감추었거나 흔적이 남아있는 조상전래의 의상과 의례단장, 민족음식, 요리와 주방가구, 로동도와 기물, 민속유희와 명절풍습, 관혼상제 등에 쓰인 물명과 어휘들을 전 지역별로 찾아내는 작업이다.
3. 동물이명, 속칭: 산짐승, 들짐승, 조류, 물고기, 곤충 등에 대하여 시대별, 지역별, 용도별에 따라 쓰인 여러가지 이름들을 찾아내는 일이다.
4. 식물이명, 속칭: 나무, 풀, 꽃, 산나물, 바다나물, 약초, 버섯 등에 대한 이명, 속칭을 발굴수집하는 일이다.
5. 현장용어: 탄광, 광산, 소금밭, 벌목장, 류벌장, 사금밭, 양어양식장과 같은 특정의 작업현장에서 학술적범위밖에서 오래전부터 로동자, 기술자들이 쓰는 통용어, 작업도구, 공구이름가운데서 문화성있는 고유어를 찾는 일이다.
6. 계층별생활어: 어린이, 청소년학생, 녀성, 남성, 지식인 등 각계층별로 통하여 쓰는 입말속어, 통용어가운데서 문화성있고 지혜있는 고유어표현들을 고르는 작업이다.
7. 문학작품의 새말: 소설, 시, 영화 등에서 작가들이 새롭게 만들어쓴 생동하고 표현성있는 어휘를 조사하는 일이다.

이 사업은 대단히 품이 많이 들고 사전편찬지식이 있는 전문가들에 의하여 진행되어야 하는만큼 여러 가지 어려움이 있을 수 있다.

그러나 이 사업을 반드시 해야 할 목적과 의의는 자못 크다.

우선 민족어유산의 발굴정리사업은 우리 인민이 창조한 귀중한 언어적재부를 더 잃어지기전에 찾아 보존리용하게 하는데 첫째가는 목적이 있다.

시대발전과 현대문명에 따라 이 사업의 시기성은 초미의 문제로 된다.

다음 민족어유산의 더 많이 찾아 고유어체계를 더 풍부히 하는것은 외래어의 람용을 직접적으로 막을수 있을뿐아니라 사람들의 언어의식과 언어생활기풍에 많은 긍

정적영향을 주게 된다.

통속적으로 이해한다면 낱말이 더 많이 밀려드는 외래어를 그냥 그대로 받아쓰기보다는 묻혀있거나 소극화된 방언이나 고유어들을 되살려쓰는편이 낫다는것이다.

또한 오늘 북과 남, 해외의 어휘의미상차이가 점점 커가고있는 때에 한뿌리를 가진 민족어유산을 발굴정리하는 사업은 언어의 통일적발전에 큰 밑거름이 된다고 볼수 있다.

그러면 언어유산의 발굴정리에서 지켜야 할 언어학적원칙은 무엇인가?

그것은 우선 유산이라고 하여 지나치게 쟁스럽거나 건전치 못한 언어요소들을 다 찾아서 모아놓지 말아야 하는것이다. 우리 말의 우수성을 자랑하고 어휘체계의 발전풍부화에 보탬이 되는것이라야 언어적재부로 되는것이지 문화성없고 몰상식하고 속된 언어수단들을 들추어낼 필요가 없는것이다.

다음으로 언어유산을 발굴하고 수집정리하는데서 어휘론적가치를 과학적으로 평가하는것이다. 어휘론적으로나 의미론적으로는 물론 단어조성적으로 볼 때 지역별발음경향과 역사주의원칙을 정확히 계측평가하고 수록정리해야만 자료가 응당한 가치를 가질수 있다.

또한 수집정리된 자료가 목적인바대로 언어의 민족성을 살리고 민족어의 통일적발전에 이바지할수 있도록 활용방도를 옳게 세워나가는것이다.

어휘정리사업이나 새로운 국어사전 또는 통일사전편찬에서 새로 발굴정리된 어휘자료들과 그의 단어조성적수법이 널리 쓰이도록 하는것도 하나의 방도일것이다.

오늘 6.15 공동선언의 기치아래 겨계의 통일열망이 그 어느때보다 높아가고 있는 때에 민족공동의 재부인 우리 말의 고유성과 우수성을 지켜나가는 사업에 기여하는것은 국어학자들의 신성한 의무라고 생각한다.

우리는 민족어유산을 발굴정리하는 어렵고 방대한 사업을 실속있게 진척시켜나감으로써 겨레앞에 지닌 자기의 책임을 다해야 할것이라고 본다.

문화적 자산으로서의 우리말 어휘

남 기 심(국립국어연구원장)

1

지금 쓰고 있는 말이거나 예전에 썼던 말이거나, 또는 중앙어이거나 지역 방언이거나 우리의 말이면 모두 우리 삶의 모습, 우리 삶의 방식을 담고 있는 그릇이요, 우리가 자연과 더불어, 또는 일정한 규범 속에서 이루어 온 우리 문화의 지표이며, 귀한 문화적 자산이다.

그런데 우리는 지금까지 우리말의 어휘를 모아서 구조화해 보려는 노력을 해 본 적이 없다. 옛날에 실학자들이 더러 ‘物譜’, ‘物名攷’, ‘事類博解’ 등과 같은 어휘집을 만들어 그러한 관심을 보인 일이 있으나, 나라를 되찾고 우리 손으로 국어 정책을 시행해 온 이래로 이런 일에 주의를 기울인 일이 별로 없다. 근년에 남쪽에서 ‘우리말 분류 사전’, ‘우리말 갈래 사전’, ‘우리 토박이말 사전’ 등이 출간되어 이런 방면에 관심을 가지게 하는 계기가 되기는 했으나, 광범위한 기초 조사 위에서 이루어진 것이 아니어서 규모가 작아, 우리말을 다 아우르는 데는 멀리 못 미치고 있다. 국립국어연구원의 표준국어대사전이 애초에 이러한 목적으로 편찬을 시작하기는 했으나 아직은 그것이 그대로 이루어지지 않은 꿈으로 남겨져 있다. 더 많은 준비와 더 많은 시간이 필요하기 때문이다. 광범위한 자료의 수집, 곧 남과 북의 모든 지역 토박이말을 비롯하여, 중국 지역의 조선말은 물론 구소련 지역의 이른바 고려말, 미주 한인 사회의 이민 언어에 이르기까지 통틀어 우리말 모두를 수집 정리하는 일은 많은 시간을 요하는 방대한 작업이요 힘든 일임에 틀림없으나 언젠가는 거쳐야 할 과제로서 오늘의 이러한 모임을 통해서 이제 우리는 그 첫발을 내디디고자 하는 것이다.

2

어휘 문제와 가장 밀접하게 연관되어 있는 것이 사전이다. 우리가 사전을 만들기 시작한 이후 100년이 채 안 되는 짧은 기간이기는 하지만 그동안 사전 편찬에 적지 않은 노력을 기울여 왔다. 그러나 지금까지의 국어사전 편찬은, 남에서나 북에서나, 널리 쓰이는 말들을 중심으로 한 실용적 목적, 혹은 규범적 목적에 초점을 맞추어 왔다. 맞춤법, 표준어 등의 언어 규범을 정착시키는 일이 무엇보다도 급했기 때문이다. 그러한 노력의 결과로 국어 규범은 상당히 정착이 된 셈이다. 이제는 우리의 언어 정책적 관심의 폭을 우리 문화 전반에 걸친 것으로, 그리고 우리 민족의 정체성을 제고하는 쪽으로 크게 확대 전개해야 하고, 사전 편찬도 이와 궤도를 같이 해야 할 필요가 있다. 이에 선행되어야 할 일이 우리의 문화적 유산으로서의 각 지역어를 발굴 수집 정리하는 일이다. 이것은 어학적 사실에 국한되었던 우리말에 대한 관심의 폭을 문화 전반적인 영역으로 확대하여 펼쳐 나가기 위한 가장 핵심적이며, 기본적인 작업이기 때문이다.

그동안 북에서 여러 차례의 사전 편찬 작업을 통해서, 그리고 오랜 동안의 말 다듬기 작업을 통해서 많은 양의 방언을 발굴하여 중앙어로 격상하기 위한 노력을 기울여서 상당한 성과를 거두어 온 것은 국어의 어휘 자산 활용의 폭을 넓혔다는 점에서 높이 평가할 일이며, 위에서 언급한 바와 같이, 근년에 ‘우리말 갈래 사전’, ‘우리말 분류 사전’, ‘한국 문화 상징 사전’, ‘한국 언어문화 사전’ 등의 특수 사전이 남쪽에서 선보이기 시작한 것 역시 언어 규범에 한정되었던 우리의 관심의 폭을 더 넓은 영역으로 넓히는 계기를 마련했다는 점에서 고무적인 일이다. 그러나 ‘우리말 갈래 사전’ 등의 업적들이 모두 기존 인쇄물을 통해서 얻을 수 있는 정보를 가공하여 만든 것들이기 때문에 새로 발굴한 어휘가 많지 않다는 것은 아쉬운 점이다. 국내에서는 아직 낱말발 사전의 편찬 같은 것이 시도되지 않았다. 국어의 어휘론 방면의 연구가 부진한 데 그 원인이 있고, 그것은 국어 연구를 하는 사람들이 가지고 있는 관심 영역의 한계를 드러내고 있는 것이기도 한다. 낱말발 식의 어휘 정리를 일상생활 영역, 농, 공, 상 등 여러 분야의 문화, 지역 문화, 역사적 유물의 영역 등으로 넓혀 조직적으로 정리하는 것은

우리의 문화 전반을 체계적, 구조적으로 조명하는 한 방법이 될 것이며, 문화의 비교 고찰을 가능하게 하는 발판이 될 것으로 생각한다.

요컨대, 광범위한 어휘의 수집, 이것은 우리에게 주어진 오늘의 과업이다.

3

옛말 자료는 이미 상당량이 전산 처리되거나 되고 있는 중으로 이 작업은 시간문제 일 뿐인 것으로 생각된다. 일반적으로 글말 자료는 소실될 염려가 적어 충분한 시간이 주어지면 정리가 될 것이다. 문제는 입말 자료의 조사 정리이다.

방언 자료는 지난 80년대에 한국정신문화연구원이 여러 해에 걸쳐 조사한 자료가 있고, 학자들이 개인적으로 조사한 자료도 있으며, 지방 자치제가 실시된 이후에 몇몇 지방 자치 단체에서 향토 문화 사업의 일환으로 조사 수집한 것도 있다. 대부분 방언 학적 연구, 국어사적 연구를 위한 필요에서 조사 수집한 것이지만, 지방 자치 단체가 주도한 조사는 문화적 시각에 초점을 맞춘 것이었다. 문화적 필요를 위한 방언 자료는 한국정신문화연구원의 구비 문학 자료도 활용할 수가 있다. 북에도 방언 조사 자료가 축적되어 있고, 중국 조선말에 대해서도 이러한 자료가 있는 것으로 알고 있다. 따라서 이들을 모두 합치면 상당한 양의 자료가 집적되어 있는 셈이다. 그럼에도 불구하고, 그러한 자료들이 모든 지역을 충분히 망라했다거나, 다양한 문화적 시각의 조명을 하기 에 충분하다거나 하기 어려우며, 최근의 급속한 방언 소멸의 속도를 감당할 수 있는 수준의 것이라고도 하기 어렵다. 더 나아가 이미 수집된 이러한 자료들이 조직적으로 안전하게 관리되고 있다고 하기도 어렵다. 방언의 조사는 지속적으로 이루어져야 하는데 그것은 아예 바랄 수도 없다.

이제 남과 북이 힘을 합쳐 전국 규모의 방언 조사를 하는데 있어서 이러한 제반 문제점을 하나하나 짚어 가면서 작업을 진행해야 할 것이라는 점을 강조하고자 한다.

중국에서의 조선어 자료 수집에 대하여

전 학 석(연변대)

1. 과제선정

1.1. 과제선정배경

중국정부의 제5차 전국 인구조사(2001.11.1)에 의하면 중국의 총 인구는 12억 9,533만명이며 그중 조선족은 192만명으로서 중국 총 인구의 0.15%를 차지한다.¹⁾

현재 중국의 조선족은 대부분 동북 3성(길림성 1,145,688명, 흑룡강성 388,458명, 료녕성 241,052명)에 살고있는데 조선족 총 인구의 92.27%(1,775,198명)를 차지한다.

동북 3성의 조선족은 주로 한개 자치주, 한개 자치현, 50개 자치향²⁾, 3,000여개 촌락에 상대적으로 집중되어 살아왔다. 그러나 20세기 80년대이후 중국의 개혁, 개방정책의 실시와 더불어 조선족사회에는 많은 변화가 일어났다. 조선족인구의 마이너스증장, 조선족의 도시이로의 이동, 조선족의 해외로의 진출 등으로 조선족의 주거형태, 조선족의 교육, 조선족의 언어위상 등에 뚜렷한 변화가 일어났다.

우선 조선족의 주거형태에 변화가 일어났다.

중국의 개혁, 개방이전에는 조선족이 대부분 농촌에 정착하여 살아왔지만 개혁, 개방이후 농촌인구가 점점 줄어들고 도시인구가 점점 늘어나고 있다. 예를 들면 1982년에 산둥성에 조선족이 930명밖에 없었으나 2000년에는 27,795명으로 늘어났다.³⁾ 이들

1) 제5차 전국 인구조사 자료 참고.

2) 金炳鎬 《中國朝鮮族人口簡論》 (1993) 23-24페이지 참고.

3) 제5차 전국 인구조사 자료 참고.

은 모두 산둥성의 도시에 살고있다.

조선족의 도시에서의 이동과 해외로의 진출은 조선족의 집거형태에서 산거(散居)형태로, 조선족이 위주였던 촌락이 한족이 위주인 촌락으로의 변화를 초래하고있다.

다음은 조선족교육에 변화가 일어났다. 조선족인구의 마이너스증장과 조선족 주거형태의 변화는 조선족학교가 줄어들고 조선족교육의 교육질이 저하하는 결과를 초래하고있다.

그 다음으로 조선족학교가 점차적으로 줄어들고 조선족학교의 교육질이 저하함에 따라 조선족학생이 한족(漢族)학교에 가는 수가 늘어나고 조선족학교에서의 교수용어를 한어(漢語)로 (조선어문 제외) 하자는 경향이 거세짐에 따라, 한어(漢語)가 공통어인 중국에서 한어를 잘 해야 한다는 사회적관념에 따라 조선어의 사회적위상은 점점 떨어지고있다.

즉 언어를 그 기능에 따라 <민족간의 통용어>, <구역 우세어>, <민족내의 교제어>⁴⁾ 등으로 층위를 나누었을 때 조선어는 연변조선족자치주내에서는 구역 우세어의 층위에 있을수 있는 언어이지만 실제로 조선어는 연변조선족자치주내에서도 <민족간의 통용어>의 층위에 있는 언어라고 할수 있을것이다. 따라서 조선어의 기능이 저하하고 비규범적인 언어사용이 늘어나고있다.

중국에서의 조선족의 교육을 살리고 조선어를 보존하고 바로 살려쓰기 위하여서는 중국 조선족사회의 이와 같은 상황을 소상히 조사하고 연구해야 한다. 뿐만아니라 우리 민족의 가장 보귀한 문화유산인 조선어를 계속 계승하고 발전시키기 위하여서는 현재 지역적으로 쓰이는 언어들을 조사하고 정확하게 기술해야 한다.

1.2. 과제래원

2003년 11월 북경에서 있는 《민족어의 통일적 발전과 방언조사 연구에 관한 학술모임》 이후 우리는 나름대로 《중국 조선족 언어사회구역 연구》라는 과제를 선정하고 조사연구사업을 추진하고있다.

4) 祝畹瑾 《社會語言學概論》 (1992) 187-188페이지 참고.

지난해 11월 30일에 《중국 조선족 언어사회구역 연구》라는 과제명으로 연변대학교 《211공정》(211工程)관공실에 연구과제를 제출하였다. 2004년 2월 6일에 연변대학교의 연구과제로 공식 선정되었다.

연구과제경비는 중점학과건설경비에서 지원하기로 하였다.

2. 과제실행

2.1. 조사항목

2.1.1. 조사지점의 일반상황 조사

2.1.2. 언어조사항목

2.1.3. 조선족인구와 조선족학교 통계

2.2. 조사지점

중국에 살고있는 조선족의 주거형태를 분류한다면 크게 대집거(大聚居), 소집거(小聚居), 산거(散居)로 분류할수 있을것이다.⁵⁾ 연변조선족자치주의 조선족 주거형태가 대집거형태에 속할것이고 장백조선족자치현과 50개의 조선족자치향의 조선족 주거형태가 소집거에 속할것이며 한족(漢族)이나 그외 다른 민족이 다수인 지역에 조선족이 몇가구 또는 몇십 가구씩 섞여사는 형태가 산거형태에 속할것이다.

이번에 행해지는 조사는 농촌일 경우에는 조선족이 100명이상 살고있는 촌(村)급 행정단위를 기준으로 조사지점을 선정하며 도시일 경우에는 500명이상 살고있는 구(區), 시(현급시)를 기준으로 조사지점을 선정한다.

5) 金炳鎬 《中國朝鮮族人口簡論》 (1993) 21-23페이지 참고.

2.3. 조사방법

2.3.1. 조사팀

- 1팀: 길림성(연변제외) 조사팀(2명)
- 2팀: 연변 조사팀(2명)
- 3팀: 흑룡강성 조사팀(2명)
- 4팀: 료녕성 조사팀(2명)
- 5팀: 내몽골자치구 및 기타 성, 시 조사팀(2명)

2.3.2. 조사방법

각 조사팀은 통일된 조사항목에 따라 현지에 내려가 직접 조사한다. 즉 직접조사방법을 위주로 하고 간접조사방법을 보조적으로 활용한다.

이번 조사에는 각 조사팀에 2명씩 도합 10명이 직접 책임지고 현지조사를 하는 외에 2004년 7~8월과 2005년 1~2월에 각각 50여명의 보조원을 동원하여 직접조사를 진행할 것이다.

3. 과제결과

3.1. 계열논문(조사보고)

다음과 같은 계열론문을 완수한다.

- 3.1.1. 중국 조선족의 언어사회에 대한 구획
- 3.1.2. 중국에서의 조선어 방언분포
- 3.1.3. 조선족 중, 소학교에서의 조선어문교육

3.1.4. 조선족의 규범 한자어 사용상황

3.2. 단행본

2005년 12월에 《중국 조선족 언어사회구역 연구》라는 제목으로 단행본을 출판한다.

[부록] 중국 조선족 언어사회구역 조사지

중국 조선족 언어사회구역 조사지

조사지점:

조사시간:

기록인:

1. 조사지점의 일반 상황

1) 본 조사지점의 간단한 개척력사:

언제, 어떤 사람들에 의하여 개척되었으며 개척초기 어떤 민족이 얼마나 살았습니까?

2) 본 조사지점의 인구(제5차 인구통계를 기준으로 삼음)

① 본 조사지점의 총인구:

② 본 조사지점의 조선족인구:

③ 조선족인구가운데서 어느 도(道)의 출신이 각각 몇프로를 차지합니까?

④ 본 조사지점의 조선족은 주로 어떤 업종에 종사합니까?

⑤ 본 조사지점의 조선족은 평소 어느 언어를 주로 사용합니까? 아래 ()안에 <V>를 표시해 주십시오.

가정: 한어() 조선어() 혼용()

사회: 한어() 조선어() 혼용()

3) 본 조사지점의 조선족은 주로 어떤 방언을 사용합니까? 사용인수의 다소에 따라 ①, ②, ③으로 표시해주십시오.

함경북도방언

함경남도방언

평안북도방언	평안남도방언
경상북도방언	경상남도방언
전라북도방언	전라남도방언
경기도방언	강원도방언
황해북도방언	황해남도방언
제주도방언	

4) 본 조사지점에는 조선족학생을 대상으로 운영하는 학교가 몇개 있습니까?

매 학교의 명칭, 학생수, 교원수, 학급수, 교수용어상황 등을 구체적으로 밝혀 적어 주십시오.

학교명칭:

학생수:

교원수:

학급수:

교수용어상황:(조선어로 강의하는가? 한어로 강의하는가? 어떤 학과목을 한어로 강의하고 어떤 학과목을 조선어로 강의하는가?)

언어사용실태조사(1)

조사시간:

조사기록인:

조사대상자(조사대상자는 반드시 로인층(60세이상)과 젊은층(30세좌우)에서 선정해야 함)

성명:

년령:

성별:

문화정도:

출신지방:

아래의 조사표에서 이 지방에서 쓰는 단어에 <<V>>를 해주십시오. 만약 다른 방언 형이 쓰이면 빈자리에 표기하여주십시오.

로인층(60세이상)과 젊은층(30세좌우)의 언어사용이 다를 경우에는 해당 단어에 표시해주십시오.

단어	<<V>>	단어	<<V>>	단어	<<V>>	단어	<<V>>	단어	<<V>>
가위		가새							
구유		구시							
겨울		저실							
가을		가슬							
김		기슴							
누에		느배		누배					
우윳		우봉		우병					
확		호박							
터울		터불							
새우		새비		새갱이		새빙개		새옹개	
머루		멀기		멀구		멀귀			
가루		갈기		가리					
수레		슬기							

단어	《V》	단어	《V》	단어	《V》	단어	《V》	단어	《V》
별레		별거치		별기					
시렁		실경		실공					
길(道)		질							
김치		짐치							
기름		지름							
기장(밥)		지장							
키(箕)		치		소보치					
흥년		승연							
흥(내)		승(내)							
형(님)		성(님)							
혀		세		세띠		세때			
전기		던기		던기					
철도		털도		털도					
장마당		당마당		당마당					
좋다		동다		동다					
이발		니발							
이마		니마		니매					
여름		녀름		너름					
일곱		닐곱		닐곱					
량식		냥식		양식		냥식			
려관		녀관		여관		너관			
량반		냥반		양반		냥반			
리발		니발		이발					
녀자		녀자		여자					
념불		념불		염불					
년령		년녕		연녕					
뇨도		노도		요도					
질다		즐다							
짐승		즘승							
싱겁다		승겁다							
찢다		쫓다							
치다(눈-)		츠다							
의사		이사		으사					
의견		이견		으견					
의복		이복		으복					
의논		이논		으논					
의문		이문		으문					
위(胃)		y(홀모음)		wi(겹모음)					
위문		ymun(홀)		wimun(겹)					
귀(耳)		ky(홀)		kwi(겹)					

단어	《V》	단어	《V》	단어	《V》	단어	《V》	단어	《V》
쥐(鼠)		fy(홀)		fw(겹)					
늪(쌀-)		ny(홀)		nwi(겹)					
외삼촌		웨삼촌		웨삼촌					
외나무다리		웨나무다리		웨나무다리					
외딸		웨딸		웨딸					
원손		웬손		웬손					
외국(外國)		웨국		웨국					

언어사용실태조사(2)

조사시간:

조사기록인:

조사대상자(조사대상자는 반드시 로인층(60세이상)과 젊은층(30세좌우)에서 선정해야 함)

성명:

년령:

성별:

문화정도:

출신지방:

아래의 조사표에서 이 지방에서 쓰는 단어에 <V>를 해주십시오. 만약 다른 방언 형이 쓰이면 빈자리에 표기하여주십시오.

로인층(60세이상)과 젊은층(30세좌우)의 언어사용이 다를 경우에는 해당 단어에 표시하여주십시오.

단어	<V>	단어	<V>	단어	<V>	단어	<V>	단어	<V>
잡니다		가꾸마		감무다		감니더		감네다	
		가우다		가왜다		가지라오			
먹습니다		먹스꾸마		먹슴무다		먹심더		먹습네다	
		먹수다		먹소이다		먹췌다			
잡니까		감둥, 감두		감네까		감네꺼		감네껴	
		가우까		가왜까					
먹습니까		먹슴둥 먹슴두		먹습네까		먹습네꺼		먹습네껴	
		먹수까							
가십시오		갑소 갑서		가십소 가십서		가이소		가시소	
갑시다		갑세다		가갑소		갑수다		갑시더	
		가입시다		가개라우		갑주			

언어사용실태조사(3)

아래의 단어의 발음을 표시하여주십시오. 해당 발음에 <V>를 해주십시오. (音高, 音長, 音強)

- 《말이》(言)에서의 《말》 : 높음(), 김(), 셈()
 《말이》(馬)에서의 《말》 : 높음(), 김(), 셈()
 《눈이》(雪)에서의 《눈》 : 높음(), 김(), 셈()
 《눈이》(目)에서의 《눈》 : 높음(), 김(), 셈()
 《밤이》(栗)에서의 《밤》 : 높음(), 김(), 셈()
 《밤이》(夜)에서의 《밤》 : 높음(), 김(), 셈()
 《일이》(事)에서의 《일》 : 높음(), 김(), 셈()
 《일이》(一)에서의 《일》 : 높음(), 김(), 셈()
 《제비》(燕子)에서의 《제》 : 높음(), 김(), 셈()
 《제비》(抽籤)에서의 《제》 : 높음(), 김(), 셈()
 《가지》(茄子)에서의 《가》 : 높음(), 김(), 셈()
 《가지》(枝)에서의 《가》 : 높음(), 김(), 셈()
 《매다》(系)에서의 《매》 : 높음(), 김(), 셈()
 《매다》(剗)에서의 《매》 : 높음(), 김(), 셈()
 《묻는다》(聞)에서의 《묻》 : 높음(), 김(), 셈()
 《묻는다》(埋)에서의 《묻》 : 높음(), 김(), 셈()

언어사용실태조사(4)

조사시간:

조사기록인:

조사대상자(조사대상자는 반드시 로인층(60세이상)과 젊은층(30세좌우)에서 선정해야 함)

성명:

년령:

성별:

문화정도:

출신지방:

아래의 조사표에서 이 지방에서 쓰는 단어에 <V>를 해주십시오. 만약 다른 형태로 쓰인다면 빈자리에 표기하여주십시오.

로인층(60세이상)과 젊은층(30세좌우)의 언어사용이 다를 경우에는 해당 단어에 표시해주십시오.

한어	한자어	<V>	한자어	<V>	한자어	<V>
街道	가도		가두			
家長	가장		학부형		학부모	
居民	거민		주민			
竟@	경색		경쟁		시합	
工齡	공령		근무년한			
工資	공자		로임		월급	
工人	공인		로동자			
導演	도연		연출			
聯歡會	련환회		친선모임			
領導班子	령도반즈		지도부			
領隊	령대		인솔자			
成本	성본		원가		본값	
城市	성시		도시			
牲畜	성축		가축		집짐승	

한어	한자어	《V》	한자어	《V》	한자어	《V》
廠長	창장		공장장			
請假	청가		말미			
招生	초생		학생모집			
粗糧	추량		잡곡			
破案	파안		사건해명			
評比	평비		평의			
表揚	표양		표창			
風收	풍수		풍작			
批發	피발		도매			
畢業證	필업증		졸업증		졸업장	
函授生	함수생		통신생			
護理員	호리원		간병원			
流氓	류망		건달		부랑자	
晚會	만회		야회			
名額	명액		인원		인수	
毛病	모병		병집		결함	
門診部	문진부		진찰부			
民辦教員	민반교원		민영교원			
密碼	밀마		비밀번호		미마	
發貨票	발화표		매상전표			
電話號碼	전화호마		전화번호			
收據	수거		령수증			
病志	병지		병력서			
產值	산치		생산액		찬즈	
上課	상과		상학		수업	
備課	비과		교수준비			
教案	교안		교수안			
護照	후조		려권			
話劇	화극		연극			
化驗單	화험단		생화학분석표			
掛號信	파호신		등기우편			
樣品	양품		견본			
業務水平	업무수평		실무수준			
余額	여액		잔고		잔액	
演出節目	연출절목		공연종목			
演員	연원		배우			
營業執照	영업집조		영업증			
外行	외항		문외한			
口糧	구량		식량			
伙食費	화식비		식비			
紅十字	홍십자		적십자			
土方法	토방법		민간방법		재래식방법	

년대별 통계(조선족인구, 조선족학교):

1) 년대별 인구통계(조선족인구)

제1차 인구통계(각 성, 시, 진(향), 촌)

제2차 인구통계(각 성, 시, 진(향), 촌)

제3차 인구통계(각 성, 시, 진(향), 촌)

제4차 인구통계(각 성, 시, 진(향), 촌)

제5차 인구통계(각 성, 시, 진(향), 촌)

2) 년대별 학교통계(조선족학교)

50년대 조선족학교 통계(각 성, 시, 진(향))

60년대 조선족학교 통계(각 성, 시, 진(향))

70년대 조선족학교 통계(각 성, 시, 진(향))

80년대 조선족학교 통계(각 성, 시, 진(향))

90년대 조선족학교 통계(각 성, 시, 진(향))

2000년대 조선족학교 통계(각 성, 시, 진(향))

제1분과 민족 방언의 공동 연구



빈 면

방언과 사회 발전

- 이른바 언어 발전 문제와 관련하여 -

김 하 수(국립국어연구원 어문규범연구부장)

1. 발전의 의미

일반적으로 ‘발전’ 혹은 ‘발달’이라는 용어는 언어 연구자들에게는 비교적 생소한 말이다. 둘 다 영어(혹은 인구어)에서 사용하는 ‘development’라는 뜻을 가지고 있으나 우리말에서의 쓰임은 약간의 차이를 보인다. 대체적으로 유의어적인 관계에 있으나 ‘발달’은 ‘근육, 심리’ 등과 같은 미시적인 대상에 대해 사용할 수가 있는 반면, ‘발전’이라는 말은 ‘사회, 조직, 국가’ 등과 같은 거시적 대상에 많이 사용한다. 그 까닭에 심리학에서 언어 습득 과정을 논의하는 과정에서는 대개 ‘언어 발달’이라는 용어가 쓰이는 반면에 사회 언어학에서는 ‘언어의 발전’이라는 개념이 상대적으로 더 유용하다.

언어를 형식적인 틀에서 관찰하는 경우에는 ‘언어의 발전’이라는 개념을 사용하기에는 적잖은 무리가 따른다. 예를 들어 형태 요소 변동이나 피동과 능동의 상호 변형이 발전과 무슨 관계인지 사실상 설명이 불가능하다. 그러나 언어 습득 과정을 관찰하는 경우에는 ‘발전’의 현상이 뚜렷하게 나타난다. 사용 음운의 양이 증가한다든지 어휘 사용의 예가 다양해진다는지 하는 경우, 또 단순한 문장 형성에서 복잡한 문장의 형성 등으로 진행되는 언어 능력의 변화를 보면 여기에는 분명히 ‘발전’ 혹은 ‘발달’한다는 개념을 아니 쓸 수가 없다.

사회 언어학에서도 ‘발전’이라는 개념을 통해 한 언어가 사회 내부에서 어떠한 기능을 맡아 가는지를 관찰할 수가 있다. 아마 대표적인 것이 방언의 지위에서 공용어의

지위로 승격하게 된 노르웨이 어를 연구한 Haugen(1966)을 들 수 있을 것이다. 다시 말해서 비공식적인 지위에 있던 언어(혹은 방언)가 일정한 사회적 동기에 힘입어 공식적인 지위를 획득하는 현상은 여러 지역의 탈식민지화 과정에서 자주 목격할 수 있다.

또 언어에 대한 역사적 연구를 하는 과정에서도 종종 중세어에서 현대어로의 ‘발전’과 같은 개념을 쓰기도 한다. 우리의 언어에서도 19세기에서 20세기로 넘어오는 과정에서 어휘와 문장 구조가 ‘질적인 변화’를 보여 주는 것을 볼 수 있다.

곧 ‘언어의 발전’이라고 하는 것은 그 형태의 변화가 아니라 형태와 의미, 그리고 쓰임새가 시대적인 요구에 부응하는 ‘질적인 변화’를 수용하는 것으로 볼 수 있다. 그리고 그 질적인 것의 내용은 시대가 새롭게 요청하는 새로운 ‘가치’에 바탕을 둔다. 이러한 발전을 겪어 새롭게 된 언어는 그 이전의 언어와 상당한 질적 차이를 보이면서 일반 언어 사용자들에게는 낯은 텍스트에 접근하기 어려워진다. 다시 말한다면 언어 대중으로 하여금 낯은 시대와 의사소통을 못하게 하고 새 시대에 적응하도록 하는 강제적 구실을 하는 것이다. 그런 점에서 언어의 발전이라는 것은 그저 자연발생적인 것이라기보다는 매우 정치적 현상이다.

또한 언어의 발전은 근본적으로 사회 발전과 그 축을 함께 한다. 사회 발전은 끊임 없이 그에 상응하는 사람의 행태의 변화를 불러오며, 그러한 행태의 변화는 언어의 변화를 불러온다. 이 언어의 변화는 마침내 사회 변화와 맞물려 언어 발전을 낳게 된다는 것이다.

2. 언어의 변화와 발전

언어는 한편으로는 변화하면서, 또 다른 한편으로는 발전한다. 물론 퇴화하는 퇴행적 발전도 가능하다. 그러나 그러한 현상은 매우 논쟁적인 다른 명제와 관련되기 때문에 이 글의 논점에서는 일단 져혀 두기로 한다.

앞에서 예를 들었던 Haugen(1972)은 언어의 발전 문제를 방언과 개별 언어의 관계를 따지면서 이론적으로 정밀화하며, 방언에서 독자적인 언어(공용어의 지위)를 획득하기 위해서는 부호화(codification), 정교화(elaboration)라는 두 가지의 주요 조건을

제시했다. 여기서 부호화라고 함은 이른바 표준화된 규범을 가리키는 것이고, 정교화라 함은 그 언어를 사회적으로 고급 기능 분야(과학 기술 등)에서 실질적으로 사용하는 것을 말한다.

다시 말해서 아무런 규범(발음, 표기 등) 없이 사용하던 방언에 일정한 규범을 부여하고 그것을 사회적으로 폭 넓게, 고급 기능으로 사용한다면 그것은 하나의 공용어의 지위를 가지게 되는 것으로 파악했다. 물론 그의 논의는 주로 유럽의 경험을 바탕으로 한 것이었지만 이와 같은 관점은 더욱 세밀한 현장 연구를 바탕으로(Ferguson(1962)), 1960년대 광범위한 식민지 해방 과정을 연구하는 연구에서도 일관되게 받아들여졌다.

이러한 관점의 성격을 파악해 보면 대략 다음과 같은 몇 가지의 '발전' 기준과 조건을 찾아볼 수 있다.

- 1) 좁은 영역의 의사소통에서 넓은 영역의 의사소통으로
- 2) 적은 사용자 집단에서 많은 사용자 집단으로
- 3) 산만한 사용에서 집중화된 사용으로, 곧 제도적 보장이 전제되는 것
- 4) 문헌, 매체 등을 통하여 폭넓은 사용의 가능성을 열어 놓고 있을 것
- 5) 외국어로의 번역이 가능할 것

조건 1)의 경우는 지역이나 영역의 한계를 넘어야 한다는 것이다. 지역의 한계를 벗어나지 못하면 방언의 비위를 벗어날 수 없고, 영역의 한계를 뚫지 못하면 특수 어휘 혹은 은어의 기능밖에 할 수 없게 된다. 조건 2)는 단순히 사용 인구의 수를 가리키는 것이 아니라 광범위한 언어 사용 대중을 뜻한다. 언어 대중 속에서 언어는 왕성하게 생성과 소멸을 하는 '자연 언어'로서의 속성을 유지하게 된다.

매우 중요한 조건이 바로 3)번 조항이다. '제도'에 의해서 해당 언어가 강제될 필요가 있다. 여기서 '제도(institution)'라고 하는 것은 법적 제도, 인습적 제도, 편이적 제도 등을 망라하여 가리킨다. 이러한 제도에 의하여, 재판, 회의 진행, 교육, 문서 작성, 종교적 예식, 집체적 노동 등에서 언어 사용이 관철된다.

더 나아가 다양한 문헌과 기록 형태를 생산하고 소비할 수 있는 구조가 만들어지고, 이를 뒷받침하는 각종 매체(인쇄, 도서, 영상, 음성, 전파, 전자 등)가 마련되어야 한다.

끝으로 이 언어를 다른 언어로 번역이 가능하도록 그 언어적 기능과 의미부가 보편성을 띠어야 한다는 것이다.

언어 자체의 변화상보다는 그와 맞물린 사회적 성격의 질적 변화를 주목하고 있는 것이다.

3. 언어 발전의 양상들

언어의 발전 양상은 역사적으로 무척 다양하게 진행되고 있다.

가장 일반적으로 관찰되는 것은 입말에서 글말로의 발전이다. 이는 언어가 문자를 매개물로 삼기 시작하면서 진행된 결과로 언어를 상황으로부터 상대적인 독립을 이루게 한다.

이와 같은 성질을 가진 것이 바로 방언에서 공통어로의 발전이다. 사용자 범위가 넓어지면서 지역적 특성이 희석되고 탈지역적 공통성이 강화된다.

그 다음으로 관찰할 수 있는 것이 통속어에서 공용어로의 발전이다. 이것 역시 입말에서 글말로 발전하고 있는 전반적 다른 양상과 같은 궤를 그린다.

매우 중요한 발전 양상은 자연 언어에서 전문 용어로의 발전에서 확인할 수 있다. 전문 용어는 기본적으로 '제도적으로 창출된 언어'이다. 자연 언어가 일상성의 구속에서 벗어나 제도의 한 부분으로 기능하게 되는 고도의 발전을 보여 준다. 곧 사람에 의해서 만들어져서 사람을 오히려 구속하는 기능을 가지게 되는 것이다.

4. 방언의 기능 재발전

만일 방언이 점점 공통어로 발전을 하고 세월이 가면 과연 방언은 살아남을 수 있을까 하는 의문은 누구든지 자연스럽게 품어 봄 직한 의문이다. 대부분의 사회에서 각 지역이 경제적으로 강하게 통합되어 나가면서 지금까지 우리가 보아 왔던 전형적인 방언, 특히 어휘 중심의 방언은 쇠퇴하고 있는 것이 사실이다. 그러나 방언의 문법적 요소, 곧 문장 어미, 존대법, 억양 등을 통한 지역인들의 유대감 확인 과정은 그리 쉽

게 사라지지 않고 있다. 아마도 앞으로의 방언 조사는 어휘카드 중심이 아닌 비분절 요소 및 표정과 손짓 등의 준언어적 요소 중심으로 흘러가지 않을까 하고 조심스럽게 예측해 본다. 다시 말해서 방언은 구체적인 언어적 의사소통의 주역에서 서로의 유대감을 확인하는 기능으로 바뀔 것이라는 뜻이다.

또 다른 한편으로 방언은 무수한 자연 언어의 생성 원천이다. 앞으로 상당 방언이 한편으로 소멸되어 가고, 또 다른 한편으로 공통어로 발전해 나갈 것이다. 특히 농어촌 지역에서 많이 관찰되는 동식물 명칭, 음식물 명칭 등은 공통어에, 특히 전문 용어에도 많은 영향을 끼칠 것이다. 마치 도시 지역의 새로운 어휘가 공업적인 양태로 생산되듯이 농어촌 지역에서는 농업적이고 생태적인 현상을 보일 가능성이 크다. 예를 들어 도시 언어는 추상적이고 관념적인 어휘가, 농어촌 지역에서는 감성이 풍부한 형용사나 부사 어휘가 발전할 가능성이 크다는 것이다.

무엇보다 중요한 것은 앞으로 사회 발전의 방향과도 관련이 되겠지만 한 언어가 국가 단위로 획일화되어 가는 것을 지양하는 태도가 강해진다면, 하나의 언어 내부에 매우 다양한 요소의 발생을 일으키는 언어의 생명력을 보장하는 기능은 결코 끝나지 않을 것이다.

<참고 문헌>

- 김영환(1978), 조선민족어발전력사연구, 과학,백과사전출판사
- Ferguson, C. A.(1962), The language factor in national development, In: Anthropological Linguistics vol. 4, no. 1, pp23-27
- Haugen, E. (1966), Language and Language Planning: The Case of Modern Norwegian. Cambridge, Mass. : Harvard University Press
- _____ (1972), Dialect, Language, Nation, In: Pride & Holmes pp97-111
- Kloss, H. (1952), Die Entwicklung Neuer germanischen Kultursprachen von 1800 bis 1950, Pohl, München

황해도방언에서 토의 쓰임에 대하여

리 기 만(사회과학원 언어학연구소 실장)

방언은 해당 지역에서 오래동안 생활해오는 사람들속에서 입말을 통하여 대대로 전해 내려오는 귀중한 언어유산이므로 그에 대한 연구는 매우 중요한 의의를 가진다.

우리 말의 방언적현상은 어음, 어휘, 문법 등 여러 측면에서 나타나지만 해당 지역의 방언적특성이 가장 뚜렷하게 나타나는것은 토라고 볼수 있다.

일반적으로 언어에서 문법적현상은 어음이나 어휘에 비하여 그 변화가 매우 더디게 이루어지는것처럼 방언에서 쓰이는 토도 어음이나 어휘와는 달리 다른 방언의 영향을 잘 받지 않는 견고성을 가지고 오래동안 보존되면서 해당 지역의 사람들속에서 쓰이게 된다.

이런 견지에서 보면 방언토는 해당 방언지역을 다른 방언지역과 가려보는 뚜렷한 징표로 된다고 말할수 있다.

이 논문에서는 황해도방언에서 토의 쓰임을 1990년초에 조사한 말법토자료에 기초하여 중점적으로 고찰하려고 한다.

말법은 말하는 사람이 듣는 사람에게 어떤 목적을 가지고 말하는가를 나타내는 문법적범주이다.

문화어의 말법은 《알림》, 《물음》, 《시킴》, 《추김》의 네가지로 이루어지지만 황해도방언에서는 《알림》, 《물음》, 《시킴》의 말법토가 쓰이며 《추김》은 거의나 문화어토에 의하여 이루어지고 있다. 때문에 여기에서는 《알림》, 《물음》, 《시킴》의 세 측면만을 보게 된다.

첫째, 알림법에 사용되는 황해도방언토로는 <-시다(-외다), -어다(-이어다), -어요, -

나요, -어(-서), -어어, -다야, -다이, -다어》 등이 쓰인다.

이 방언의 알림법토들은 말을 듣는 상대가 누구인가 하는데 따라 서로 다르게 쓰인다.

방언토 《-시다(-외다)》는 말을 듣는 상대가 말하는 사람보다 우인 경우에 많이 쓰이는데 주로 전형적인 황해도방언의 소유자들속에서 남녀의 구별이 없이 두루 쓰인다.

○ 성님 말이집이 갔시다.

이 레문은 황해남도 신천군 새길리에 사는 65살난 로인이 《성님이 어데 갔시까?》라고 묻는 말에 대답하는 문장이다.

그러나 방언토 《-어다(-이어다), -이요, -다어》는 말을 듣는 상대가 웃사람일 때나 동년배인 경우에도 쓰인다.

- 아바진 산술 돌아보어다.(황해남도 재령군 읍 54살, 남자)
- 들채녀석은 도에 갔다어.(황해남도 신원군 계남리 63살, 여자)
- 이자 작업총활 했이요.(황해북도 은파군 묵천리 57살, 여자)

우의 레문들에 쓰인 토 《-어다(-이어다), -다어》는 주로 60세이상의 사람들속에서 많이 쓰이며 《-이요》는 50대의 사람들과 그아래의 젊은 사람들속에서도 널리 쓰인다.

방언토 《-어(-서), -어어, -다야》는 말을 듣는 상대가 말하는 사람과 같거나 아래인 경우에 그 구별이 불명확하게 쓰이고 있다.

- 돼지발 손질하어.(황해남도 해주시 광석동 67살, 여자)
- 신문을 받서.(황해남도 벽성군 읍 59살, 남자)
- 버스가 학교에 있다야.(황해북도 사리원시 철산동 50살, 남자)
- 약열 먹어어.(황해남도 삼천군 군산리 54살, 여자)

우의 레문에서 본 방언토 《-어(-서)》는 보통 나이가 많은 사람들속에서 주로 쓰이

며 《-어어, -다야》는 어린이들로부터 늙은이들속에서 널리 쓰이고 있다.

둘째, 물음법에서 사용되는 황해도방언토로는 《-ㄱ니까(-습니까), -ㄱ네까(-습네까), -ㄱ데까(-습데까), -시까(-외까), -까(-이까), -지러, -다나, -이, -ㄴ》 등이 쓰인다.

방언토 《-ㄱ니까(-습니까), -ㄱ네까(-습네까), -ㄱ디까(-습디까), -시까(-외까)》는 문화어의 물음법토 《-까》의 변종으로서 황해도방언에서 《높임》의 뜻으로 활발히 쓰이고 있다.

이 토들은 황해남도의 해주, 신원, 삼천, 신천, 재령, 태탄, 벽성, 장연, 과일군과 황해북도의 은파, 린산군을 비롯한 일부 지역들에 분포되어있는데 전형적인 황해도방언소유자들은 물론 세세대 학생들에서도 널리 쓰인다.

- 아자씨 읍니까? (황해남도 해주시 옥계동 40 살, 여자)
- 성님이 갑디까? (황해남도 벽성군 읍 49 살, 남자)
- 어디 갑네까? (황해남도 태탄군 성남리 67 살, 남자)
- 누굴 찾았시까? (황해남도 삼천군 읍 54 살, 여자)

방언토 《-까(-이까)》는 위의 토들과는 달리 《높임》과 《같음》의 계선이 없이 두루 쓰이고 있다.

- 값이 얼마까? (황해남도 재령군 읍 35살, 남자)
- 이게 심불콩이까? (황해북도 은파군 류정리 51 살, 여자)

방언토 《-다나, -이, -지러, -ㄴ》는 말을 듣는 상대가 말하는 사람과 동급인 경우에 많이 사용된다.

이 토들의 쓰임을 보면 《-지러》는 고유한 황해도방언지역들에서 두루 쓰이지만 《-다나, -이, -ㄴ》은 서북방언과 중부방언의 교차지역들에서도 다같이 널리 쓰이고 있다.

- 발김 매야지러? (황해남도 삼천군 읍 50 살, 여자)

토 <<-지러>>는 문화어의 토 <<-지러>>와 거의 같은 뜻으로 쓰이면서도 그것보다는 더 강한 물음을 나타내는데 남자들보다 여성들속에서 더 많이 쓰이고 있다.

- 서해잡문어러 바러 갔다나? (황해남도 송화군 읍 72 살, 남자)
- 입성 해와이? (황해북도 신계군 태율리 57 살, 남자)
- 음마이 일 나간? (황해북도 평산군 탄교리 47 살, 남자)

셋째, 시킴법에 사용되는 황해도방언토로는 <<-가라, -나라>>가 있다.

이 토들은 문화어토 <<-가라, -너라>>의 어음적변종으로서 말을 듣는 상대가 말하는 사람보다 아래인 경우에 쓰이고 있다.

- 할마니한데 가가라. (황해북도 서흥군 읍 52 살, 여자)
- 어서 이리 오나라. (황해남도 과일군 읍 73 살, 여자)

이 토들은 황해남북도의 거의 모든 지역에서 널리 쓰이고 있다.

또한 황해도방언에서는 알림법토 <<-어, -어어>>가 억양의 도움으로 시킴의 문법적 의미를 나타내는 경우가 있다.

- 일 시작해보어. (황해북도 토산군 읍 51 살, 남자)
- 어서 앉어어. (황해남도 연안군 읍 67 살, 여자)

이 토들은 남녀의 구별이 없이 늙은이로부터 새세대들속에서 널리 쓰이는데 주로 중부방언의 교차지대라고 할수 있는 황해북도의 금천, 토산군과 평산군의 일부 지역 그리고 황해남도의 연안, 배천, 청단, 봉천군에 분포되어있다.

이상에서 진행한 황해도방언에서 쓰이고있는 말법토의 쓰임에 대한 중심적인 고찰을 통하여 황해도방언분포의 지역적특성은 크게 두가지로 나누어보게 된다.

첫째로, 황해도방언은 주로 구월산줄기의 서남쪽과 불타산줄기, 수양산줄기와 멸악

산줄기, 정방산줄기를 축으로 하고 재령강을 중심으로 분포되어있다.

말법토의 쓰임을 통하여 황해도방언의 분포지역을 보면 황해남도의 해주시와 신원군, 신천군, 삼천군, 벽성군, 재령군, 태탄군, 룡연군, 송화군과 황해북도의 사리원시와 봉산군, 서흥군, 린산군, 은파군 등이다.

그러므로 고유한 황해도방언에 대한 조사연구는 위에 지적된 시, 군에 속한 농촌지역을 지점으로 하여 진행할 때라야 정확히 조사할수 있을것이다.

황해도방언에 대한 조사연구를 많은 지역에서 다 진행하기 어려운 경우에는 황해남도의 해주시, 삼천군, 재령군, 강령군과 황해북도의 봉산군, 은파군, 린산군의 농촌지역을 대상으로 정하고 조사하면 황해도방언의 특성을 비교적 정확히 조사할수도 있다고 본다.

그것은 위에서 지적된 시, 군들이 평안도방언이나 경기도방언의 영향을 적게 받은 지역으로서 황해도방언의 특성을 보존하고 있기때문이다.

둘째로, 황해도방언은 북동쪽으로는 서북방언지역인 평안도 방언과 교차되고 남쪽으로는 방언지역인 경기도방언과 교차되는 지역적특성을 가진다.

우선 황해북도의 황주군, 송림시, 연탄군, 연산군, 수안군, 곡산군, 신평군과 평안남도과 린접해 있는 은률군, 은천군, 안악군은 평안도방언과 황해도방언이 다같이 쓰이는 지역이다.

이 지역에서 물음의 뜻을 나타내는 고유한 황해도방언토 <<-가이>>와 전형적인 평안도방언토 <<-ㄴ>>이 다같이 두루 쓰이고있는것은 그 대표적인 실례이다.

○ 축구에서 이견. (곡산군 초평리 50 살, 남자)

○ 축구에서 이겨이. (수안군 읍 45 살, 남자)

그러나 이 지역들에서는 황해도방언의 고유한 물음법토 <<-까(-이까)>>가 쓰이는 경우를 찾아보기 어렵다. 이것은 이 지역들이 평안도방언의 영향을 비교적 많이 받은 지역이라는것을 보여준다.

또한 황해북도의 금천군, 토산군, 장풍군, 개풍군, 평산군의 일부 지역과 황해남도의

청단군, 연안군, 봉천군은 경기도방언의 영향을 많이 받은 지역이다.

이 지역들에서는 황해도방언과 중부방언이 다같이 쓰이지만 평안도방언의 영향을 받은 지역들과는 달리 황해도방언과 중부방언인 경기도방언이 다같이 쓰이고 있다.

그러므로 황해도방언과 중부방언의 교차현상을 조사하려면 황해북도의 개풍군, 금천군, 토산군이나 황해남도의 연안군, 봉천군을 대상으로 정하여야 하며 황해도방언과 서북방언인 평안도방언의 교차현상을 조사하려면 황해북도의 연산군과 황주군이나 황해남도의 은률군, 은천군을 대상으로 정해야 할것이다.

이상에서 본바와 같이 황해도방언은 자기의 고유한 특성을 가지면서도 일부 지역에서는 다른 방언들과 교차되어 쓰이는 특성을 가진다.

우리는 앞으로 방언지역들에 대한 정확한 인식을 가진데 기초하여 방언조사를 체계적으로, 과학적으로 진행해나감으로써 우리 인민들속에 묻혀있는 좋은 말들을 찾아내어 민족어를 더욱 발전시켜 나가야 할것이다.

지역어의 조사 원칙과 조사 방법

최 명 옥(서울대학교 국어국문학과 교수)

1. 서론

이 글은 지역어의 조사 원칙과 조사 방법에 대해 논의함을 목적으로 한다. 그러기 위해서 '지역어'와 '지역어의 조사'에 대한 개념을 분명히 할 필요가 있으며, 지역어의 조사 목적과 그 목적을 달성하기 위한 지역어 조사의 고려 사항을 명시할 필요가 있다.

'지역어'란 어떤 지역에서 사용되는 한 언어의 하위어를 말하며 독립된 언어 체계를 가지는 한 언어의 변종인 '방언'(방언 구획에 의해 결정된)과 구별된다. 그리고 '지역어의 조사'란 일반적으로 해당 지역어에 대한 언어학적 조사를 말하지만 넓게는 무형 문화재로서의 해당 지역어를 조사하는 것이다.

해당 지역어에 대한 언어학적 조사를 위한 지역어의 조사 목적은 언어 체계를 구성하는 요소(음운, 문법, 어휘)를 중심으로 해당 지역어가 언어학적으로 다른 지역어와 어떤 공통성과 특수성을 가지는가를 밝힐 수 있는 언어 자료를 조사하는 것이다. 그러므로 이러한 지역어의 조사 목적은 해당 지역어를 하나의 방언으로 가정함을 전제한다. 한편 무형 문화재로서의 지역어를 조사하는 목적은 해당 지역의 언어문화와 언어 문화 이외의 그 지역의 전통문화를 밝힐 수 있는 언어 자료를 조사하는 것이다.

이러한 조사 목적에 부합하는 지역어를 조사하기 위해서는 다음과 같은 사항이 고려되어야 한다. 즉 ① 조사 지점, ② 제보자, ③ 조사자, ④ 조사 항목, ⑤ 질문, ⑥ 전사(轉寫)와 녹음(錄音) 등이다.

따라서 이 글은 위에 제시된 여섯 개의 사항을 중심으로 지역어의 조사 원칙과 조사 방법에 대한 논의로 구성된다.

2. 지역어의 조사 원칙

하나의 방언을 전제로 어떤 지역어를 조사하기 위해서는 ① 조사 지점, ② 제보자, ③ 조사자, ④ 조사 항목, ⑤ 질문, ⑥ 전사(轉寫)와 녹음(錄音)이라는 여섯 개의 사항이 고려되어야 한다. 먼저 지역어 조사를 위해 ‘조사 지점’을 첫 번째 고려 사항으로 한 것은 조사의 대상이 지역어이기 때문이다. 조사 대상이 되는 지역어가 독립된 방언임을 전제로 할 때, 그러한 지역어는 선정되는 조사 지점과 밀접한 관계를 가진다. 상업이 성한 지역을 선정한다거나 새로 개발된 지역을 선정한다면, 그러한 지역의 언어는 독립된 언어 체계를 가진 방언이 될 수 없기 때문이다.

다음으로 ‘제보자’가 지역어 조사를 위한 고려 사항이 되는 것은 선정된 제보자가 어떤 사람인가에 따라서 조사된 언어의 성격이 달라질 수 있기 때문이다. 제보자가 그 지역의 출신이 아니라든가 외지 출입이 잦은 사람이라면 그러한 제보자를 통해 수집된 언어 자료는 그 지역의 전형적인 언어 자료가 될 수 없다.

다음으로 ‘조사자’가 지역어 조사를 위한 고려 사항이 되는 이유는 누구나 지역어 조사를 할 수 있는 것이 아니기 때문이다. 위에 제시된 여섯 개의 고려 사항 중 조사자를 제외한 것들은 모두 조사자의 책임 아래 선정되고 시행된다. 그러므로 조사 지점, 제보자, 조사 항목의 선정 능력이나 적합한 질문 능력이나 제보자의 말에 대한 정확한 전사 능력이 없는 조사자는 조사 목적에 부합하는 지역어를 조사할 수 없다.

또 ‘조사 항목’이 지역어 조사를 위한 고려 사항이 되는 이유는 그것이 조사하려는 지역어의 내용이 되기 때문이다. 어떤 내용으로 구성되는 지역어를 조사할 것인가. 어휘 중심의 지역어를 조사할 것인가. 문법 형태 중심의 지역어를 조사할 것인가. 활용이나 곡용이 포함되는 지역어를 조사할 것인가. 음운론적 연구는 물론 형태론적 연구와 통사론적 연구가 가능한 지역어를 조사할 것인가. 언어학적 연구를 위한 지역어를 조사할 것인가. 무형 문화재로서의 지역어를 조사할 것인가. 이러한 질문에 적합한 지역

어는 선정된 조사 항목에 의해 결정된다.

한편 '질문'이 지역어 조사를 위한 고려 사항이 되는 이유는 질문 형식과 질문 방법에 따라 제보자가 질문에 대한 이해를 할 수도 있고 하지 못할 수도 있으며, 조사된 지역어가 실제와 같을 수도 있고 다를 수도 있기 때문이다. "소나무 잎을 갹아먹고 사는 벌레를 무엇이라고 합니까?"라고 질문하면 제보자가 질문의 내용을 잘 이해할 수 있다. 그 결과 '송충이'에 대한 그 지역어를 조사할 수 있다. 그러나 "잘게 부스러진 물건을 무엇이라고 합니까?"라고 질문하면 '물건'이라는 단어 때문에 제보자가 질문 중의 '무엇'이 '부스러기'라는 것을 이해하기 어렵다. 이 경우에는 해당 지역어를 조사할 수 없다. 또 '파랭이'나 '포리'라고 하는 지역에서, 표준어 '파리'에 대한 지역어를 조사하는 경우, 조사자가 "파리를 여기서는 무엇이라고 합니까?"라고 질문하면 제보자의 답은 대개 질문에 있는 대로 '파리'가 된다. 이러한 답은 제보자들이 평소에 자기들이 사용하는 말을 표준어에 비하여 품위가 낮다고 생각하고 있기 때문에 나오는 것인데, 이 경우에 제보자의 답은 실제하고 다르다.

지역어 조사를 위한 고려 사항 중에서 마지막으로 거론될 것은 '전사(轉寫)'와 녹음(錄音)이다. 여기서 말하는 '전사'란 체계적이고 일관된 방식으로 언어음을 기록하는 방법이며 '녹음'은 제보자의 음성을 녹음테이프에 담는 것이다. 앞에서 논의된 다섯 개의 사항도 중요하지만, '전사'는 지역어 조사의 최종 결과라는 점에서 가장 중요한 사항이라고 할 수 있다. 최종으로 '전사'된 것이 실제와 같지 않다면 다른 사항이 아무리 완벽한 것이었다고 해도 그것은 무의미하기 때문이다. 표준어 '꺾어서'에 해당하는 지역어를 제보자가 [꺽'꺼서(k'ɕʰŋk'ɕsɕ)]라고 했는데, 그것을 [꺽'꺼서(k'ɕʰk'ɕsɕ)]나 [꺽'꺼'서(k'ɕʰŋk'ɕ'sɕ)]나 [꺽'꺼'서(k'ɕʰŋk'ɕ'sɕ)]라고 전사하거나 표준어 '고양이'에 해당하는 지역어를 제보자가 [에'앵구(E'Engu)]라고 했는데, 그것을 [앵'구(E'ngu)]나 [에'앵'구(E'E'ngu)]라고 전사한다면, 표준어 '붓-(注)'의 활용형 '붓고'에 해당하는 지역어를 제보자가 [부'코(pu'kho)]라고 했는데, 그것을 [부'꼬(pu'k'o)]라고 전사한다면, 그것들은 모두 실제와 일치하지 않으므로 잘못이 되고 만다.

'녹음'은 이러한 잘못을 확인할 수 있으며 전사(轉寫)에서 제외되는 제보자의 음성을 모두 기록하여 필요할 때에는 언제나 되풀이하여 들을 수 있으며 사실 여부를 확인할

수 있다는 점에서 중요하다. 그런데 녹음기가 작동을 하지 않는다거나 녹음기는 작동 하여도 녹음테이프가 돌아가지 않아 녹음을 하지 못한다면, 녹음되는 소리를 너무 작 게 했거나 주변의 소음이 너무 커서 녹음된 음성을 제대로 들을 수가 없다면, 그러한 '녹음'은 아무 소용이 없게 된다. 또 녹음테이프를 아끼기 위해 조사 항목에 대한 제보 자의 답만 녹음하거나 휴식 시간에는 녹음기를 꺼 버린다면, 조사 항목 외의 중요한 자료를 다 놓치게 될 것이다.

따라서 이상의 고려 사항을 포함하는 지역어 조사의 원칙은 다음과 같아야 한다. 첫째, '조사 지점'은 대상 지역의 언어 실체를 대표하는 즉 대상 지역의 언어를 전형적으 로 보여 주는 지점이어야 한다. 둘째, 제보자는 전형적인 그 지역어의 사용자여야 한 다. 셋째, 조사자는 전형적인 해당 지역어를 조사하고 전사할 수 있는 사람이어야 한 다. 넷째, 조사 항목은 한편으로는 음운 면, 문법 면, 어휘 면에서 해당 지역어의 특수 성과 다른 지역어와의 공통성을 보여 줄 수 있어야 하며, 다른 한편으로는 단어나 형 태는 물론 문장 단위도 조사될 수 있는 것이어야 한다. 다섯째, 질문은 해당 지역어의 전형적인 어형, 어법이나 문장 등이 조사될 수 있도록 하는 것이어야 한다. 여섯째, 전 사는 제보자의 제보 내용을 충실하게 반영하도록 해야 하며 녹음은 조사 중에 하는 제 보자의 말 전체를 명료하게 들을 수 있도록 해야 한다.

3. 지역어의 조사 방법

지역어의 조사 방법이란 위에 제시된 여섯 개의 고려 사항에 적용되는 원칙을 지키 도록 하는 구체적인 내용이다. 따라서 이 장에서의 논의 내용은 여섯 개의 고려 사항 에 부가된 원칙의 구체화에 대한 것이다. 그런데 그들 고려 사항에 부가된 원칙의 구 체화는 조사나 연구 목적에 따라 달라진다. 예를 들어, 사회 방언학적인 조사나 연구를 목적으로 한다면, 조사 지점은 사회 계층마다 한 지점씩 정해야 하며, 제보자는 계층마 다 노년층과 청소년층으로 나누어 한 사람씩 선정하고, 조사 항목은 계층 간의 차이를 보여 줄 수 있는 것으로 선정해야 한다. 그리고 방언 구획을 위한 조사나 연구를 목적 으로 한다면, 먼저 군 단위로 한 지점씩 정하고, 제보자는 노년층으로 하며, 조사 항목

은 음운 면, 어휘 면, 문법 면을 포함하여 방언 체계를 보여 줄 수 있는 것으로 선정해야 할 것이다. 그리고 그 결과 방언 경계가 예상되는 곳에서는 방언 경계를 밝힐 수 있도록 좀 더 지점을 많이 정하여야 할 것이다.

이 장에서는 소멸 위기에 있는 전통 지역어의 조사나 연구를 목적으로 논의를 전개하기로 한다.

3.1. 조사 지점

전통 지역어의 조사나 연구를 목적으로 하는 경우에, 조사 지점은 다음 몇 가지 기준에 따라 선정될 수 있다. 도(道)를 단위로 하여 지역어를 조사하는 경우에, 먼저 도 단위의 방언 구획이 밝혀진 경우에는, 일차 하위 방언 중의 하나를 택하고 그 하위 방언 구획 내에 목(牧)과 도호부(都護部), 군(郡), 현(縣) 등이 있었던 곳이 있으면 그곳을 조사 지점으로 정한다. 대개 행정의 중심 지역은 언어의 중심 지역이 되기 때문이다. 다음으로 해당 방언 구획 내에 행정의 중심 지역이 없는 경우에는, 논이 많고 큰 저수지가 있는 마을을 조사 지점으로 정한다. 이런 마을은 오래전에 형성되어 그곳에서 사용되는 언어도 역사가 길 것이 예상되기 때문이다.

그러나 가끔은 이러한 예상이 빗나가는 경우가 있다. 그러므로 확실하게 적합한 조사 지점을 정하기 위해서는, 해당 지역의 군청이나 면사무소의 담당 직원을 찾아 오래전에 형성된 농촌 마을로 주민의 이동이 적고 노년층이 많으며 그 지역의 언어를 대표할 수 있는 곳을 안내해 주도록 부탁하는 것이 좋다.

3.2. 제보자

이 경우의 제보자는 해당 지역의 전형적인 말을 사용하는 사람이어야 한다. 그러한 사람은 안정된 직업을 가진 그 지역의 토박이로서 말하기를 좋아하고 언어에 대해 민감하며 발음 기관이 양호하고 건강해야 한다. 안정된 직업은 일반적으로 농업을 뜻한다. 오랫동안 타향에서 생활을 했거나 외지를 자주 왕래하는 사람은 다른 지역어의 간섭에 의해 전형적인 그 지역어를 사용하지 못하는 경우가 흔하다. 그러나 농업에 종사

하는 사람은 외부로 왕래하는 경우가 적으므로 다른 직업에 종사하는 사람보다는 비교적 순수한 자기 지역어를 사용한다.

그 지역의 토박이란 대개 3대 이상 그 지역에 거주해 오는 집안의 사람을 말한다. 이 경우에도 어머니나 할머니 등 부인들이 모두 그 지역에서 출생한 사람이어야 한다. 아이들은 출생하여 혼자 걸어 다니기 전까지는 어머니와 함께 있으면서 어머니의 말을 듣고 언어를 습득한다. 그러므로 어머니나 할머니 등 부인들이 다른 지역의 출신인 경우에는 그 지역어의 영향을 받게 되므로 그런 사람은 순수한 해당 지역어의 사용자라고 할 수 없다.

그리고 말하기를 좋아한다는 것은 능동적으로 말하는 것을 뜻하며 일방적으로 자기 생각만 계속하여 늘어놓는 것을 뜻하지 않는다. 그리고 언어에 대해 민감하다는 것은 조사 항목에 대한 답만 아니라 그와 관련된 다른 사실까지도 알려 줄 만큼 자기 언어에 대해 잘 알고 있다는 것을 의미한다.

한편 제보자의 발음 기관이 양호해야 한다는 것은 지역어형을 정확하게 기록하고 녹음해야 하기 때문이다. 제보자의 발음 기관이 양호하지 못하다면, 제보자는 그 지역어의 언어 사실을 정확하게 알려 줄 수 없기 때문이다. 그리고 제보자가 건강해야 한다는 것은 지역어의 조사가 오랜 시간과 기간에 걸쳐 이루어지기 때문이다.

3.3. 조사자

지역어의 조사에 적합한 조사자는 개인적 임기응변이 있고 상대방의 심리 파악 능력이 있어야 하며, 언어학적 지식과 조사 항목에 대한 적절한 질문 방법을 알고 있어야 한다.

개인적 임기응변이란 조사 중에 부닥치는 어려움을 즉시 해결해 내는 것을 말한다. 어떤 조사 항목에 대한 설명을 몇 번이나 듣고도 적합한 답을 하지 못할 때, 말을 바꾸어 다른 조사 항목으로 넘어간다거나 조사 항목에 대한 사전적 설명으로는 제보자를 이해시킬 수 없을 때, 조사자가 엉뚱한 답을 하여 제보자로부터 정답을 끌어내게 하는 것 등이 그에 해당한다. 예를 들어 표준어 ‘앉다(坐)’나 ‘씹다(齧嚼)’에 대한 지역어형을

조사하려고 할 때에, ‘앉다’나 ‘썩다’에 대한 사전적 설명보다는 “‘서다’의 반대는 무엇입니까?”나 “밥은 마십니까?”라고 물어서 제보자로부터 답을 쉽게 끌어낼 수 있다.

조사자가 상대방의 심리를 파악하는 능력을 가져야 한다는 것은 제보자가 곤란한 처지에 있을 경우에 그것을 빨리 감지하여 그 곤란을 해결해 줄 수 있어야 하기 때문이다. 장시간 조사하는 도중에 다리가 저리다거나 피곤하여 제보자가 좀 쉬고 싶을 때에 조사자가 잠시 휴식을 제안한다거나 더 이상 조사를 진행할 상황이 아닐 경우에는 비록 조사한 시간이 얼마 되지 않았다 하더라도 조사자가 조사를 중단하는 것 등이 그에 해당한다. 이렇게 조사자가 제보자의 곤란을 적절하게 해결해 준다면, 제보자는 더욱 친절하게 적극적으로 조사에 응해 줄 것이다.

그리고 조사자가 언어학적 지식을 갖추고 있어야 한다는 것은 조사자가 조사되었거나 조사되는 언어 자료를 분석할 수 있으며 그것들에 대한 공시적·통시적 해석을 할 수 있고 필요에 따라서는 조사 항목을 추가할 수 있어야 하기 때문이다. 제보자가 표준어 ‘바꾸-(換), 가두-(囚)’ 등이 어미 ‘-고, -지, -더라, -으면, -어도’와 통합할 때의 지역어형을 각각 [바꾸^ㄴ코, 바꾸^ㄴ치, 바꾸터^ㄴ라, 바꾸^ㄴ우머, 바까^ㄴ아도]와 [가두^ㄴ코, 가두^ㄴ치, 가두터^ㄴ라, 가두^ㄴ우머, 가다^ㄴ아도]라고 하면서, 표준어 ‘나누-(分)’가 동일 어미와 통합할 때의 지역어형은 [농^ㄴ구고, 농^ㄴ구지, 농^ㄴ구더라, 농^ㄴ구머, 농^ㄴ가도]라고 한다면, 조사자는 그 두 유형의 활용형의 차이를 즉시 분석할 수 있어야 한다. 즉 표준어에서 2음절을 가진 동사 어간으로서 어간의 말 음절 구조가 ‘자음소’와 ‘우’인 것 즉 Cu인 것은 해당 지역어에 ‘XCuh-’와 ‘XCu-’로 존재한다.

이러한 분석을 통해서 그들 동사의 원래 어간형은 모두 XCu-였지만, 어간 성조가 LH인 것은 해당 지역어에서 어간 말에 ‘ㅎ’가 첨가되어 XCuh-로 변했고 어간 성조가 HL인 것은 해당 지역어에 그대로 남아 있다. 조사자는 이러한 통시적 해석을 한 다음 동일한 2음절 어간 동사로서 어간의 말 음절 구조가 ‘우’인 것 즉 u인 것은 해당 지역어에 어떤 어간형으로 존재하며 그 활용형은 어떠한가를 조사해야 한다. 그래야 표준어에서 ‘우’로 끝나는 2음절 어간 동사에 대한 해당 지역어의 체계적 조사가 가능하게 된다.

그리고 조사자가 조사 항목에 대한 적절한 질문 방법을 알고 있어야 한다는 것은 제

보자가 자연스러운 조사 대상 지역어형을 말할 수 있도록 해야 하기 때문이다. 구체적인 조사 대상이 있는 경우와 없는 경우에 따라 질문이 달라야 할 것이며 해당 주제와 관련된 많은 어휘를 알고자 할 경우와 해당 조사 항목에 대한 지역어형이 가지고 있는 의미나 용법을 알고자 할 경우에 따라 질문이 달라야 할 것이다. 그리고 곡용형이나 활용형을 알고자 할 경우와 제보자가 조사 항목을 이해하지 못하기 때문에 제보자에게 어떤 암시를 줄 경우에 따라서도 질문이 달라야 할 것이다.

이 밖에 조사자는 조사되는 자료의 참과 거짓을 판정하고 조사할 주제를 풍부하게 할 뿐 아니라 제보자와의 친밀도를 강화하기 위해서 조사 대상 지역에 대한 지식을 갖추고 있어야 한다. 그러기 위해서 조사 대상 지역어에 대한 기존의 조사 자료를 검토하여 그 특징과 성격을 파악하고 군지(郡誌)나 읍지(邑誌) 등을 통해서 그곳에서 출생한 유명 인사나 그곳의 유적이거나 유물과 민속에 대해 파악해야 한다. 그리고 길고 지루한 조사 시간 중에 제보자가 즐겁게 조사에 응할 수 있도록 하기 위해 배우(탈런트)적인 기질도 지니고 있어야 한다.

3.4. 조사 항목

지역어 조사를 위한 조사 항목은 조사 목적에 따라 차이가 있다. 다시 말하면, 언어학적 연구를 위한 조사인가 무형 문화재로서의 지역어를 보존하기 위한 조사인가에 따라 조사 항목에 차이가 있게 된다.

먼저 언어학적 연구를 위한 조사 항목은 그 지역어의 언어 체계를 파악할 수 있도록 조사 항목이 구성되어야 한다. 즉 음운 면, 문법 면, 어휘 면의 조사 항목이 그것이다. 이것들은 각각 공시적 연구를 위한 것과 통시적 연구를 위한 것으로 구분할 수 있다. 그러나 여기서는 공시적 연구를 위한 것으로 그 범위를 한정하기로 한다.

음운 면의 조사 항목은 ㉠ 음운 체계를 조사하기 위한 것과 ㉡ 음운 과정을 조사하기 위한 것으로 구성되어야 한다. ㉠을 위한 조사 항목은 자음소 체계, 모음소 체계(여기에는 단모음소 체계를 조사하기 위한 것과 이중 모음소 체계를 조사하기 위한 것이 포함됨), 유음소 체계, 활음소 체계를 조사하기 위한 것으로 구성되어야 하며 ㉡을 위

한 조사 항목은 곡용 조사 항목과 활용 조사 항목으로 구성되어야 한다.

문법 면의 조사 항목은 ㉠ 문법 형태소를 조사하기 위한 것과 ㉡ 통사적인 것을 조사하기 위한 것으로 구성되어야 한다. ㉠을 위한 조사 항목은 곡용 어미를 조사하기 위한 것과 활용 어미를 조사하기 위한 것으로 구성되어야 한다. 활용 어미를 조사하기 위한 것은 다시 종결 어미, 연결 어미, 선어말 어미, 전성 어미를 조사하기 위한 것으로 구성되어야 한다. ㉡을 위한 조사 항목은 존비법, 부정법, 피동법, 사동법 등을 조사하기 위한 것으로 구성되어야 한다.

어휘 면의 조사 항목은 그 기준이 무엇이냐에 따라 여러 가지 하위 항목으로 구성될 수 있다. 단어의 문법적 성질을 기준으로 하면, 명사, 동사, 부사, 관형사 등 품사별 조사 항목으로 구성되어야 하며 의미를 기준으로 하면, 농사, 음식, 가옥, 의복, 민속, 인체, 육아(育兒), 인륜(人倫), 동물, 식물, 자연 등의 조사 항목으로 구성되어야 한다.

다음으로 무형 문화재로서의 지역어를 조사하기 위한 조사 항목은 해당 지역어의 언어 체계를 파악할 수 있는 것과 음식, 복식, 주거, 민속 등 그 지역의 전통문화를 파악할 수 있는 것으로 구성되어야 한다. 전통 지역어는 그 지역의 문화를 형성하고 표현하는 매체이다. 일반적으로 지역어를 무형 문화재라고 하는 이유가 거기에 있다. 그러므로 무형 문화재로서의 지역어를 보존하기 위해서는 그 지역의 전통문화를 파악할 수 있는 다양한 주제들이 조사 항목에 포함되어야 한다.

3.5. 질문

조사 항목에 대한 해당 지역어형을 알아내기 위해서 조사자는 제보자에게 질문을 한다. 이 경우에 조사자가 명심해야 몇 가지가 있다. 첫째, 조사자와 제보자가 충분히 친숙한 상태가 아닌 경우에는 제보자가 해당 지역어형을 제대로 말하지 않는다는 것이다. 그러므로 제보자는 조급하게 조사 항목에 대한 조사를 해서는 안 된다. 그러기 전에, 그 지역의 전통문화를 파악할 수 있는 다양한 주제들에 대해서 제보자의 설명을 듣는 것으로 조사를 시작하는 것이 좋다.

이런 순서로 조사를 하는 데에는 몇 가지 이점이 있다. 제보자와 상당한 시간에 걸

처 대담하는 동안에 두 사람이 친숙하게 됨으로써 나중에 하게 될 조사 항목에 대해서 제보자로부터 실제의 답을 얻을 수 있다는 것, 그리고 제보자의 설명에서 문장 단위의 지역어를 조사할 수 있다는 것, 그 지역어에 대한 음운 면, 문법 면, 어휘 면의 일부 조사 항목에 대한 답을 미리 조사할 수 있다는 것 등이 그것이다.

둘째, 조사자는 제보자를 스승과 같이 대해야 한다는 것이다. 제보자가 학력이 낮거나 초라하다고 하여 함부로 대한다면, 제보자는 적극적으로 조사에 임하려고 하지 않을 것이며 심한 경우에는 조사를 거부할 것이다.

셋째, 어휘 항목을 조사할 때에는 조사 항목이 되는 단어를 절대로 조사자가 먼저 말해서는 안 된다는 것이다. 제보자들은 대개 자신들이 사용하고 있는 지역어가 표준어나 중앙어에 비하여 품위가 떨어진다는 생각을 가지고 있다. 그러므로 조사자가 먼저 조사 항목의 단어를 말한다면, 제보자는 자기 지역에서도 그 단어를 사용한다고 답하는 것이 일반적이다. 예를 들어, 조사자는 “‘파리(蠅)’를 여기서는 무어라고 합니까?”, “혹시 ‘포리’라고 하지 않습니까?”, “‘포리’라고 합니까 ‘파리’라고 합니까?”, “‘포리’라고 하지요?” 등과 같은 질문을 해서는 안 된다. 그렇게 되면, 옳은 지역어형을 조사할 수 없다.

넷째, 제보자가 조사 항목에 대해 알지 못하거나 잘못된 답을 할 경우에는 설명을 추가하여 두 번 정도 질문을 하되, 그래도 옳은 답을 얻지 못하면 다음 조사 항목으로 넘어가야 한다는 것이다. 한 가지 조사 항목에 대해 너무 시간을 끌면 지루해져서 제보자가 조사에 대한 흥미를 잃어버리게 되기 때문이다.

다섯째, 조사자와 제보자는 전사에 필요한 상을 가운데 두고 마주 앉아야 한다. 이렇게 앉을 때에 조사자는 제보자가 /집 ㅁ 은/ 등에서 ‘으’를 원순 모음으로 발음하는지 ‘위, 외’ 등을 단모음으로 발음하는지 이중 모음으로 발음하는지를 확인할 수 있다.

이상 몇 가지 사실을 기억하면서 질문에 대해 논의하기로 한다. 질문은 조사 항목이 어떤 것이냐에 따라서 그 방법이 달라진다. 조사 항목이 전통문화 파악을 위한 것이라면, 그 내용을 자세히 알 수 있도록 질문해야 한다. 예를 들어, 주제가 ‘된장 담는 것’이라면, 자료는 무엇이며, 그 자료를 어떻게 하여 ‘메주’를 만드는가, 메주는 어떻게 띄우며 메주가 뜰 때에 벌어진 사이에 나는 머리털 같은 파란 것은 무엇이며, 띄운 메주를

어떻게 하여 ‘된장’으로 만드는가. 이러한 질문이 계속되어야 한다. 그러려면 조사자가 된장 담는 절차와 방법에 대해서 미리 알고 있지 않으면 안 된다.

그러나 조사 항목이 어휘에 대한 것이라면, 질문의 방법이 달라져야 한다. ① 조사 항목이 조사 장소에 있는 경우에, 질문 방법은 “이것은 무엇입니까?”, “저것은 무엇입니까?”와 같다. 이러한 질문 방법을 명명법(命名法)이라 한다. 이것은 가장 간단하면서 제보자가 처음으로 조사 항목에 대한 답을 하게 하는 질문 방법이다. 따라서 조사 항목에 해당하는 것이 추상적인 것이거나 조사 장소에 없는 것이라도 그에 대한 ‘그림’이나 ‘사진’을 준비하여 그것을 보여 주고 이 방법으로 질문을 한다면 제보자로부터 쉽게 답을 얻을 수 있다.

② 조사 항목이 조사 장소에 없는 경우에, 질문 방법은 “자기 아버지의 아버지를 부를 때에 무엇라고 합니까?(표준어 ‘할아버지’에 대한 지역어형을 물을 때)”, “며느리가 시아버지를 부를 때에 무엇라고 합니까?(표준어 ‘아버님’에 대한 지역어형을 물을 때)”와 같다. 이러한 질문 방법을 완성법(完成法)이라 한다. 조사 항목에 대한 설명을 제시하여 제보자가 답하게 한다는 것이 이 질문 방법의 핵심이다.

③ 조사 항목이 특정 단어가 아니라 어떤 주제와 관련된 많은 단어를 찾는 것인 경우에, 질문 방법은 “‘콩’을 가지고 만들 수 있는 것들로는 어떤 것이 있습니까?”, “봄에는 밭에 어떤 채소들을 심습니까?”와 같다. 이러한 질문 방법을 열거법(列擧法)이라 한다.

④ 조사 항목이 곡용형이나 활용형인 경우에, 질문 방법은 “두 아이가 있는데 내가 보니 이 아이도 물을 어찌하고 저 아이도 물을 어찌하더라. 너는 절대로 물을 어찌하면 안 된다. 애야, 이 물은 네가 다 어찌해라.”(‘마시-(飲)’의 활용을 조사할 때)”와 같다. 이러한 질문 방법을 대치법(代置法)이라 한다. 이 경우에 조사자는 먼저 조사 항목에 대한 지역어형을 조사해야 한다. 조사 항목이 ‘마시-(飲)’인 경우에, 조사자는 제보자에게 “밥은 씹어 먹는데, 물은 어찌합니까? 물도 씹어 먹습니까?”하고 물으면 제보자는 쉽게 그에 대한 지역어형을 답할 것이다. 그 다음에 제시된 것과 같은 질문을 하여 ‘마시-’의 활용형을 조사한다.

그런데 이러한 질문 방법은 제보자에게 익숙하지 않기 때문에 미리 답하는 방법을

교육해야 한다. 예를 들어, ‘막-(防)’의 활용형을 조사할 경우를 보자. 먼저 제보자와의 담화에서 표준어 어미에 대한 지역어형을 확인한다. 다음에 표준어 어간에 대한 지역어형을 조사한다. “구멍에서 물이 새는데 물을 안 새게 하려면 구멍을 어떻게 해야 합니까?” 제보자가 “ 막아야지.”라고 한다면, 표준어 ‘막-’에 대한 지역어형도 ‘막-’임을 알게 된다. 다음으로 어간이 동사인 경우에는 ‘어찌하-’를 사용하고 어간이 형용사인 경우에는 ‘어떠하-’를 사용하여 어간과 지역어의 어미가 통합하는 적절한 문장을 만들어 제보자가 지역어 활용형을 답하게 한다.

표준어 어미 ‘-으니까, -고, -더라, -으면, -어라’에 대한 지역어형이 각각 ‘-응게, -고, -더라, -으면, -어라’라면, 어간 ‘막-’과 지역어 어미가 통합하는 적절한 문장을 제시한다. “내가 물이 새는 구멍을 어찌항게 이 아이도 구멍을 어찌하고 저 아이도 구멍을 어찌하더라. 너는 절대로 구멍을 어찌하면 안 된다. 애야, 네가 구멍을 어찌해라.” 조사자는 제보자에게 “‘어찌하-’ 대신 ‘막-’을 넣어서 말해 주십시오. (내가 물이 새는 구멍을) <어찌항게> 대신 <막응게>라고 말해 주시고, (이 아이도 구멍을) <어찌하고> 대신 <막꼬>라고 말해 주십시오. 그리고 (저 아이도 구멍을) <어찌하더라> 대신 <막떠라>라고 말해 주시고 (너는 절대로 구멍을) <어찌하면> 대신 <막으면>이라고 말해 주십시오. 그러면 (애야, 네가 구멍을) <어찌해라> 대신 어떻게 말하겠습니까?” 이 경우에 답을 제대로 하면 같은 식으로 예를 들어 연습을 하고 그렇지 않으면 조사자가 답을 알려 주고 질문을 이해하여 옳은 답을 할 때까지 연습을 되풀이해야 한다.

곡용형의 조사도 활용형의 조사와 같은 방식으로 하되, 반드시 ‘-이, -에서, -을/를’이 연결된 형을 조사해야 한다. 지역에 따라서는 표준어 ‘술(鼎)’ 등의 곡용형이 [소시, 소술, 소세서]인 것도 있고 [소시, 소술, 소테서]인 것도 있기 때문이다.

⑤ 제보자가 조사 항목을 이해하지 못하거나 조사 항목을 이해해도 해당 지역어형을 잘 생각해 내지 못할 경우에, 질문 방법은 “‘이(虱)’의 ‘알’을 무엇이라고 합니까? 혹시 ‘서캐, 켜가래, 켜가리, 헤에기’라고 하지 않습니까?”, “‘각자’ 또는 ‘제각기’에 해당하는 이곳 말은 무엇입니까? 혹시 ‘지주굼 지주굼, 지주움 지주움, 시시마꿈, 시제마꿈’이라고 하지 않습니까?”와 같다. 이러한 질문 방법을 제시법(提示法) 또는 암시법(暗示法)이라 하는데, 이 질문 방법은 위의 여러 질문 방법을 사용하여도 물음에 대한 답을

얻지 못할 때에, 몇 가지 가능한 지역어형을 제시하여 제보자가 그중 어느 하나를 선택하게 하거나 제시되지 않은 지역어형을 생각하게 하는 것이다.

⑥ 조사 항목에 대한 지역어형은 아는데, 그것의 의미나 용법을 모르는 경우에, 질문 방법은 “‘아뻬(표준어 ‘시아버지’에 대한 호칭)은 누가 누구를 부르는 말입니까?”, “‘새뻬(작은 민물 새우)와 ‘징게미(검고 큰 민물 새우)’는 어떻게 다릅니까?”, “‘짚지(젖으로 담은 물 없는 김치)’는 무엇을 말합니까?”와 같다. 이러한 질문 방법을 역질문법(逆質問法)이라 한다.

어떤 조사 항목이든지 위의 여섯 가지 질문 방법으로 질문한다면 해당 지역어형을 알아낼 수 있을 것이다.

3.6. 전사(轉寫)와 녹음(錄音)

3.6.1. 전사

좋은 제보자를 구하고, 그 제보자를 상대로 적절한 질문을 하여, 조사 지역에서 사용되는 전형적인 지역어형을 얻어 내는 것으로서 방언 자료 조사는 일단 성공적이라 할 수 있다. 그러나 여기까지의 성공도, 조사자가 제보자의 발음을 정확하게 듣지 못하거나, 정확하게 들었다고 해도 그것을 정확하게 전사하지 못한다면 그 결과는 실패로 돌아가고 만다. 조사자가 범하게 될 이러한 잘못을 방지하기 위해서 조사자는 조사 기간 동안 자신이 다른 천체(天體)에서 온 사람이라는 마음 자세를 가져야 한다. 달리 말하면, 조사자는 제보자의 입에서 나오는 말을 의미로 받아들이지 말고 음성으로 받아들이야 한다. 이러한 마음 자세를 가지면, 조사자의 선입관 때문에 제보자의 입에서 나오는 말을 의심하거나 잘못 듣는 경우를 예방할 수 있다.

제보자의 말을 전사하는 방법 중 중요한 것은 다음과 같다.

① 제보자의 입에서 나오는 말을 소리 나는 대로 전사한다. 전사는 국제 음성 기호로 하는 것이 원칙이지만, 그러한 기호에 숙달되어 있지 않으면 제보자의 말을 국제 음성 기호로 전사하기는 어렵다. 이런 경우에 가능한 것은 한글로 전사하는 것이다. 예를 들면, 표준어 ‘젓-(潤)’과 어미 ‘-(고, 는다, 더라, 어도)’가 통합할 때의 지역어형을

‘적꼬, 전는다, 절떠라, 저저도’와 같이 전사한다는 말이다.

수집된 지역어 자료는 수집한 사람만이 아니라 그 지역어에 관심을 가지는 있는 그
리고 관심을 가지게 될 모든 사람의 연구 자료로 사용된다. 그러므로 아래의 자료 ㉠
과 같이 지역어 자료를 표준어의 표기법대로 전사한다거나 자료 ㉡과 같이 음성형이
예측될 수 있는 것은 표준어의 표기법대로 전사하고 그렇지 못한 것은 소리대로 전사
한다거나 해서는 안 된다. 음성형의 예측은 민족어에 대한 음운론적 지식이 있는 사람
에게는 가능하겠지만, 그렇지 않은 사람에게는 불가능할 것이다. 따라서 모든 사람에
게 공정한 지역어 자료는 자료 ㉢과 같이 소리대로 전사한 것이다.

-자료-

- ㉠ 큰아부지<큰아버지>는 기운이 시고<세고> 농사일일<을> 잘하시고 음악은 일 없고
<필요 없고>. 큰엄매<큰어머니>는 강썸데 조강지처가 아녀라<아닙니다> <뿌리깊은
나무(1992: 21)>
- ㉡ 그런 걸 일본 싸람이 들어와서 다:: 때려 부시구 공원을 맨들어놔어. 그러구 인제
사:직동에서 영천으루 올라가는 터:널을 맨들었지 않아요? 그 길 맨드느라구 다::
때려 부셔서 그렇게 해 놔어 [:는 장음을 ::는 :보다 더 긴 장음을 나타냄] <국립국어
연구원(1997: 153)>
- ㉢ 가실' 게느 딸' 로' 내' 애눅코, 보' 페느 메' 늘로 내' 애논는다<가을에는 딸을
내어 놓고, 봄에는 며느리를 내어 놓는다>. 가실' 콩' 주근 앵 무' 거도(무' 가
도), 봉' 콩' 주근 잘 몽는' 다<가을 콩죽은 안 먹어도, 봄 콩죽은 잘 먹는다>. <김
주석·최명옥 편저(2001)>

② ‘어’와 ‘으’, ‘에’와 ‘애’가 변별되지 않는 지역어나 ‘위’와 ‘외’가 단모음소이거나 이
중 모음소인 지역어, 그 밖에 표준어 ‘호미’에 대해 [ho̞mi]라고 하는 지역어를 전사할
때에는 <이 지역어에서는 ‘어’와 ‘으’가 구별되지 않는다>거나 <‘위’는 단모음소다>라
거나 <모음소 사이의 ‘o(ŋ)’은 그 앞과 뒤의 모음소를 비모음화시키고 발음되지 않는
다>거나 하는 주(註)를 달아 주는 것이 좋다

③ 한 조사 항목에 대한 지역어형이 여러 개가 있으면 그것들을 모두 기록하고 부등

호(>)로써 사용 빈도를 표시한다. 그리고 고행(古形)인가 신형(新形)인가, 노년층이 사용하는 것인가 청소년층이 사용하는 것인가 등을 구별하여 표시한다.

예. 표준어 ‘아궁이’의 방언형: 부직’(㉠,㉡) < 부석(□□,□□)

④ 단어의 의미나 용법에 설명이 필요할 경우에는 해당 단어의 오른편 어깨에 1, 2, 3 등으로 표시하고 아래 위의 빈칸에 설명을 적어 둔다.

예. 덕석1, 땃방석2: 1. 네모난 것, 2. 둥근 것.

⑤ 부득이하게 제보자를 바꿀 경우에는 제보자 사항을 기록하고(색을 달리할 것), 조사 내용을 명백히 구별한다. 왜냐하면 제보자에 따른 지역어 차이가 있을 수 있기 때문이다. 그럴 때에 각 제보자의 인적 사항이나 자료를 구별해 두지 않는다면, 나중에 그것이 연구 자료로 활용될 경우에 한 자료 안에서 발견되는 그러한 지역어 차이를 설명하기 어렵거나 설명할 수 없을 것이다.

예. 1-35까지(제보자 A), 36 이후(제보자 B). 제보자 A: 강원도에서 4년간 거주함. 공무원 경력 6년. 제보자 B: 경상남도에서 5년간 교사로 근무. 그 기간 이외에는 농업에 종사함.

⑥ 음장 지역어인 경우에는 반드시 음장을 표시하고(예. 자료 ㉠), 성조 지역어인 경우에는 성조 표시를 한다(예. 자료 ㉡). 음장의 경우, 장모음은 모음 다음에 :나 ::(아주 길 때)를 표시하거나 같은 모음을 더 쓰고, 성조의 경우, 고조(高調)는 모음 뒤에 “를 표시하고 상승조(上昇調)는 해당 모음 뒤에 ~를 표시하며 저조(低調)는 아무 표시를 하지 않는다. 또는 음성형 다음에 H(고조)나 R(상승조)나 L(저조)로 표시한다.

예. 음장: ‘사 : 람’ 또는 ‘사아람’ (사람, 人), 성조: ‘말도~’ 또는 ‘말도(LH)’, ‘말 ~ 도~’ 또는 ‘말도(RH)’

3.6.2. 녹음

녹음은 전사한 자료가 정확한가의 여부를 확인하고 조사 장소에서 전사하지 못하는 제보자의 말을 기록하여 조사가 끝난 뒤에 유용하게 사용하기 위해서 절대적으로 필요하다. 그러나 녹음기 조작을 잘못하거나 점점을 소홀히 하는 경우에 녹음이 안 되거나 녹음 상태가 좋지 않아 자료로 사용할 수 없게 되기도 한다. 여기서는 녹음을 위한 녹음기 조작과 점점에 대해 간단히 언급하기로 한다.

① 가능하다면 제보자가 녹음한다는 사실을 모르게 하는 것이 좋다. 녹음하는 사실을 제보자가 알게 되면, 자연스런 지역어를 말하지 않는 것이 일반적이기 때문이다. 제보자가 녹음한다는 사실을 모르게 하기 위해서는 녹음기나 마이크를 보이지 않게 하고 녹음기 자체나 마이크에 달려 있는 '잠시 멈춤(pause) 버튼'을 손으로 조작하여 녹음을 하면 된다.

② 조사자와 제보자가 충분히 친숙해진 다음에는 제보자에게 녹음해야 하는 이유를 설명하고 녹음기를 공개적으로 사용하여 녹음할 수 있다. 이런 경우에도 가능하다면 녹음기를 제보자가 볼 수 없게 하는 것이 좋다.

③ 조사자는 미리 사용할 녹음기의 성능과 각 부분의 기능, 배터리의 작동 시간 등을 확인하여야 하며, 제보자를 만나기 전에 녹음 버튼만 누르면 녹음할 수 있도록 해 두어야 한다. 음성 크기는 중간보다 조금 작게 해 두고 자동 역전(自動逆轉, autoreverse)이 가능하게 해 두는 것이 좋다. 소리를 크게 해 두면 주변의 잡음이 크게 녹음되어 녹음 상태를 나쁘게 하기 때문이다.

④ 마이크는 본체에 부탁시키는 것보다는 선으로 연결되어 있으며 모든 방향의 소리를 받아들일 수 있는 것이 좋다. 마이크에 '잠시 멈춤 버튼'이 있거나 이어폰에 '원격 조정(리모컨)' 장치가 있는 것이 좋다. 이 경우에는 녹음기를 가방 속에 넣어 놓고 손으로 녹음기의 작동을 조정할 수 있기 때문이다. 마이크의 위치는 제보자 가까이 두는 것이 좋으나 마이크의 성능이 좋은 경우에는 그리고 줌 마이크인 경우에는 조사자 가까이 두어도 무방하다.

⑤ 녹음테이프는 90분 동안 녹음할 수 있는 것이 좋다. 대개 90분 정도 조사하고 30분 정도 쉬는 것이 적절하기 때문이다. 녹음테이프는 앞뒤 면을 손바닥에 두세 번 탁탁 친 뒤에 녹음기에 장착한다. 그렇게 하지 않으면 테이프가 서로 붙어서 돌아가지 않는 경우가 있게 된다.

⑥ 녹음이 진행되는 동안 가끔 녹음기가 녹음 상태에 있는가, 테이프가 돌아가고 있는가, 배터리의 전지 소모 상태가 어떠한가를 점검해야 한다.

⑦ 녹음은 특별한 경우가 아닌 한 조사가 진행되는 동안 계속해야 하며 쉬는 동안에도 제보자가 말을 하는 경우에는 계속해야 한다.

4. 결론

지금까지 필자는 지역어의 조사 원칙과 조사 방법에 대해 논의했다. 지역어의 조사는 조사자가 제보자를 상대로 필요한 조사 대상 지역어형을 알아내는 것이지만, 그러기 위해서는 조사 지점, 제보자, 조사자, 조사 항목, 질문, 전사(轉寫)와 녹음(錄音)과 같은 사항들이 고려되어야 한다. 이들 사항들이 모두 고려되어야 지역어의 조사가 가능하게 된다. 그러므로 지역어의 조사 원칙은 이들 사항들에 부과되는 당위성이며 지역어의 조사 방법은 이들 사항들이 지닌 당위성을 구현하는 방법이다.

이 점에서 필자는 지역어의 조사 원칙을 다음과 같이 보았다. 조사 지점은 대상 지역의 언어 실체를 대표하는 즉 대상 지역의 언어를 전형적으로 보여 주는 지점이어야 한다. 제보자는 전형적인 그 지역어의 사용자여야 하고 조사자는 전형적인 해당 지역어를 조사하고 전사할 수 있는 사람이어야 한다. 그리고 조사 항목은 한편으로는 음운면, 문법면, 어휘면에서 해당 지역어의 특수성과 다른 지역어와의 공통성을 보여 줄 수 있어야 하며, 다른 한편으로는 단어나 형태는 물론 문장 단위도 조사될 수 있어야 한다. 또 질문은 해당 지역어의 전형적인 어형, 어법이나 문장 등이 조사될 수 있도록 하는 것이어야 한다. 전사는 제보자의 제보 내용을 충실하게 반영하도록 해야 하며 녹음은 조사 중에 하는 제보자의 말 전체를 명료하게 들을 수 있도록 해야 한다.

이러한 조사 원칙을 구현하기 위한 지역어의 조사 방법은 다음과 같이 보았다. 조사 지점은 과거에 중요 행정 중심 지역이 있었던 곳이나 논이 많고 큰 저수지가 있는 곳으로 한다. 그런 곳이 적합하지 않으면 군청이나 면사무소의 담당 직원에게서 적합한 지역을 추천받는다. 제보자는 안정된 직업을 가진 그 지역의 토박이로서 말하기를 좋아하고 언어에 대해 민감하며 발음 기관이 양호하고 건강해야 한다. 조사자는 개인적 임기응변이 있고 상대방의 심리 파악 능력이 있어야 하며, 언어학적 지식과 조사 항목에 대한 적절한 질문 방법을 알고 있어야 한다.

조사 항목은, 언어학적 연구를 위한 것일 경우, 그 지역어의 언어 체계를 파악할 수 있도록 구성되어야 한다. 무형 문화재로서의 지역어를 조사하는 경우, 해당 지역어의

언어 체계를 파악할 수 있는 것과 음식, 복식, 주거, 민속 등 그 지역의 전통문화를 파악할 수 있는 것으로 구성되어야 한다.

질문은 상황에 따라 방법이 다르다. 조사 항목이 조사 장소에 있는 경우에는 명명법을 사용해야 하며 조사 항목이 조사 장소에 없는 경우에는 완성법을 사용해야 한다. 그리고 조사 항목이 특정 단어가 아니라 어떤 주제와 관련된 많은 단어를 찾는 것인 경우에는 열거법을 사용해야 하고 조사 항목이 곡용형이나 활용형인 경우에는 대치법을 사용해야 한다. 한편 제보자가 조사 항목을 이해하지 못하거나 조사 항목을 이해해도 해당 지역어형이 잘 생각해내지 못할 경우에는 제시법 또는 암시법을 사용해야 하며 조사 항목에 대한 지역어형은 아는데, 그것의 의미나 용법을 모르는 경우에는 역질문법을 사용해야 한다.

전사는 제보자의 입에서 나오는 말을 소리 나는 대로 하되 한 조사 항목에 대한 지역어형이 여러 개가 있으면 그것들을 모두 기록하고 각각에 대한 정보를 표시해야 한다. 단어의 의미나 용법에 설명이 필요할 경우에는 해당 단어의 오른쪽 어깨에 1, 2, 3 등으로 표시하고 아래 위의 빈칸에 설명을 적어 두어야 한다. 제보자를 바꿀 경우에는 제보자 사항을 기록하고(색을 달리할 것), 조사 내용을 명백히 구별해야 한다. 음장 지역어나 성조 지역어인 경우에는 반드시 음장이나 성조를 표시해야 한다.

녹음은 제보자가 모르게 하되 조사자와 제보자가 충분히 친숙해진 다음에는 제보자에게 녹음해야 하는 이유를 설명하고 녹음기를 사용하여 녹음할 수 있다. 조사자는 사용할 녹음기의 성능과 각 부분의 기능, 배터리의 작동 시간 등을 확인해야 하며, 조사 전에 녹음할 준비를 완료해 두어야 한다. 음성 크기는 중간보다 조금 작게 해 두고 자동 역전(自動逆轉, autoreverse)이 가능하게 해 두는 것이 좋다. 마이크는 선으로 연결되어 있는 것으로서 모든 방향의 소리를 받아들일 수 있고 '잠시 멈춤 버튼'이 있는 것이 좋다. 녹음테이프는 90분 동안 녹음할 수 있는 것이 좋다. 녹음테이프는 앞뒤 면을 손바닥에 두세 번 탁탁 친 뒤에 녹음기에 장착해야 한다. 녹음이 진행되는 동안 가끔 녹음 상태를 점검해야 하며 특별한 경우가 아닌 한 조사가 진행되는 동안 또는 쉬는 동안에도 제보자가 말을 하는 경우에는 계속해야 한다.

〈참고 문헌〉

- 국립국어연구원(1997), 서울 토박이말 자료집(1).
- 김병제(1959), 조선어 방언학 개요(상), 평양: 과학원 출판사.
- 김영환(1982), 조선어방언학, 평양: 김일성종합대학출판사.
- 김주석·최명옥 편저(2001), 경주 속담·말 사전, 서울: 한국문화사.
- 뿌리깊은나무(1992), “에이 째한 사람!” 내가 나보고 그라요(민중자서전 20).
- 崔明玉(1980), 慶北東海岸方言研究(民族文化叢書 4), 嶺南大出版部.
- 韓國精神文化研究院(1979), 方言 1.
- 韓國精神文化研究院(1987-1995), 韓國方言資料集 I-IX.
- 한영순(1967), 조선어 방언학, 평양: 김일성종합대학출판사.
- 藤原與一(1962), “第三章 共時方言學,” 方言學, 東京: 三省堂.
- Nida, E. A.(1963), Morphology(Second ed.), Ann Arbor: University of Michigan Press.
- Orton, H.(1960), “An English Dialect Survey: Linguistic Atlas English”, Orbis 9.
- Samarin, W. J.(1967), Field Linguistics, New York: Holt, Rinehart and Winston.

민족어유산으로서의 조선어방언의 조사원칙과 방법

김 성 근(사회과학원 언어학연구소 연구사)

오늘 북과 남의 언어학자들에게는 역사적인 6.15공동선언의 정신을 높이 받들고 우리 인민의 귀중한 민족문화유산인 방언을 모두 조사발굴하여 나라의 문화발전에 적극 기여하며 이 과정을 통하여 서로 협력하고 마음을 합치여 나라의 통일위업을 앞당겨야 할 명예롭고도 성스러운 과업이 나서고 있다. 이 방대한 사업은 어느 한쪽 학자들의 힘만으로는 불가능하며 북과 남의 모든 언어학자들이 적극 참가하여 서로 힘을 합칠 때만이 가능하다. 더우거나 현재 나라가 북과 남으로 갈라져있는 조건에서 어느 한쪽의 힘만으로는 방대한 이 사업을 해내기가 어렵다.

모든 학자들이 다 아는바와 같이 방언은 민족문화의 귀중한 재부이며 언어유산이다. 방언은 조선민족성원들의 오랜 창조적 활동과정에 생겨났으며 그것은 사람들의 사회생활에 적극 이바지하면서 전하여오고 있다. 역사발전과정에 방언도 한자리에 머물러 있는것이 아니라 그것이 사람들의 입말속에 존재하면서 부단히 변이를 일으킨다. 더우거나 과학문화와 정보통신수단이 발전하고 국어교육이 강화되면서 문화어에 의한 방언의 수평화과정이 다그쳐져 방언의 본래의 모습을 변화시키거나 소멸과정을 촉진시키고 있다.

이러한 조건하에서 방언을 잘 조사하고 장악하여야 귀중한 민족문화유산을 영원히 보존하고 발전시킬수 있다.

다른 한편 방언을 잘 조사하고 장악하여야 귀중한 민족문화유산을 영원히 보존하고

발전시킬수 있다.

다른 한편 방언을 조사장악하는것은 방언에 있는 수많은 좋은 말을 찾아 문화어의 어휘구성을 풍부히 하고 민족의 유구성과 단일성을 과학리론적으로 론증하는데도 의의가 있다.

방언조사를 잘하기 위해서는 조사원칙과 방법을 보다 합리적으로 정하여야 한다.

이번에 진행하는 방언조사원칙에서 중요한것은 우선 지난날 조사한것을 령으로 하고 완전히 새롭게 시작하는것이다.

이것은 지금 있는 방언자료가 대부분 30년대지 40년전에 조사한 자료라는것과 관련된다. 그때 조사한 자료는 현대적기술수단이 없이 그냥 조사하였기때문에 정확성이 부족하다. 음성표기에서도 부족점이 많다. 표기를 국제음성학회에서 제정한 발음기호로 표기하지 않고 우리 글자로 표기하였기때문에 표기가 정확하지 못하다.

또한 지금 있는 방언자료가 방언학자가 직접 한것도 있지만 많은것이 간접조사로 진행한것이다. 지방에 있는 교원들을 동원시키거나 대학생들이 방언실습을 하면서 조사한것들이다. 지금의 시점에서 보면 자료의 신빙성을 가늠하기가 힘들다. 실지 방언은 방언학자가 나가 직접 조사하고 기록하여야 정확성을 담보할수 있다.

방언자료에 대한 정보를 줌에 있어서도 지난날 자료에는 도만 밝혀져있는것이 많으며 극상해서 군까지 밝히고 리까지 밝힌것은 없다. 이러한 정보자료를 가지고는 자료의 신빙성을 담보할수 없다.

또한 지난날의 자료는 편파성을 띠고있는것이다. 지난날 많은 경우 동북방언과 룡진방언이 특수하다고 하여 방언조사를 함경남북도에 대하여는 많이 하고 황해남북도나 평안남북도, 강원도에 대한 조사를 소홀히 하였다. 물론 지나날 함경도에 대한 조사를 많이 하였다고하여 이번 조사에서 빼놓아야 할 지역으로는 되지 않는다. 지난날의 조사자료를 백지화하고 새로운 립장과 관점에서 조사를 진행하는것만큼 황해남북도, 평안남북도, 자강도, 량강도, 강원도에 대한 조사와 함께 함경남북도에 대한 조사를 면밀히 하여야 할것이다.

다음으로 방언을 조사하는데서 방언과 관련한 정보를 최대한 장악하는 원칙을 지키는것이다.

여기에는 방언어휘의미의 세분화현상, 문법적형태와 의미의 호상관계, 행정구역의 변화로 생긴 언어변이현상 등 여러가지 언어정보자료를 장악하여야 진전으로 방언조사를 원만히 하였다고 볼수 있다. 특히 문화어에 대응되는 방언어휘와 의미에서 나타나는 미세한 차이, 대상, 현상은 있어도 문화어어휘에는 없고 방언에만 있는 어휘부류들을 조사하여야 한다.

레컨대 평안도지방에서 봄철씨불임을 할 때 조나 기장 같은것을 심어 씨불임이 잘 되도록 하기 위하여 이랑을 쪼고 씨를 뿌린 뒤 사람이 발로 밟아주는 일을 《안자구》라고 하고 결혼후 3일만에 신부가 본가집을 다녀오는것을 《뒤자구》라고 한다. 지금 《안자구》, 《뒤자구》가 사전에 올라있지 않지만 얼마나 좋은 말인가. 이런 어휘들을 많이 찾아내면 문화어어휘를 풍부화시키는데도 좋을것이다.

또한 방언을 조사하는데 있어서 방언구획을 정확히 긋는데 모를 박는것이다.

지금까지 방언구획을 긋는데 있어서 함경도, 량강도는 동북방언, 희령, 온성지방은 룡진방언, 평안도와 자강도, 황해도 북부지역은 서북방언, 강원도는 중부방언이라고 하여왔다. 이런 구분만으로는 방언구획을 정확히 그을수 없다. 행정구역이 변경되고 방언들이 호상 영향을 주어 방언구역계선이 불명확한데가 많다. 룡진지방에도 동북방언이 많이 침투해들어갔으며 광복후 원산, 안변군이 강원도로 넘어오면서 방언의 수평화현상이 일어나 현재 원산이나 안변사람들의 말에서는 함경도방언을 찾아보기가 힘들다. 이러한 현상은 황해도남부지방에서도 나타난다. 보통 지금 황해도 토산, 금천, 배천, 청단, 연안, 웅진으로부터 이남지역은 중부방언이고 그 이북지역은 서북방언이라고 하고있는데 이것도 명확히 구획된 계선이 아니다.

명확한 방언계선을 긋기 위하여서는 이 지역들에 대한 방언조사를 면밀히 하여 등어선을 그어야 한다. 등어선을 정확히 긋는 사업이 품이 들고 어려운 조사과정을 거쳐야 하는 일이지만 방언구역을 정확히 정하기 위한 중요한 공정의 하나이다.

그러면 방언조사방법에 대하여 보기로 하자.

조사방법에서 중요하게 나서는 문제는 조사자가 현지에 나가 직접 조사하는것이다. 이때 조사자는 방언에 대한 지식뿐만아니라 어음론에 대한 일정한 지식이 있어야 한다. 방언과 어음에 대한 학술적지식이 있어야 방언소유자의 말을 듣고 제때에 정확히

판단하여 기록할수 있다.

레컨대 《개구리》를 황해남도 배천지방에서는 《떡자구》라고 하고 황해남도 재령, 룡연, 강령지방에서는 《떡자구》라고 하며 평남도 온천지방에서는 《떡자귀》라고 한다. 《떡자구》, 《떡자구》, 《떡자귀》를 가려내려면 일정한 어음론적지식이 있어야 한다. 이 어음적차이를 직접 들어야 그것을 잘 가려낼수 있다. 만약 조사자가 직접 조사하지 않고 농촌교원들이나 대학생들에게 의뢰해서 간접적으로 한다면 정확히 조사하지 못할뿐아니라 조사한 자료조차도 믿을수 없게 된다.

같은 지방에서도 방언의 변이가 나타날수 있다. 이때는 여러 사람의 말을 통해서 확인하여 전형적인 어휘나 형태를 찾아내야한다. 이 모든것은 조사자가 직접 조사에 참가할 때만이 가능하다.

방언을 조사하는 방법으로서는 다음으로 방언소유자들의 자연발화음을 택하는것이 다. 자연발화음에서만이 진짜 방언이 표현될수 있다. 그렇지 않고 조사자가 《여기서는 노루를 무엇이라고 합니까?》라는 식으로 묻는다면 정확한 방언어휘가 나올수 없다. 자연발화음을 택하여야 《노루》를 《놀가지》라고 한다든가, 《고라니, 노로, 노리, 놀각지, 놀기, 놀개지, 놀개이, 놀갱이》 등 그 지방에서 쓰는 말이 튀어나올수 있다. 자연발화음을 택한다고 하여 그가 말하는대로 그냥 내버려둔다는것을 의미하지 않는다. 방언조사는 방언소유자 2~3명을 모아놓고 그들끼리 말하도록 하면서도 조사자가 요구하는 어휘나 형태들이 튀어나오도록 말을 유도해나가야 한다. 그러면서도 말의 주동은 방언소유자들이 되도록 하여야 한다. 그렇지 않고 조사자가 주동이 되어 이야기가 진행된다면 진정한 의미에서의 방언을 조사하기 힘들다. 왜냐하면 우리 나라에서 70세이하되는 사람들은 광복후 소학교나 중학교에서 정규교육을 받은 사람들이기 때문이다. 그들은 대개 사투리와 문화어를 가릴 일정한 능력을 어느정도 가지고 있다. 자연발화상태에 두지 않으면 문화어로 말하거나 서툰 방언으로 말하게 된다.

방언조사자가 자기가 의도하는 어휘를 수집하기 위하여 그림을 보여줄수도 있다. 그림을 보고 자기가 말하도록 하면 조사자가 요구하는 방언현상을 찾아낼수 있다.

다음으로 방언조사대상자문제다. 조사대상을 50살이상 70살까지를 잡되 절대적으로 거기에 국한시킬 필요는 없다. 왜냐하면 우리 나라 형편에서 60대, 70대도 광복후 소학

교나 중학교에서 정규교육을 받은 사람들이며 문맹자가 없기 때문이다. 50대 이상의 늙은이들을 대상으로 하되 그 고장 토착민으로서 정착하여 산 사람이면 충분하다. 만약 늙은이들의 발음이 똑똑치 않아 조사하기 힘든 경우가 생기면 30대, 40대의 사람을 보조적으로 조사대상자로 잡을수 있을것이다. 지식정도의 측면에서 보면 70대이하나 30대, 40대는 광복후 다 정규교육을 받았다는 측면에서는 같다. 이런 측면에서 중년들도 보조적인 조사대상자로 정할수 있을것이다.

또한 조사방법에서 최소한의 품을 들이면서도 최대한의 성과를 거둘수 있게 조사지점을 잘 선택하는것이다.

조사지점을 잘 선택하는것은 그 지역방언의 특성을 밝혀내기 위한 관건적고리로 된다. 조사지점선택문제가 중요하게 나서는것은 같은 방언구역이라 하여 방언이 다 같은 것이 아니는데 있다.

레컨대 서북방언에서 문화어어휘인 《할머니》에 대응되는 방언은 각이하다.

- 큰마니(평북 신의주, 선천, 벽동, 동림, 삭주, 대관, 정주, 통천, 구성, 운전, 철산, 의주, 태천, 박천, 자강도 위원)
- 클마니(평북 녕변, 구장, 향산, 운산, 대관)
- 클만(평북 의주, 피현)
- 할마이(황남 재령, 신천, 태탄, 황북 사리원, 린산, 평산, 서흥, 신계, 수안, 곡산)
- 할만(평남 통강, 온천, 대동)
- 할만(황남 벽성, 태탄)

보는바와 같이 평북도에서는 대부분 《큰마니》, 《클마니》, 《클만》이라고 하며 황해도에서는 《할마이》, 《할만》이라고 하며 평남도에서는 《할만》이라고 한다.

이와 같은 경우에 신의주, 선천, 벽동, 동림, 삭주, 대관, 정주 등 지역에서 조사한다면 다같이 《큰마니》로 말할것이며 황해도 재령, 신천, 태탄, 사리원, 린산, 평산 등 지역에서는 《할마이》로 말할것이다.

이처럼 같은 방언들이 있는 지역을 조사지점으로 정한다면 공연한 노력만이 낭비하

게 될것이다. 그리하여 이러한 현상을 막고 다양한 방언자료를 얻기 위하여서는 미리 조사지점에 대한 료해사업을 잘하여야 한다.

또한 같은 어음론적현상이라도 서로 달리 나타나는 경우들이 많다.

레컨대 《개,》 《개》, 《셋》, 《넷》을 재령, 태탄, 은룰, 황주, 연탄을 비롯한 황해도지방과 남포, 룡강, 온천, 강서, 성천, 강동을 비롯한 평안도지방에서는 《가:이》, 《거:이》, 《서:이》, 《너:이》라도 하지만 《새》를 《사:이》라고 하는 지방은 벽성, 태탄, 재령을 비롯한 황해도남부지방과 룡강, 온천, 남포를 비롯한 평안도방언 구역에 속하는 지역이다. 강동, 대동, 성천지방에서는 《새》를 《사:이》라고 하지 않는다.

이와 같이 방언의 분포특성을 초보적으로 료해한 조건하에서 조사지점을 정하는 문제가 나선다.

이상에서 방언조사원칙과 방법문제를 제기하였다. 여기에 북과 남, 해외에 있는 여러 방언학자들이 참석한것만큼 더 합리적인 조사원칙과 방법이 있을수 있다고 본다.

여기에 모인 모든 학자들은 같은 민족의 성원인것만큼 힘과 지혜를 한데 모두어 우리 민족의 귀중한 언어문화유산을 조사발굴해나가야 할것이다.

방언조사어휘선정에서 제기되는 문제

오 철 만(사회과학원 언어학연구소 연구사)

방언조사에서는 조사방법과 함께 조사내용문제가 중요하게 제기된다.

방언조사에서의 조사내용은 방언조사를 위한 문화어항목을 무엇으로, 얼마나 정하는가 하는 문제이다. 방언조사어휘항목은 결국 방언에서의 문화어어휘항목이다.

방언조사에서 방언조사를 위한 어휘항목선정은 매우 중요한 의의를 가진다.

방언조사어휘항목을 어떻게 선정하는가에 따라 로력과 시간을 적게 들이면서도 해당 방언구역들의 방언적특성을 원만히 반영하는가 못하는가 하는것이 좌우되기 때문이다.

물론 방언조사에서는 조사되는 어휘항목이 많을수록 수집되는 방언어휘의 량도 많아질수 있고 조사연구결과도 더 정밀하고 정확할것이겠지만 거기에 드는 로력과 시간을 타산하지 않을수 없으며 당면단계에서는 일정한 수의 어휘항목을 가지고도 해당 방언들의 특성이 다 반영되도록 하기만 하면 될것이다.

방언조사를 위한 어휘항목선정은 음운, 어휘, 문법의 매 체계에서의 방언적특성이 충분히 반영되도록 잘 고려하여 선정하여야 한다.

그러자면 지금까지 조사수집한 방언자료들에 대하여 연구해볼 필요가 있다.

물론 지금까지의 연구자료를 령으로 치고 새로운 시점에서 다시 재정립한다는 조사원칙을 내세웠지만 제한된 수의 어휘항목으로 전국의 방언들의 방언적특성을 다 반영하도록 하기 위하여서는 지금까지 조사수집된 방언자료를 검토하여 방언들의 어휘부류별특성, 지역별로 나타나는 특성, 언어학적특성 등을 미리 파악하고 그에 기초하여 어휘항목들을 합리적으로 선정하여야 할것이라고 본다.

우리가 지금까지 수집한 방언자료들을 총 묶어놓은 방언어휘자료집에는 3,000여개의 문화어어휘항목에 1만여개의 방언어휘가 올라있다.

이 방언어휘항목들은 지역적으로 도까지만 밝혀주었으며 모두 29개의 어휘부류로 가르고 모든 어휘항목을 해당 어휘부류에 소속시켜 배열해놓았다.

그것을 구체적으로 본다면

1. 사람과 관련된 부류말(지칭어)(약 170개)
2. 사람의 행동, 성격, 감각, 정서와 관련된 말(약 160개)
3. 사람의 신체와 관련된 말(약 140개)
4. 병과 관련된 말(약 35개)
5. 몸차림과 관련된 말(약 90개)
6. 음식과 관련된 말(약 120개)
7. 곡식과 관련된 말(약 45개)
8. 열매와 관련된 말(약 20개)
9. 남새와 관련된 말(약 20개)
10. 야생식물과 관련된 말(약 125개)
11. 짐승과 관련된 말(약 60개)
12. 새와 관련된 말(약 45개)
13. 물고기와 관련된 말(약 65개)
14. 곤충과 관련된 말(약 50개)
15. 집구조, 건축과 관련된 말(약 95개)
16. 민속 및 놀이와 관련된 말(약 60개)
17. 농사일과 관련된 말(약 50개)
18. 부업생활과 관련된 말(약 80개)
19. 사람이 리용하는 도구, 기구, 기물과 관련된 말(약 275개)
20. 방향, 위치와 관련된 말(약 50개)
21. 지리, 지형지물과 관련된 말(약 100개)
22. 차례, 수량, 계산과 관련된 말(약 45개)
23. 천문, 기상과 관련된 말(약 55개)
24. 계절, 시간과 관련된 말(약 65개)

25. 광물, 금속과 관련된 말(약 5개)

26. 동사(약 530개)

27. 형용사(약 200개)

28. 부사(약 180개)

29. 기타(약 50개)

로 되어있다.

이상의 매 어휘부류들을 고찰해보면 그중에서 사람들의 생활과 가장 밀착되어있는 기초어휘들에 대한 방언어휘들이 빠져있는것을 찾아볼수 있다.

실례로 사람과 관련한 부름말부류에서는 《작은아버지》, 《작은어머니》, 《제수》, 《딸》, 《이모》, 《이모부》 사람의 신체구조와 관련된 항목에서는 《손》, 《눈동자》, 《입》, 《젖》 등에 대한 방언어휘가 올라있지 않다. 음식과 관련한 어휘부류에서는 주식의 하나인 《죽》에 대응한 방언어휘가 빠져있으며 《밥》에 대응하는 방언어휘는 《밥》(함경) 하나만이 올라있다.

이러한 기초어휘들에 대한 방언어휘가 수집되지 않은것은 실지 그에 대응한 방언어휘가 없어서 그랬을수도 있고 또 목적의식적으로 해당한 어휘항목을 설정하고 조사하지 않은 이상 지금까지 수집한 자료들이 자연발생적으로 수집된것으로 하여 조사되지 못하고 빠져나갔을수도 있다.

그러므로 앞으로의 어휘항목선정에서 이러한 기초어휘들의 방언어휘존재유무에 대해 주의를 돌릴 필요가 있겠다고 생각한다.

둘째로, 지역별에 따라서 자연지리적조건, 사회경제적조건과 관련되는 지방적특성에 따라 방언어휘가 나타나는것을 알수 있다.

수집된 방언어휘자료가운데는 함경도지방에서 쓰이는 《재칼이》, 《제리》라는 방언어휘가 있는데 《재칼이》는 감자를 문질러서 가루내는데 리용하는 도구를 말하며 《제리》는 감자를 찌는데 리용하는 나무로 성글게 엮은 도구를 말한다. 이런 도구들은 예로부터 고산지대로서 감자를 많이 재배하여 왔고 또 그 가공방법도 다양하게 발전한 함경도지방에만 고유한것이며 문화어에나 다른 방언에는 이에 대응하는 어휘가 없다.

또 문화어 《닷》에 대응하는 방언어휘는 《달》(함남), 《닥》(평북)으로 나타났고 《닷줄》에 대응하는 방언어휘는 《드제비》(함경)로 나타났으며 돛에서 내려 생기는 주름살을 가리켜 말하는 방언어휘 《내림주름》(함경)도 있는데 이것들은 다 바다를 낀 지대에서나 나타날 수 있지 자강도나 량강도, 황해북도 같은 내륙지대에서 이에 대응한 방언어휘를 찾아낼수 있으리라고 기대하기는 어렵다.

그것은 방언어휘발생의 조건으로 볼 때 옹당한것이다.

사람들이 늘 접촉하고 상대하는 대상현상이라야 거기에 자기들의 생활정서를 반영한 고유한 표식을 계기로 삼은 명명이 지어지게도 되고 해당 어휘를 자기 지방의 고유한 언어적체계속에서 사용하게 되지 그렇지 않은 경우에는 해당한 방언어휘가 산생될 수 없는것이다.

이러한 특성 역시 전국적판도에서 동일한 어휘항목에 의한 방언조사어휘항목선정에서 고려해볼 여지가 있지 않겠는가고 생각한다.

셋째로, 자료집에 오른 방언어휘들을 통하여 지역별로 차이나는 방언들의 전통적인 언어학적특징들이 문화어나 이웃방언의 영향으로 점차 자기의 본래의 특징을 잃어버리고있음을 볼수 있었다.

특히 룡진방언에서 제일 심한데 룡진방언은 동북방언구역안에 들어가있는 방언섬으로써 이웃방언인 동북방언의 영향을 받기가 매우 쉬웠으며 그로 하여 점차 자기의 전통적인 방언적특성을 잃어버리고 있다.

실례로, 룡진방언은 서북방언에서처럼 《ㄷ, ㅌ》는 물론 《ㅎ》에서도 구개음화가 일어나지 않는것이 전통적특성이였으나 이제 와서는 거의다 구개음화가 일어나는것으로 나타났다.

례: 형님-성님 형제-성데

협동-섭중 형편-생편

힘-심

휴가-후가 철쭉꽃-턴디고디

방언들사이에 서로 영향을 주고받는 이러한 경향은 앞으로 더욱 심해질 것이며 그로 하여 방언구역들간의 경계를 확정하기도 어렵게 될것이다.

그러나 위에서 언급한 이 모든것은 어디까지나 자연발생적으로 수집한 자료들을 모아놓은데 대한 분석이다.

우리가 앞으로 전국적판도에서 방언분포의 개략적상황을 공시적으로 파악한다는 목적을 내세우고 동일한 원칙과 방법에 기초하여 방언조사를 진행하려는것만큼 그에 따라 방언조사어휘항목을 잘 선정하여야 할것이다.

이러한 목적의식적인 사업에 맞게 방언조사어휘항목을 선정하는데서 다음의 몇가지 문제가 제기된다.

첫째로 방언조사어휘선정에서 체계성을 보장하는데서 나서는 문제이다.

일부 경우에는 사람들의 일상생활에서 가장 많이 쓰이는 기초어휘에 그에 대응하는 방언어휘가 수집된것이 없는것으로 나타났다.

례컨대 주식물의 기초어휘인 <밥>, <떡>, <죽>가운데서 <떡>은 <시더구> (평북, 자강), <시더기> (강원), <시덕> (함남, 량강)으로 나타났고 <밥>은 <뱌> (함경) 하나로 나타났으며 <죽>은 하나도 올라있지 않다.

이런 경우에 앞으로 방언조사어휘선정에서 체계성을 보장하기 위해 <밥>, <떡>, <죽>을 다 올리겠는가.

우리가 보는데에 의하면 이러한 체계성을 절대화할 필요는 없을것 같다.

<죽>에 대응한 방언어휘가 나타나지 않은것은 그에 대한 방언어휘가 수집되지 않고 빠졌다고보다도 원래 그에 대응한 방언어휘가 없었다고 보아야 할것 같다. 왜냐하면 <죽>은 우리 인민들의 생활과 밀접히 련관된 기초어휘로서 많이 쓰이는것인데 지금껏 그에 대응하는 방언어휘가 하나도 수집되지 않고 있었다는것은 믿기 어려우며 민속학적으로도 그에 대응한 다른 어휘가 고증된것이 아직까지 없는것으로 알고 있다.

아마 발음상변종이 나타난다고 하여도 앞모음화에 의한 <적>과 같은 정도의 차이 밖에 날수 없겠다고 생각된다.

또 사람의 신체구조와 관련된 어휘인 <손>과 <발>만 보더라도 <손>의 경우에는 방언어휘가 수집된것이 하나도 없고 <발>의 경우에는 <발족대기> (량강), <족

대기》(함북)라는 두개 방언어휘만 올라있을뿐이다. 여기서 《족대기》는 한자인 《발족(足)》자와 고유어뒤붙이 《대기》가 덧붙은것으로 추정되므로 사실상 고유어방언어휘라고 보기도 어렵다.

《손》이나 《발》에 그에 대응하는 다른 방언어휘가 있다면 하다못해 다른 어휘와의 합침말관계에서라도 그런 방언 어휘가 나와야겠는데 술한 합침말들가운데서도 새로이 나타나는 방언어휘가 고찰된것이 없다. 있다면 《손가락》의 경우 《소꾸락》, 《송까락》, 《송꾸락》으로 나타나는 경우가 있는데 이것은 합침말을 이루면서 일종의 어음변화가 일어난것이지 《손》이 독자적으로 쓰이거나 합침말의 뒤에 오는 경우(《오른손》, 《왼손》)에는 《소》나 《송》으로 나타나는 일이 없다. 이것은 《소》나 《송》이 《손》에 대응하는 고유한 방언어휘가 아니라는것을 보여준다고 말할수 있다.

결론적으로 말하여 《손》과 《발》에 대응하는 고유한 방언어휘는 없다고 해도 무방할것 같다.

합침말의 경우에도 체계성여부문제에서는 같다고 생각한다.

지금까지 수집한 《손》, 《발》과 관련한 합침말의 경우를 보면 두 경우에 거의 꼭 같이 나타나는것을 볼수 있다.

례: 손툇:손툇, 손툇, 손툇

발툇:발툇

손등:손잔덩

발등:

손목:손목치, 손목찌기

발목:

손바닥:

발바닥:발바당

손가락:손구락, 손꾸락, 쏘꾸락, 소꾸락, 송꾸락, 송꼬락, 송까락

발가락:발고락, 발꼬락, 발꾸락

보다싶이 이 두 경우에는 합침말의 관계에서 나타나는 어음론적 특성이 거의 같다는 것을 알 수 있으며 이런 경우에는 구태여 체계성보장을 위해 두 갈래의 어휘들을 다 올릴 필요는 없으며 발음변종이 많이 나타나는 어느 한 합침말만 선정해도 되지 않겠는가 하는 것이다.

이러한 실례들을 얼마든지 더 찾을 수 있다.

《귀구멍》, 《코구멍》, 《땀구멍》을 보아도 《귀》, 《코》, 《땀》(《땀》에 대한 방언어휘는 나타나지 않았다)이 독자적으로 쓰일 때의 방언어휘에 비해 합침말로 쓰일 때의 방언어휘에서 새롭게 더 나타나는 어음론적 특성이 별로 없다는 것을 알 수 있다.

례: 구멍:구냐, 구녕, 구녁, 구녕, 구농, 구내, 구내:, 구야~, 구양, 구여~, 구영, 궁가이,

궁강, 궁거, 궁그, 궁기, 궁개

귀:귀때기, 구이, 기, 기띠, 귀띠, 귀띠지

귀구멍:귀꾸양

코:코때, 코이

코구멍:코구녕, 코구망, 코궁기, 코꾸내, 코꿍개, 콧구야~, 콧구양, 콧구녕, 콧기멍,

콧꿍기

땀:

땀구멍:땀군녕

이러한 합침말들을 방언조사어휘로 선정할 필요는 그리 제기되지 않으리라고 본다.

그러나 합침말의 경우에도 합쳐지는 말들이 개개로 쓰일 때의 방언어휘보다 새로운 어휘가 나타날 때에는 제외시키지 말아야 한다.

실례로 《기와집》에 대한 방언어휘들 가운데는 《지와집》, 《재:집》 등과 함께 《돌집》(황복)도 있는데 이것은 기와가 돌과 같이 굳은 물건이라는 표식을 계기로 잡아 새롭게 이름지은 것으로서 기존어휘들과 구별되며 따라서 문화어 《기와집》은 방언조사어휘항목에 포함시켜야 할 것이다.

방언조사어휘항목선정에서 제기되는 문제는 다음으로 경향적으로 같은 방언적현상

인 경우에도 방언적현상의 특성에서 중복되는 어휘들을 빠짐없이 다 주겠는가 하는것이다.

실례로 <개>, <개>를 <가이>, <가이>로 발음하는 현상을 들수 있다.

례: 개:가이 평안, 자강, 황해, 강원, 량강
 개:거이 평안, 자강, 황해, 강원, 함남
 새:사이 평안, 황해, 강원
 매:마이 평안, 자강, 황해

보다싶이 이 네개의 어휘들의 방언적특성이 나타나는 지역은 기본상 공통적으로 정해져있다. 이런 경우 어느 지방에서 <개>는 <가이>로 발음하고 <새>는 <사이>가 아니라 <새>로 발음한다고 하여도 <개>를 <가이>로 발음하는이상 그 지역에서는 <개>를 <가이>로 발음하는 방언적특성이 작용한다고 보아야 하지 그렇지 않다고 볼수는 없을것이다. 다만 그러한 특성이 나타나는 어휘수에서의 차이에 의한 경향성으로 지각될뿐인것이다.

그러므로 동일한 방언적현상에 대해 그와 관련되는 똑같은 어음론적조건의 어휘들을 다 줄 필요는 없다고 생각한다.

이렇게 하면 방언조사어휘항목을 펴 줄일수 있을것이며 2,000개미만의 어휘항목을 가지고도 전국적으로 방언현상들의 특성을 반영하는데 지장이 없으리라고 본다.

앞으로 북과 남의 언어학자들은 힘을 합쳐 민족의 언어유산인 방언조사연구를 훌륭히 결속하여 민족어의 통일적발전에 이바지하여야 할 것이다.

방언 조사 항목 선정의 원칙과 방향

이 상 규(경북대학교 국어국문학과 교수)

1. 한민족 문화유산으로서의 방언

1.1. 한민족어는 여러 방언의 총합

지난 세기는 우리 민족에게 있어서는 큰 고난과 시련의 시기였다. 19세기 후반부터 외세 침략 세력들에게 시달리다가 급기야 1910년부터 45년간 일제의 혹독한 침략과 압박을 받았다. 그 기간 동안 일제 침략어를 강요당하는 수난기를 거치면서도 굳건하게 민족의 모국어를 지켜 낸 것은 우리 민족의 강한 불굴의 정신과 자존의 정신이 있었기 때문이다.

개화기 이후부터 우리는 국문과 한문 그리고 국한문 혼용 등의 다양한 문체적 변용을 거쳐 한글의 규범과 틀을 다듬어 왔다. 그러나 그 후 현재까지 약 50여 년간은 분단이라는 상황 속에서도 한민족 언어 분열의 위기를 슬기롭게 잘 대처하고 있다. 그러나 남북간의 언어 규범의 차이로 인한 한민족 언어의 이질화 문제를 이제 우려하지 않을 수 없는 상황이다.

한편 오늘날은 방언의 위기 시대라고 말할 수 있다. 농경 사회를 거쳐 산업 정보화 사회로 변화하면서 통신과 교통수단이 급속도로 발달되었을 뿐만 아니라 인터넷을 통한 정보 매체의 교류가 활발해짐에 따라 한민족 언어의 다양성을 이루고 있는 여러 지역 방언들이 매우 급속도로 소멸되어 가고 있다. 남북 모두 그동안 언어를 통일해야 한다는 언어 정책이 강조되면서 기층 농민들의 생활 감정이 묻어 있을 뿐만 아니라 언

어 발달 과정의 중요한 정보를 갖고 있는 지역 방언이 마치 열등한 것으로 인식되기도 하였다.

방언은 규범적인 언어와 적대적이거나 대립적인 개념이 아니라 특정 지역의 독립된 언어 체계의 단위이다. 지역에 따른 언어 차이를 방언 또는 지역어(지역 방언)라 부르고, 계층에 따른 것을 계층 방언, 성별에 따른 것을 성별 방언, 세대에 따른 것을 세대 방언이라 부르기도 한다. 따라서 방언을 어떤 표준적인 언어 형식에서 이탈되거나 정확하지 않은 언어 변종으로 규정하여 버려야 할 대상으로만 이해해서는 안 된다. 방언을 배제하고 특정한 규범적인 언어만을 중심으로 하는 경우 한민족어의 다양성이 무시되는 동시에 이에 따라 우주관은 점점 좁아지게 될 것이며 지역 문화의 다양성마저도 거부될 것은 분명한 이치이다.

과거의 전통을 구성하는 주요한 요소에 대한 검토 없이는 식민지 시대의 어두운 그림자를 쉽게 벗어나지 못하리라는 예상처럼 규범에 따른 오류를 따지는 규범 중심의 연구 수준을 뛰어넘어 방언을 포함하는 총체적인 한민족어 일반에 대한 성찰과 심층적 연구가 되어야 할 것이다. 언어의 규범을 지나치게 강조함으로써 지역 말씨, 곧 함경도, 평안도, 전라도, 경상도, 제주도 등의 여러 방언이 규범적인 언어와 배타적인 열등한 관계에 놓이게 된 현실의 구조를 바꾸어야 한다. 그러기 위해 한민족의 곁고운 언어 유산인 방언을 수집 보존하는 일을 남북이 함께 해야 한다는 점에 대해 남북학자들이 인식을 같이했다는 사실은 역사적 기록으로 남을 일이 아닐 수 없다.

한민족의 모든 화자들은 적어도 한 가지 이상의 방언을 사용하고 있다. 따라서 한민족의 언어는 한반도와 도서 지역 및 해외 동포들이 사용하는 모든 말과 글을 모두 포괄하는 개념이다. 오늘날 한민족어의 방언은 크게 서북 방언(평안도), 동북 방언(함경도 영흥 이북), 서남 방언(전라도, 추자도), 동남 방언(경상도), 중부 방언과 제주도 방언과 해외 지역 동포들의 방언으로 구성된다. 이들 모든 지역 방언들의 총합을 한민족어라고 규정하여야 한다.

1.2. 방언 수집의 의의

2003년 11월 5일부터 10일까지 북경에서 개최된 “남북 언어 동질성 회복을 위한 제2차 남북 국제 학술회의”에서 남북 언어학자들은 소멸해 가는 언어 유산으로서의 한민족어의 방언을 조사하여 수집하는 일은 매우 시급한 일이며 또 중요한 과제라는 점에 대해 인식을 같이 하고 이를 실질적으로 추진하기 위해 몇 가지 사항을 합의한 바가 있다. 이 합의의 기본 정신은 방언이 점차 소멸해 가는 한민족의 언어 유산이라는 점을 근거로 하여 방언 조사를 추진해야 할 당위성이 무엇인지 다시 한번 강조한 것이다.

우리말이란 한민족 언어를 구성하는 여러 하위 방언을 다 모아 놓은 총합이다. 따라서 한민족어의 특성을 파악하기 위해서는 이를 구성하고 있는 여러 하위 방언의 특성이 규명되어야 할 것이며, 이들 하위 방언들의 공통적인 특성은 곧 한민족 언어의 보편적 특성이라고 할 수 있을 것이다. 그러므로 공시적 측면에서 이러한 하위 방언에 대한 균질적인 연구가 이루어지는 경우에 방언 상호 간의 내적 비교가 가능하게 된다. 방언의 공시적인 연구를 통해 개별 방언의 독자적인 체계를 밝힐 수 있으며, 나아가서 이들 개별 방언의 독자적인 체계를 비교함으로써 공통어인 한민족어의 보편적인 체계를 수립할 수 있다. 또한 이들 자료를 토대로 하여 한민족어의 공통 사전 편찬 등의 기초 사업으로서의 성과도 얻을 수 있으며 나아가서 한민족의 동질성과 정체성을 확인하는 토대가 될 수 있을 것이다. 이번 남북 방언 조사 사업이 남북 한민족이 하나되는 통일의 선도적 기틀을 마련하는 매우 중요한 기반이 되리라 확신한다.

남북 공동 방언 조사 결과는 한민족어의 문화유산으로서의 가치뿐만 아니라 언어의 본질적인 연구를 위해서도 많이 활용될 수 있을 것이다. 함경도, 평안도, 자강도, 강원도, 전라도, 경상도 등 여러 하위 방언에 대한 조사는 한민족 언어에 대한 종합적이고도 체계적인 연구를 위해서 필수적으로 필요한 것이다. 또한 한민족어를 구성하고 있는 여러 방언에 대한 조사와 연구가 국어학 일반과 결코 유리되거나 독립적일 수는 없다.

방언 조사와 연구를 통해 공통어의 통시적 변화 과정에 대한 정밀화를 꾀할 수 있으며 아울러 한민족어의 점진적 변화 과정에 대한 정밀한 설명이 가능하다. 그러므로 방

언 연구는 문헌자료를 통한 통시적 연구의 한계를 보완할 수 있으며, 동시에 개별 방언의 통시적 변화를 이해하는 데 이바지할 수 있다.

5. 남북 공동 방언 조사의 항목 선정 원칙과 방향

5.1. 남북 공동 방언 조사의 1차 합의

2003년 11월 5일부터 10일까지 남북한 연구자들이 함께 모인 남북 언어 동질성 회복을 위한 제2차 남북 국제 학술회의 제1분과 “민족 방언의 공동 연구” 분과에서 남북간의 방언 연구의 현황과 그 전망에 대한 진지한 학술회의를 통해 소멸 위기에 직면해 있는 문화유산으로서 한민족 방언 자료를 남북 공동으로 조사하고 그 결과인 자료를 교환을 할 것을 합의하였다. 남북학자 간에 합의된 중요한 내용은 아래와 같다.

첫째, 소멸해 가는 방언을 한민족 언어의 문화유산으로서 수집 보존해야 한다. 현재 소멸위기에 처해 있는 한민족의 방언은 매우 중요한 민족 문화유산으로서 무형 문화재와 같은 가치를 지닌 것이라는 점에 대해 남북한 학자들이 인식을 같이 하였다. 이러한 공통된 인식을 바탕으로 하여 단계적으로 남북 학자들이 방언 조사를 수행하여 조사된 자료를 서로 교환하여 공유함으로써 한민족어의 동질성을 유지하는 근간으로 활용할 수 있을 뿐만 아니라 한민족의 언어 자산으로서 길이 보존할 수 있다.

둘째, 방언 수집을 위해 2000여 개의 방언 조사 항목을 정하여 2004년부터 남북 공동으로 현지 조사를 하고 또 그 결과를 서로 교환하는 남북 방언 조사 사업을 지속적으로 추진한다.

방언 조사의 항목은 방언을 전통적인 문화유산이라는 인식 아래 소멸 단계에 있는 항목으로 표제어로 정하되 그 분량을 약 2,000항목으로 정한다는 원칙에 합의하였으며 이 조사 항목을 남북에서 서로 검토한 뒤에 최종 확정하기로 합의하였다. 1차로 지난 4월 29일 어휘 항목 일람표를 남측 국립국어연구원에서 제작하여 북측 사회과학원에 전달한 바 있다.

셋째, 방언 조사 항목은 소멸되어 가고 있는 어휘 조사뿐만 아니라 자연 발화 환경을 연출하여 매우 자연스러운 자연 발화도 함께 수집하며 직접 현지 조사 방식으로 조사하며 조사 결과에 대한 전사는 한글 전사를 원칙으로 하기로 합의하였다.

방언 조사를 위해 어휘 조사뿐만 아니라 자연 발화를 조사하기 위한 조사 방법으로 자연 발화 환경을 연출하여 자연 발화를 채록할 수 있는 질문지를 고안하기로 합의하였다.

이상의 합의 내용은 남북 모두 방언을 한민족의 언어를 문화 자산으로 인식하고 이를 보존하여 역사의 기록물로 후대에 물려줄 수 있는 기반이 조성된 것으로 평가할 수 있을 뿐만 아니라 남북간에 언어의 동질성을 유지할 수 있는 기초를 마련하는 매우 획기적인 전기를 마련해 줄 바탕이 될 것으로 확신한다. 따라서 남북 모두 이 사업이 성공적으로 이루어질 수 있도록 노력해야 할 것이다.

5.2. 항목 선정 원칙과 과정

남측에서는 지난 2003년 11월 9일에 남북 언어학들 간에 합의한 남북 공동 방언 조사 사업의 구체적인 추진을 위해 2004년 2월 국립국어연구원에서 남측 방언 조사 추진 위원회를 구성하여 방언 조사 항목 선정 작업과 질문지 제작을 시작하였다. 그 결과 1차로 선정한 방언 조사 항목을 2004년 4월 29일 북측 사회과학원에 전달하였다. 그 이후 항목에 대한 세부적인 검토와 함께 5차례에 걸친 항목 조정 및 항목 분류, 질문지 제작 등의 결과를 토대로 하여 항목의 선정 원칙과 과정에 대해 보고하고자 한다.

5.2.1. 항목 선정 원칙

남북 한민족의 방언 조사를 위해 먼저 남측에서 방언 조사 항목을 선정하기로 하였다. 제3차 남북한 학자들 간에 합의된 남북 공동 방언 조사 원칙에 근거하여 직접 현지 조사를 원칙으로 하며 항목은 1) 자연 발화 조사, 2) 어휘 조사, 3) 음운·문법 조사 등 3으로 구분하고 항목의 수는 약 2,000여 항목의 표제어로 구성한다. 조사 항목은 방언을 한민족 문화유산으로 규정, 사라져 가는 방언을 수집하는 것을 기본으로 하는 동

시에 해당 지역 방언의 특징을 가장 뚜렷이 나타내 줄 수 있는 항목을 우선 선정 기준으로 하였다.

① 직접 조사를 위한 항목 선정

직접 조사나 간접 조사에서 조사 항목이나 질문지를 작성하지 않고 현지 자료를 조사하는 방법도 있을 수 있으나 대체로 방언 조사 목적에 따라 적절한 조사 항목을 미리 선정하고 또 질문지(questionnaire)를 구성해야 한다. 남북 직접 현지 조사를 전제로 한 항목을 선정해야 자료의 균질성을 유지할 수 있다.

② 조사 항목의 수

전국을 대상으로 조사한 외국의 경우 조사 항목 수는 프랑스 언어 지도가 1,920개, 루마니아 언어 지도가 2,100개, 이탈리아-스위스 언어 지도가 2,000개 등으로 이루어져 있다. 오구라 신포이의 조선 지역 방언 조사에서는 조사 일정에 따라 조사 항목을 늘여, 재조사 과정에서 조사가 안 된 지역을 보충하는 방식을 이용했다. 한국정신문화연구원에서 작성한 <한국 방언 질문지(韓國方言調査質問紙)>는 1,782개의 항목으로 구성되었다. 이러한 큰 규모의 질문지는 대상 방언의 전반적 성격을 파악하기 위한 것으로 그 내용도 방언의 어휘 형태·음운·문법 등의 여러 분야를 망라하고 있다.

조사 일정과 소요되는 경비 등을 고려하여 조사 목적에 적합한 숫자의 항목을 마련해야 한다. 제3차 학술회의에서 남북한 학자들 간에 합의된 남북 공동 방언 조사 추진 원칙에 근거하여 자연 발화 조사와 어휘 조사와 음운·문법 항목으로 크게 3구분하여 2,000여 항목의 표제어를 1차로 선정하였다.

③ 조사 항목의 내용

조사 항목 중 가장 많은 분량을 차지하는 것은 어휘 조사 부문이다. 음운이나 문법에 관한 언어 현상은 어휘·형태보다 다양하게 분화되지 않기 때문에 조사 항목이 적을 수밖에 없다.

<영국 언어 지도 질문지>(A Questionnaire for a Linguistic Atlas of English)에서는 1,270개의 조사 항목 가운데 음운이 365개, 통사 41개, 어휘 형태 792개였다. 이처럼 통

상적으로 어휘 형태 항목이 가장 큰 비중을 차지하고 다음으로 음운, 통사 차례로 항목을 배분하는 것이 일반적이다.

조사 항목의 내용은 조사 목적에 적합하도록 사전에 면밀한 계획에 따라 선정해야 한다. 이번 조사는 남북 모두 사라져 가는 방언들을 문화유산이라는 측면에서 이들을 수집할 필요가 있다는 공동의 인식에서 출발한 것이기 때문에 주로 일상의 삶 속에서 사용되는 방언을 우선하였다.

이번 남북 공동 방언 조사에서 자연 발화 조사와 어휘 조사는 전통적인 삶의 방식 속에서 소멸되어 가고 있는 어휘나 자연 발화를 수집하는 데 주력하였으며, 음운·문법 조사는 개별 지역 방언의 독자적인 특성을 밝혀낼 수 있는 항목으로 구성하였다. 특히 음운·문법 조사 항목은 개별 방언의 체계를 비교하여 공통어인 한민족어의 보편적인 특질을 밝혀내는 것을 전제하여 선정하였다.

④ 조사의 난이도

조사 항목의 성격에 따라 조사가 쉬운 것과 어려운 것 간의 차이가 있다. 어휘 항목은 조사가 쉬운 것들이 많으나 음운과 문법에 관한 조사는 간단하지 않다. 예를 들면 동사의 활용형이나 부사어에 관한 조사는 결코 쉽지 않다. 그러나 조사하기 어렵다고 꼭 조사해야 할 항목을 뺄 수는 없는 일이다. 가능한 모든 방법을 고안하여 의도한 방언형을 이끌어 내어야 하는 것이 조사자의 의무이기도 하다.

5.2.2. 남북 방언 조사를 위한 항목 선정 과정

이번 방언 조사를 위한 조사 항목의 선정을 위해 남북에서 간행된 각종 방언 자료집과 사전을 토대로 하여 총체적인 조사 대상 항목의 데이터베이스화를 구축하였다. 이를 토대로 남측의 조사 위원으로부터 언어 발달사와 유관하거나 또는 언어 문화유산으로서의 주요성을 갖는 항목을 여러 차례에 걸쳐 선정하도록 하였다.

① 항목 선정의 준비

지금까지 보고된 남북 측 방언사전의 항목과 각종 사전류 그리고 문화관광부의 21세기 세종계획의 한민족언어정보화의 <남북한 방언 검색 프로그램>에 등재된 항목을

종합한 것을 기준으로 하여 조사 위원별로 항목을 선택한 결과를 계량화하여 항목 데이터베이스에 나타내었다. 각 항목별 여러 가지 정보를 다음과 같은 약어로 나타내어 항목에 대한 정보를 표시하여 국립국어연구원에서 자료를 관리하고 있다.

방언 조사 항목 앞에 위와 같은 정보를 붙여서 9명의 조사 위원(자문 교수 2명, 조사 위원 7명)들로부터 취사선택하는 과정을 거쳐서 어휘, 음운, 문법의 각 항목에 아래와 같이 항목별 선정 기준을 계량화하였다. 이와 함께 조사 항목이 영역별로 중복되는 것을 구별하기 위해 아래와 같은 부호를 부여하였다.

항목	분류	비고
101	101	101
102	102	102
103	103	103
104	104	104
105	105	105
106	106	106
107	107	107
108	108	108
109	109	109
110	110	110
111	111	111
112	112	112
113	113	113
114	114	114
115	115	115
116	116	116
117	117	117
118	118	118
119	119	119
120	120	120
121	121	121
122	122	122
123	123	123
124	124	124
125	125	125
126	126	126
127	127	127
128	128	128
129	129	129
130	130	130
131	131	131
132	132	132
133	133	133
134	134	134
135	135	135
136	136	136
137	137	137
138	138	138
139	139	139
140	140	140
141	141	141
142	142	142
143	143	143
144	144	144
145	145	145
146	146	146
147	147	147
148	148	148
149	149	149
150	150	150
151	151	151
152	152	152
153	153	153
154	154	154
155	155	155
156	156	156
157	157	157
158	158	158
159	159	159
160	160	160
161	161	161
162	162	162
163	163	163
164	164	164
165	165	165
166	166	166
167	167	167
168	168	168
169	169	169
170	170	170
171	171	171
172	172	172
173	173	173
174	174	174
175	175	175
176	176	176
177	177	177
178	178	178
179	179	179
180	180	180
181	181	181
182	182	182
183	183	183
184	184	184
185	185	185
186	186	186
187	187	187
188	188	188
189	189	189
190	190	190
191	191	191
192	192	192
193	193	193
194	194	194
195	195	195
196	196	196
197	197	197
198	198	198
199	199	199
200	200	200

〈그림-1〉 항목별로 여러 가지 정보를 첨부한 항목을 데이터베이스화한 결과

② 조사 항목 분류표

이번 남북한 방언 조사를 위해 예비 작업으로 선정한 2,070항목을 조사 영역에 따른 조사 항목의 분류표는 다음과 같다.

③ 방언 조사 항목의 배열

방언 조사를 위해 예비로 선정한 항목 2,070항목을 가지고 조사 항목의 유기적인 관련성을 고려하여 질문 항목을 적절하게 배열해야만 능률적인 조사를 할 수 있다. 항목의 구성 방법은 일반적으로 의미론의 영역에 속한다. 방언 분화형이 나타나리라고 예상되는 어휘나 음운, 문법 항목의 범주가 결정되면 그것이 동일한 주제에 집중될 수 있도록 의미적으로 유사한 부류끼리 모이도록 해야 한다. 곧 비슷한 항목을 순차적으로 배열함으로써 조사의 흐름을 효과적으로 유지할 수 있다.

유사한 의미 영역의 항목을 같은 자리에 모이게 한다는 것은 기본적인 원칙이다. 그리고 처음에 시작하는 항목들은 가급적 일상적이고 자연스러운 분위기를 만들 수 있는 것이 좋다. 항목 배열의 유기적인 연결 관계의 타당성에뿐만 아니라 질문지의 항목 간의 유기적인 면을 고려하여 질문 내용이 서로 관련성이 있도록 하였다.

영국이나 미국의 방언 조사에서는 농업 기술, 식물군, 동물군, 날씨, 사회 활동, 친족 관계 등과 같이 유사 의미 영역을 집단화하여 질문지를 구성하였다. 한민족어의 경우 지리 언어학적인 목적으로 방언을 조사하는 것이므로 농사일에 대한 조사를 시작으로 하는 것이 바람직하다. 농사에 관한 항목도 농기구 이름, 농작물 이름, 농사일의 이름 등으로 세분될 수 있는데 이들을 각각 하나의 무리로 묶어 조사하는 것이 이상적이다. 이처럼 항목 배열은 의미상 유사한 것끼리 묶여 흐름이 자연스럽게 이루어져야 한다. 예를 들어 농사에 관한 질문을 하다가 갑자기 인체에 관한 질문을 하는 식의 방식은 자연스럽지 못하다.

이러한 기준과 원칙을 고려하여 어휘 항목의 하위 분류를 아래와 같이 11개 영역으로 하위 배열하였다. 그리고 항목의 배열의 순서를 나타내기 위해 하위 항목의 분류 고유 번호를 배당하여 전체 항목의 데이터베이스 둘째 칸에 고유 번호를 나타내었다.

1. 농경
 1. 경작-01
 2. 타작-02
 3. 도정-03

4. 곡물-04

5. 채소-05

2. 음식

1. 주식/부식-06

2. 반찬/별식-07

3. 부엌/그릇-08

3. 가옥

1. 방/가구-09

2. 건물-10

3. 마당-11

4. 마을/가게-12

4. 의복

1. 복식/장식-13

2. 침선/세탁-14

5. 민속

1. 세시풍속-15

2. 농경용품-16

3. 생활용품-17

6. 인체

1. 머리/얼굴-18

2. 상체-19

3. 하체-20

4. 질병/생리-21

7. 육아

1. 발달-22
2. 재롱/놀이-23

8. 인륜

1. 가족/결혼-24
2. 친척/타인-25

9. 동물

1. 고기(민물/바다)-26
2. 곤충/벌레-27
3. 가축-28
4. 짐승/날짐승-29

10. 식물

1. 꽃/풀-30
2. 나무-31
3. 열매/과실-32

11. 자연

1. 산/들-33
2. 강/바다-34
3. 시후-35
4. 날씨/방향-36

위와 같은 분류 방식에 따라 어휘 항목의 배열은 아래와 같은 방식으로 항목을 구성하였다.

④ 조사 항목의 분류와 항목 수

5차례에 걸친 어휘 항목의 분류 및 조정 작업을 거쳐 현재 잠정적으로 확정된 방언

조사 항목의 숫자는 아래와 같다. 2004년 4월 25일 북측에 전달한 1차 예비 선정 한 항목은 어휘 1,433항목, 음운·문법 600항목과 자연 발화 항목(35항목)이었으나 그 이후 항목 재배열과 질문지 구성 과정을 통해 잠정적으로 확정 한 항목의 분류표는 다음과 같다.

어휘 항목 총 1,356항목, 음운(366항목)·문법 항목(326항목) 총 692항목을 포함하여 총 2,048항목이며 자연 발화 항목 136항목을 포함하면 2,184항목으로 구성되어 있다. 주 항목 이외에 보조 항목의 수는 135여 항목이며, 조사자의 재량에 따라 조사하지 않아도 될 <참고> 항목도 다수 포함하였다.

어휘	1,356항목
음운·문법	692항목
음운	366항목
문법	326항목
자연 발화	136항목
총계	2,184항목

<도표-1> 지역어 조사 질문지 항목 총괄표

앞으로 남북 양측에서 항목에 대한 검토 과정을 거치고 또 예비 조사를 거쳐 최종적인 조사 항목을 확정할 수 있을 것이며 2004년부터는 본격적인 조사를 시행할 수 있을 것이다.

언 차이가 예상되기 때문에 ① 붉은색 열매, ② 흰색 열매, ③ 찰수수, ④ 메수수와 같은 항목을 그리고 0413의 '옥수수' 항목도 ① 찰옥수수, ② 메옥수수를 보충 항목으로 처리하였다.

<예> 0411의 '수수'

t 0411 수수

[보충] ① 붉은색 열매

② 흰색 열매

③ 찰수수

④ 메수수

보충 항목 가운데 선택식 질문지 제작을 고려하여 여러 개의 어형을 제시하여 선택하는 방식의 항목도 고려하였다. 곧 1422의 '매듭'의 경우 아래와 같이 보충 항목의 표제어를 "(매듭을) 맺는다/짓는다/엮는다/묶는다" 와 같은 항목도 배열하였다.

1422 매듭

[보충] (매듭을) 맺는다/짓는다/엮는다/묶는다

조사 항목의 방언형이나 용법상의 차이를 확인하기 위하여 참고 질문을 덧붙이기도 했는데, 이는 <참고>로써 표시하였다. <참고>는 조사 항목을 확장하여 더 많은 조사가 가능하도록 배려한 것이다. 조사자가 시간적 여유를 보아 가며 <참고> 조사는 하지 않아도 좋다. 따라서 <참고>는 자료 정리 시에는 제외되지만 보다 정밀한 연구를 위한 자료로 활용할 수 있도록 하였다. 0202 '벚단' 항목에 "<참고> 벚단의 크기에 따른 어형 확인. (1) 한 줌 정도 (2) 타작 단 (3) 한 아름 정도(무논에서)"를 덧붙여 두었다.

또한 조사의 통일을 기하며 또 조사 과정에 특별히 필요한 유의 사항을 강조하고자 할 때에는 <주의>로 표시하였다. 예를 들면 0907 '감추다'의 방언 분화형으로 '숨기다' 형이 공존하는 지역이 있다면 이들이 이형동의어인지 확인하지 않으면 조사가 잘못된

것으로 생각할 수도 있다. 따라서 “<주의> ‘숨기다’와의 의미 차이 확인”과 같은 표시를 하였다.

질문은 때로 그림, 실물의 가리킴 또는 동작 시늉 등의 방법을 활용하였는데 질문상에 <그림>, (지시 내용), (시늉 내용) 등으로 표시하였다. 예를 들면 “0125 ‘발독’<그림 14> (발독을 가리키면서) 여기를 무엇이라고 합니까? (발의 가장자리에 둘러 있는 독으로 흔히 발과 발 사이의 경계가 된다.)”와 같이 표시하였다.

질문에 대해 좀 더 보충적인 정보를 설명할 필요가 있는 경우에는 ()에 질문 내용 풀이를 보충하여 조사에 편리하도록 배려하였다.

어휘 편에서 곡용형 조사와 활용형 조사는 명사형인 경우 기저형의 확인이 필요하거나 용언형의 경우 불규칙적인 활용형을 확인하기 위해 곡용이나 활용의 정보를 첨가하였다.

0107	보습
	-이
0108	벗
	-이
0109	끝다
	-고
	-었다

주 항목에는 일련번호를 배정하였으나 곡용형이나 활용형의 경우나 보충 항목에서는 하위 번호를 부여하지 않고 주 항목에 종속되도록 하였다.

제2편 음운·문법 편은 음운 조사와 문법 조사로 구성되어 있다. 음운 조사는 음운 조사 방법에 대한 간단한 설명과 (1) 음운 체계 조사와 (2) 음운 과정 조사로 되어 있다. 음운 체계 조사는 단모음 체계, 이중 모음 체계, 장모음과 성조 체계에 대한 최소대립쌍을 조사할 수 있는 항목으로 구성되어 있으며 음운 과정에 대한 조사는 곡용과 활용형의 조사 항목으로 구성되어 있다. 문법 조사는 문법 전 영역에 걸친 조사 항목을 문

장 제시형의 질문 방식으로 조사하도록 구성하였다.

제3편 자연 발화 편은 조사 방법과 136항목에 걸친 구술 조사 내용을 담고 있다. 자연 발화 편은 실지 현지 조사에서 활용될 것을 고려하여 앞으로 제1편으로 옮겨 질문지를 구성하는 방안도 검토되어야 할 것이다.

5.3. 항목 선정에서 제기된 문제

남북 공동 방언 조사를 위한 항목 선정 과정에서 양측의 어문 규범 차이에서 기인하는 표제어의 불일치 문제의 해결 방안을 모색해 볼 필요가 있다. 남측의 문화관광부에서 기획한 21세기 세종계획의 일환으로 제작한 <한민족 정서법 변환 프로그램>을 이용하여 몇 가지 제기되는 문제점에 대해 살펴보자.

5.3.1. 남북 항목의 표제어가 없는 경우

남북 측에서 공동으로 조사할 방언 항목을 선정하는데 몇 가지 남은 문제가 있다. 북측의 <조선말대사전>에 표제어가 없는 경우(0)나 남측의 <표준어국어대사전>에 표제어가 없는 경우(11)가 있다. 주로 남북간의 복합어의 인정 범위의 차이에 따라 남측에서 복합어로 인정되는 어휘가 북측에서는 인정되지 않는 경우이다. 아래의 인용 자료는 국립국어연구원에서 제작한 데이터베이스 자료를 활용한 것이다.

① 조선말대사전에 표제어가 없는 경우

묵은세배, 으름, 삼층장, 선달그믐날, 쇠죽바가지, 담뱃대꽃이, 김매다, 쉬슬다, 빨래하다, 구겨지다, 잃어버리다, 길들이다

② 남측 사전인 <표준어국어대사전>에 표제어가 없는 경우

똥&줄, 돼지&머리, 미역&감다, 사발&뚜껑, 장독&뚜껑, 주발&뚜껑, 방귀&뀌다, 송아지&부르는&소리

③ 남북 공통으로 다듬은 말의 차이가 있는 경우

지난날에 또는 낡은 사회에서나 봉건 사회에서 사용되던 어휘를 버리고 새로 다듬어진 어휘들에 대한 항목 표제어를 어떻게 처리해야 할 것인가도 문제가 된다.

< 전날에 >

가게, 눅, 서방님, 인두

< 낚은 사회에서 >

머슴, 무당, 품삯

< 봉건 사회에서 >

사모, 유견

5.3.2. 남북 모두 다듬은 말의 차이로 인한 차이가 있는 경우

언어 다듬기의 결과로 표제어가 남북한 간에 서로 일치하지 않는 경우가 있다. 남북이 표제어는 유사하지만 다듬은 결과 다른 경우, 남측의 한자 말이나 외래어 표제어가 북측의 고유어로 다듬어진 경우, 거의 유사한 말의 경우가 있다. 또 표제어가 방언 차이에 의해 다른 경우도 있다. 그 외에도 맞춤법이나 표기법의 차이에 의해 표제어가 다른 경우도 있다. 곧 음라우트형을 반영한 경우나 사잇소리를 표기법에 반영한 경우와 그렇지 않은 경우의 차이 등의 유형이 있다.

① 남북이 표제어가 유사하지만 다듬은 결과 달라진 경우

고누/꼬니

귀리/귀밀

종달새/종다리

누룽지/가마치

썸바귀/사라구

노른자위/노란자

용마름/룽마름

내년/래년

이자(利子)/리자

② 남측의 한자 말이나 외래어 표제어가 북측의 고유어로 다듬어진 경우

숙부/작은아버지

백부/큰아버지

계모/이붓아버지

망월/보름달

채소/남새

홍수/큰물

홍시/물렁감

희충/거위

③ 거의 유사한 말의 경우

백설기/흰설기

갈고리/갈구리

겨드랑/겨드랑이

곰팡이/곰팡이

등우리/등주리

땅벌/따벌

또아리/따리

맷돌/매돌, 망돌

떡등구미/등구미

엉덩이/엉덩이

요즈음/요즘

가뭇/가물

경기/경풍

이튿날/①초이튿날, ②다음날

허파/폐

기어코/기어이

나흔날/초나흔날

④ 표제어가 방언 차이에 의해 다른 경우

옥수수/장냉이

낙숫물/기스락물

식해/식해

- ⑤ 기타, 움라트우트나 사잇소리 표기를 맞춤법이나 표기법에 반영하는가 등의 요인에 의해 다른 경우

가자미/가재미

우렁이/우렁이

구더기/구태기

윗마을/웃마을

가르마/가리마

다슬기/와라

소방서/소방대

아내/안해

응달/능달

개숫물/개수물

곶간/고간

- ⑥ 남북간의 표제어와 뜻풀이가 뒤바뀐 경우

‘반디불’에 대한 뜻풀이가 남측에서는 ‘개똥벌레’ 또는 ‘반디’로 되어 있으며 북측에서는 ‘반디’에 대한 뜻풀이가 ‘반디벌레’로 되어 있다. 따라서 ‘개똥벌레’를 ‘반디’로 공통 표제어를 조정할 수 있다.

< 공통 표제어로 조정 >

개똥벌레 →반디

방개 →물방개([남]물방개(뜻풀이)=방개, [북]기름도치(뜻풀이)=물방개)

생인손 →생손([남] 생인손(뜻풀이)=생손, [북] 생손얹이(뜻풀이)=생손)

이와 유사한 예로 남측의 ‘부리망’이라는 표제어가 북측에는 ‘꾸레미’이며 그 뜻풀이로 ‘부리망’으로 되어 있어 ‘소입막이’를 ‘부리망’으로 표제어를 조정하였다.

< 항목 조정 >

(소입막이) →부리망

(부태허리) →부티

(나중에) →나중

위에서 제기한 표제어 지정의 문제는 남북 양측의 합의를 도출해야 할 과제이라고 판단된다. 그러나 이 문제는 그렇게 심각한 문제가 되지 않으리라 본다. 현행 규범에 따라 동일한 내용의 표제어를 각각의 규범에 따라 표기하거나 병렬식으로 나타낼 수 있을 것이다. 질문지 구성에서도 질문지 성격을 근본적으로 변형시키지 않는 범위 내에서 조사 대상 지역의 자연스러운 발화 형식으로 바꾸어 물을 수도 있기 때문에 남북 방언 공동 조사를 위해 표제항의 통일 문제는 합의에 의해 해결될 수 있다.

5.4. 어휘사 정보와 그림

남북 공동으로 조사할 지역어 조사 질문지에는 국어 변천을 고려한 어휘사 정보를 약 600여 개 첨부하여 조사자로 하여금 조사 항목의 조사 요점을 미리 파악할 수 있도록 배려하였다.

전 세계에서 방언 조사 질문지에 이러한 역사적 정보를 명시한 사례는 쉽게 찾아볼 수 없다. 이번 조사는 여러 명의 조사자가 각기 다른 시점에서 시차를 둔 조사 방식이 될 것이다. 따라서 조사 항목에 대한 보다 이해가 없으면 조사 자료에 대한 신뢰성이 떨어질 수도 있다.

예를 들면 “2119 ‘부스름’ 항목에서 ‘브스름’이라는 역사적 정보만 있다면 ‘허다(瘡, 弊)’의 방언 분화형인 ‘헌디’형은 마치 의미와 어형이 다른 것으로 인식하여 조사가 잘못 된 것으로 판단할 수 있다. 따라서 2119 ‘부스름’ 항목의 방언 조사 결과가 ‘부스름’ 계열과 ‘허다’ 계열이 있다는 역사적 정보를 조사자가 사전에 미리 예상하고 있는 경우 유리할 것이다. 2121 ‘곰기다’ 항목에서도 만일 ‘힐다’라는 방언형이 나타날 경우 ‘허다(瘡, 弊)’에 대한 역사적 정보를 모르고 있는 경우 조사의 잘못으로 생각하기가 쉽다.

이처럼 조사 항목에 어휘사 정보를 부가함으로써 조사자가 조사 항목에 대한 역사적 정보를 사전에 파악해 둬으로써 보다 효율적인 조사를 할 수 있도록 배려한 것이다.

질문지에서 보다 효율적으로 질문을 하기 위해 많은 항목에서 <그림>을 활용한 질

문지로 제작하였다. 앞으로 질문지에 활용할 그림 자료를 수집 정리가 끝나면 책자 형식과 함께 그림 이미지 파일로 제작하여 예비 조사에 활용할 수 있도록 할 것이다.

6. 마무리

2003년 11월 5일부터 10일, 남북한 연구자들이 소멸 위기에 직면해 있는 한민족 방언 자료를 문화유산으로 인식하고 이를 남북 공동으로 조사하여 그 자료의 교환을 합의한 것은 통일로 향하는 어문 정책의 단초를 연 일이라 평가할 수 있다.

이러한 합의를 바탕으로 한 항목 선정의 원칙과 방향을 방언을 한민족의 문화유산으로 인식하고 소멸해 가는 자료를 수집하는 동시에 한민족어를 구성하는 여러 지역의 방언의 체계적인 이해를 위해 조사 항목을 선정하고 이를 바탕으로 균질적인 방언 조사를 위해 질문지를 구성하였다.

앞으로 조사 항목 및 조사 질문지에 대한 남북 양측의 정밀한 검토와 함께 적절한 조사 지점을 선정하여 예비적인 조사를 거쳐 조사 항목과 질문지를 최종적으로 확정 한 다음 본 조사를 시작해야 할 것이다.

이번에 제안한 <지역어 조사 질문지>는 여러 단계를 거쳐 조사 항목을 조정하고 이를 분류한 결과를 토대로 하여 제작하였다. 시차를 두고 여러 명의 조사자가 조사하는 방식을 취할 남북 방언 조사는 자료의 균질성을 유지하는 것이 가장 중요하다.

동일한 질문지를 기준으로 한 면담은 다른 지역 현지 조사자들이 다양한 환경 아래에서 그리고 일정 기간의 차이를 두고 수행하는 조사를 최대한 동질적인 흐름으로 이끌어 낼 수 있다. 남북 방언 조사를 위해 미리 준비된 질문지를 이용한 격식적 질문법이 매우 유리할 것이다. 따라서 이번에 제작한 질문지를 가지고 각각의 예비적인 조사를 통해 여러 가지 항목과 질문지 상에 나타나는 문제점을 앞으로 보완한 뒤에 완벽한 최종 질문지를 제작해서 본조사에 임해야 할 것이다.

이와 함께 조사 지역 및 조사 제보자 선정 원칙, 한글 전사에 따른 세부적인 방법에 대한 양측의 합의를 앞으로 도출해야 할 것이다. 따라서 향후 일정 조정과 실무적인 일을 효율적으로 추진하기 위해 이번 회의 기간 동안 남북 양측의 실무 추진위를 구성

하여 지속적인 협의를 해 나가는 방안을 모색할 필요가 있다. 보다 효율적인 남북간의 방언 조사를 위해 1) 중장기 조사 계획과 일정(년차 조사 일정과 조사 지점), 2) 조사 항목과 질문지 검토와 최종 확정, 3) 조사 방법(한글 전사 방식, 제보자 선정 기준과 원칙), 3) 자료 처리 방식(녹음 및 자료의 디지털화)와 같은 공동 의제를 실무 추진위에서 긴밀하게 논의할 수 있기를 기대한다.

동북방언의 어음론적특성에 대하여

홍 석 희(사회과학원 언어학연구소 실장 학사)

현대적인 음성녹음설비들과 음성분석설비들이 갖추어진 조건에서 지난 시기에 진행된 방언연구정형을 료해하고 현실적요구에 맞게 새로운 방법론을 세워 방언음성을 조사하고 그의 어음론적특성을 밝히는것은 우리 민족의 언어유산인 방언을 생생한 모습으로 보존하고 리론적으로 연구하여 체계화하는데서 매우 중요한 사업의 하나이다.

해당 방언의 발음상특성은 력사적으로 형성된 발음토대에 기초하고있는만큼 어휘적 측면이나 문법적측면에 비하여 상대적으로 공고하다.

그러므로 해당 방언의 어음론적특성을 구체적으로 밝히는것은 조선어의 력사적발달과정을 해명하고 우리 말의 고유한 특성을 적극 살려나가는데서 중요한 의의를 가진다.

1. 동북방언의 어음론적특성에 대한 간단한 개괄

1.1. 말소리구성

동북방언에는 문화어에 대응하는 모음 <ㅏ>, <ㅓ>, <ㅗ>가 없다.

방언적인 <ㅓ-ㅓ> 대응과 <ㅓ-ㅓ> 대응현상에 의하여 <ㅏ>와 <ㅓ>가 발음되는것이지만 그것은 어디까지나 <ㅓ>, <ㅓ>의 위치상변종으로서 발음된것이지 낱소리로서 존재하는것은 아니다. 동북방언에 낱소리로서 <ㅏ>와 <ㅓ>가 존재하지 않기 때문에 일상 입말에서는 <ㅏ>를 <ㅓ>로 대응시키고 <ㅓ>를 <ㅓ>로 대응

시키거나 <ㅈ+ㅣ>로 대응시키고 있다.

외삼촌-애삼촌 회의-해이

귀중하다-기중하다 위대한-우이대한

동북방언에는 문화어에는 없는 옛날말 <븅, △, 햏, ·>의 역사적 변종이라고 할수 있는 말소리들이 있다.

○ <븅>의 역사적변종으로 볼수 있는 말소리

고우니-고부니 노을-노불

○ <△>의 역사적변종으로 볼수 있는 말소리

가을-가슬, 가실 마을-마슬, 마실

○ <햏>의 역사적변종으로 볼수 있는 말소리

켜다-써다

○ <·>의 역사적변종으로 볼수 있는 말소리

마을-ㅁ실, 말: 가을-ㅁ슬, ㅁ실

1.2. 구개음화현상

동북방언지역은 구개음화가 가장 적극적으로 일어나는 지역이다.

동북방언지역에서 나타나는 구개음화현상으로는 <ㄱ-스>형, <ㅋ-츠>형, <ㅎ-ㅅ>형, <ㄷ-스>형, <ㅌ-츠>형이 있다.

이중에서 <ㄱ-스>형이 가장 활발하게 일어나며 그 다음은 <ㅎ-ㅅ>형이다.

○ <ㄱ-스>형의 구개음화

김치-짐치 기와-지와 길-질

○ <ㅎ-ㅅ>형의 구개음화

향나무-상나무 혀-세 형-성

○ <ㅋ-츠>형의 구개음화

키-치

<ㄷ-스>, <ㅌ-츠>형의 구개음화는 <ㄷ>, <ㅌ>가 받침으로 들어간 소리마디의 뒤에 단어조성적인 뒤붙이 <ㅣ>와 이어질때 일어나는데 이것은 문화어규범으로 인정된것이다.

해돋이-해도지 같이-가치

1.3. 모음대응현상

방언적인 모음대응현상은 우리 나라 전반적지역에서 나타나는 현상이지만 지역에 따라 일정한 차이가 있다.

동북방언지역에서 나타나는 모음대응현상을 보면 <ㅏ-ㅓ>대응, <ㅓ-ㅏ>대응, <ㅗ-ㅜ>대응, <ㅜ-ㅗ>대응, <ㅜ-ㅓ>대응, <ㅓ-ㅜ>대응, <ㅓ-ㅣ>대응이 있다.

이중에서 <ㅏ-ㅓ>대응, <ㅗ-ㅜ>대응, <ㅜ-ㅗ>대응은 동북방언지역을 포함하여 우리 나라 전반적지역에서 매우 적극적으로 나타나는 현상이다. 동북방언지역에서는 <ㅓ-ㅜ>대응과 <ㅓ-ㅣ>대응이 특징적으로 나타난다.

◦ <ㅓ-ㅜ>대응

누구-눠기, 뉘귀 구경-귀경

◦ <ㅓ-ㅣ>대응

두부-두비, 디비 춥다-칩다

1.4. 말소리의 보존과 탈락

동북방언에는 단어의 소리마디사이에서 자음이 약화되거나 없어지는 현상이 있는 반면에 단어의 가운데서 역사적인 음 <ㄱ>을 그대로 보존하여 발음하는 현상이 있다.

시다-시구다

시령-실경

나무-낭구

동북방언에서는 단어의 첫머리와 가운데에서 역사적으로 쇠여오던 <ㄴ>이 반모음 <ㅇ>나 <ㅣ>의 앞에 놓이는 경우 빼버리고 <ㄴ>을 발음하는 현상이 있다.

니탄-이탄

아니다-아이다

가니까-가이까

동북방언에서는 단어의 끝에 오는 <ㄴ>을 빼칠 때 그 앞의 모음을 코소리로 내거나 길게 발음한다.

우리는-우리느: 거기는-거기느:

한편 동북방언에서는 받침으로 들어간 자음 <ㅇ>을 발음하지 않고 빼치면서 그

소리마디를 올림소리로 만든다.

강냉이-강내이 꽃송이-꽃쏜이

1.5. 말소리줄이기

동북방언에서 특징적인것은 단어안의 말소리들을 탈락시켜 줄이거나 짧게 발음하는 현상이 있다.

가지않겠어요-자제이오 옹지않아요-올체이오

1.6. 운률

동북방언의 말은 속도가 빠르고 문장발음속에서 단어의 악센트가 일정하게 작용하며 말소리흐름의 높낮이곡선의 굴곡이 매우 심한것으로 특징지어진다.

2. 심화시켜야 할 문제점과 연구방안

2.1. 문제점

① 동북방언에 존재하는 어음론적인 현상들이 여러 저서들을 통하여 기본적으로 언급되었지만 다른 방언들과의 차이점이 구체적대상들에서 정량적인 값으로 잘 나타나지 않고있다.

실례로 옛날말 《·》의 역사적변종으로서 나타나는 말소리가 동북방언과 동남방언에서 어느정도 차이나는가 하는것을 정량적으로 나타내지 못하고 있다. 또한 동북방언에서 《ㄱ》를 《T+I》로 대응시켜 발음한 결과에 생긴 《우이대한》의 어음구성상 특성이 서북방언에서 《ㄱ》를 《T+I》로 대응시켜 발음한 결과에 생긴 《니파구이》의 어음구성상특성과 구체적으로 얼마나 차이냐며 같으면 어느정도로 같은가에 대한 해석이 부족하다.

② 동북방언에서 어음론적특성의 지역적특성이 구체적으로 나타나지 않고있다.

동북방언지역은 강원도의 북부지방과 룡진지역을 제외한 함경남북도와 량강도를 포함하고 있다.

전지역적인 특성과 함께 부분적지역들의 특성이 나타나야 한다.

실례로 옛날말 ≪·≫의 역사적변종으로 볼수 있는 말소리는 량강도의 삼수, 갑산, 혜산, 보천지방과 함경남도의 정평, 함주지방에서 나타나고 다른 지방에서는 다른 모음으로 대체된것으로 조사자료가 수집되었다. 그러므로 이것을 놓고 동북방언지역에 ≪·≫가 존재한다고 하기는 어렵다.

③ 문화어에서와 같이 동북방언의 어음구성상 특성을 어음론적으로 체계화하지 못하고 있다.

해당 방언에 존재하는 말소리들은 역사적과정을 거쳐 형성된 발음토대에 기초한것으로서 자기의 독특한 소리값을 가지고있다.

방언의 입말은 그 근처에 깔려있는 말소리로서 실현된다. 그러므로 해당 방언에 존재하는 말소리들의 음가를 정확히 규정하여 어음론적으로 체계화하는것은 매우 중요한 일이다.

④ 동북방언의 운률적특성이 매우 소약하게 언급되고있다.

말의 속도와 악센트, 억양은 해당 방언의 어음론적특성을 규정하는데서 매우 중요한 요소이다. 특히 동북방언의 억양은 다른 방언의 억양과 다른 독특한 특성을 가지고 있다. 그러므로 이에 대한 연구가 심화되어야 한다.

2.2. 연구방안

앞에서 제기한 문제점들을 해결하기 위하여서는 다음과 같은 연구사업들이 필요하다고 본다.

① 방언적인 입말들을 구체적으로 나타낼수 있는 정밀전사방법을 연구하여 제정하여야 한다.

문화어로서는 방언음성을 정밀전사하기 어렵다. 왜냐하면 방언은 문화어의 지역적 변종이기때문이다. 그러므로 방언음성을 표기함에 있어서 문화어의 글자외에 영문이

나 기타 기호같은것을 첨부하여 그에 대한 주석을 주고 국내적인 규범으로 제정하는 것이 필요하다.

② 방언의 어음론적특성을 연구함에 있어서 언어지리학적고찰방법을 밀접히 결부시켜야 한다.

방언지역들간의 차이, 한 방언지역안에서의 방언현상의 분화를 구체적으로 표현하기 위해서는 개별적인 어음론적현상들을 가지고 그의 지역분포를 언어지리학적으로 표현해주어야 하며 그렇게 해야 그것을 통해 그 방언지역의 어음론적특성을 더 잘 나타낼수 있다.

③ 방언연구에 실험음성학적인 연구수법들을 적용하여야 한다.

지금까지의 방언연구자료들을 보면 많은 경우 방언소유자의 발성음을 청취하고 연구자의 주관적판단에 의하여 그의 어음론적특성들을 서술하고 있다.

현대적인 음성분석설비들이 갖추어진 조건에서 해당 방언에 존재하는 말소리들의 구체적인 음가를 분석하고 그 방언의 음운구성체계를 과학적으로 고찰하여야 하며 여러가지 어음론적인 현상들을 실험분석하고 정량적인 분석자료에 기초하여 해당 방언의 특성을 론증하고 다른 방언들과 비교고찰하여야 한다.

④ 전국적규모에서 방언의 어음론적특성을 비교고찰할수 있는 자막대기로 될수 있는 방언조사의 표준방안을 작성하여야 한다.

방언에는 지역별로 다종다양한 어음론적인 현상들이 나타나고 있는만큼 여기서 특징적인것들과 보편적인것들을 갈라 이것을 합리적으로 방언조사의 표준방안에 반영하여야 한다.

⑤ 동북방언의 억양을 구체적으로 고찰하기 위한 방법론적문제를 제기하여야 한다.

억양을 연구하기 위하여서는 초보적으로 조선어에 존재하는 여러가지 문장류형을 찾고 그에 기초하여 동북방언에 존재하는 어음론적특성들을 나타내는 말마디들로서 표준문장을 작성하는것이 필요하다.

방언억양은 어디까지나 방언소유자들의 자연발화상태에서 나타나는것이므로 미리 작성된 문장발성음을 끌어내기 위한 언어환경을 마련하는 사업이 중요하다. 즉 그 말을 얻기 위한 연출대본이 있어야 한다.

이상에서 동북방언의 어음론적특성과 그에 대한 연구를 심화시키는데서 나서는 문제들에 대하여 간단히 저의 소견을 말하였다.

우리는 단순히 방언음을 수집하여 역사자료로 보관하자는데만 목적이 있는것이 아니라 그를 통하여 우리 민족어의 단일성, 우리 민족어의 역사발달과정을 론증하고 우리 민족어의 우수성을 더욱 살려나가자는데 목적이 있는것만큼 보다 현실적이고 과학적인 연구조사방안들을 심중히 토의하고 교환하여 방언조사연구사업에 적극 기여하여야 할것이다.

서북방언(평안남도방언)의 지리언어학적특성과 방언조사에서 제기되는 문제에 대하여

백 운 혁(사회과학원 언어학연구소 연구사)

민족어의 순수성과 고유성을 고수하고 북남언어의 통일적발전을 이룩하는데서 기본은 인민들속에 묻혀있는 고유어들을 찾아내어 적극 살려쓰는것이다. 그러자면 고유어들이 많이 묻여있는 방언에 대한 조사를 진행하여야 한다.

오늘 토론회에서는 서북방언에 속하는 평안남도방언의 지리언어학적특성과 평안남도방언조사에서 제기되는 문제에 대하여 고찰하려고 한다.

1. 평안남도방언의 지리언어학적특성

오랜 역사적기간에 걸쳐 평안남도, 평안북도, 자강도가 하나의 행정구역으로 있었는데 서북방언을 형성하는데 커다란 영향을 주었다는것은 다 아는 사실이다. 행정구역이 방언현상의 지역적분포에 주는 영향이 크다고 하여 자연지리적조건이 완전히 무시되는것은 아니다.

평안남도는 청천강과 묘향산줄기를 경계로 하여 평안북도와 자강도, 랑림산줄기, 북대봉산줄기, 아호비령산줄기를 경계로 하여 평안남도와 강원도, 서백년산줄기, 언진산줄기, 대동강을 경계로 하여 황해남북도와 이웃하고 있다. 이러한 자연지리적조건은 평안남도에 고유한 방언현상 그리고 평안남도와 이웃지역과의 호상접촉을 보여주는 방언현상을 낳게 하였다.

그러면 자연지리적조건에 의해 생긴 평안남도방언의 지리언어학적특성은 구체적으

로 무엇인가.

그 특성은 첫째로, 평안남도에만 고유한 방언현상이 나타나고있다는것이다.

평안남도에만 고유한 방언현상은 문법, 어음, 어휘의 측면에서 다 나타나고있다.

이음토 <-니, -니까>에 대응되는 방언토에는 <-니까니, -니까니, -니까니>가 있다. 조사된 자료에 의하면 평안북도에서는 <-니까니, -니까니>가 쓰인것으로 나타나며 평안남도에서는 <-니까니, -니까니>와 함께 <-니까니>도 쓰인것으로 나타나고있다.

문화어 <호미>에 대응되는 서북방언어휘에는 <허무, 허미, 호무, 호메, 호메, 헤미, 회미> 등이 있다. <호메>는 평안남도와 평안북도에서 쓰인것으로 나타나고 있다. <호메>외에 평안남도에서는 <허메>, <호메>가, 평안북도에서는 <호무>가 쓰인것으로 나타나고있다. 문화어어휘 <귀머거리>에 대하여 평안남도에서는 <귀떡데이>라고 하지만 평안북도에서는 <떡초시>라고 하며 문화어어휘 <색동저고리>에 대하여 평안남도에서는 <동달이저구리>라고 하지만 평안북도에서는 <까치저구리>, 자강도에서는 <까치단저고리>라고 하고있다.

즉 방언이음토 <-니까니>, 방언어휘 <허메, 호메, 귀떡데이, 동달이저구리>는 평안남도지역에서만 쓰이고있다.

우의 사실은 평안남도방언에는 서북방언에 고유한 방언현상도 있지만 그 지역에만 고유한 방언현상도 있다는것을 보여주고있다.

그 특성은 둘째로, 린접지역과의 부단한 접촉과정을 보여주는 방언현상이 나타나고 있다는것이다.

자연지리적조건은 사람들이 살며 활동하는데 이러저러한 영향을 준다. 사회가 발전하지 못했던 이전시기에는 자연지리적조건이 사람들의 생활과 활동에 직접적으로 더 큰 영향을 주었다. 바로 이런것으로 하여 오랜 역사적기간 떨어져사는 주민집단들사이에 생활풍습과 언어의 차이도 생기게 되었다. 이러한 지역적주민집단들사이의 관계는 역사의 흐름과 더불어 그의 장애로 되어있는 자연지리적조건들을 극복하는것으로 실현되었으며 이 접촉과 교제는 곧 방언들이 서로 영향을 주고있는 과정으로 되었다.

평안남도방언과 린접지역의 방언과의 부단한 접촉을 보여주는 방언현상은 적지않게 찾아볼수 있으나 여기서는 대표적인 몇개만을 들어보이려고 한다.

문화어알림법토 <<-ㄴ니다>>에 해당하는 방언토에는 <<-ㄴ니, -ㄴ, -슴, -ㄴ무다, -무다>>가 있다. 이미 조사된 자료에 기초하여 방언알림법토의 분포를 보면 <<-ㄴ니, -ㄴ, -슴>>은 역사적으로 오래 고착된 함경도행정경계선을 따라 등어선이 그어지고 있다. 즉 동북방언지역이다. 이와는 달리 <<-ㄴ무다, -무다>>는 오래 고착된 평안도행정경계선을 거의 벗어나고 있다. 즉 등어선은 장자강, 강남산줄기, 승적산줄기, 적유령산줄기, 피난덕산줄기를 따라 서남쪽으로 내려오다가 다시 묘향산줄기를 따라 동북쪽으로 올라뻗은 다음 대동강을 타고 서남쪽으로 내려와서 평안북도의 경계선을 따라 서해기슭에 이르고있다. 구체적으로 지역을 짚는다면 <<-ㄴ무다, -무다>>는 평안북도전역과 평안남도의 개천, 덕천, 녕원군들에 분포되어있다. 이를 통해서 알수 있는것은 <<-ㄴ무다, -무다>>는 평안북도에서만 쓰이였고 평안남도의 일부 지역인 개천, 덕천, 녕원군에 이 방언토가 나타난것은 평안북도와 이웃한 자연지리적조건에 의해 초래된 현상이라는 사실이다.

문화어어휘 <<새우>>에 대하여 평안북도에서는 <<새옹지>>라고 하고 함경도에서는 <<새비>>라고 한다.

조사된 자료에 의하면 <<새옹지>>는 평안남도의 문덕, 안주, 개천군에서도 쓰이고있으며 <<새비>>는 평안남도의 대흥군에서도 쓰이고있다.

평안남도에 고유한 방언이 아님에도 불구하고 평안남도의 방언에 <<새옹지>>, <<새비>>가 나타난것은 그와 이웃하고있는 평안북도와 함경도방언의 영향을 받았기때문이다.

평안남도방언과 이웃지역과의 호상접촉을 보여주는 방언현상이 제일 강하게 나타나는 지역은 평안남도의 남부이다.

평안남도의 남부지역은 황해남북도와 린접해있다.

문화어어휘 <<팽이>>를 황해남도 황주, 송림시, 평안남도 룡강, 온천, 순천, 녕원, 맹산, 안주, 문덕, 강동, 개천군, 평안북도의 녕원, 구장, 향산, 박천, 구성, 동립, 정주군, 자강도의 희천, 동신군에서는 <<세리>>라고 하고 황해남북도와 평안남도의 룡강, 남포, 온천군에서는 <<페이>>라고 하고있다. 문화어어휘 <<턱>>을 평안남도의 남포, 룡강, 온천군, 황해남북도에서는 <<턱수가리>>라고 하고있다.

방언어휘 《세리》는 평안도에 고유한 방언어휘이고 《페이》나 《턱수가리》는 황해도에 고유한 방언어휘이다. 이것이 린접지역에 영향을 주어 《세리》는 황주, 송림 시에서도 쓰이게 되었고 《페이》나 《턱수가리》는 평안남도의 남포, 룡강, 온천군에서도 쓰이게 되었다. 황해남북도의 북부지역과 평안남도의 남부지역과의 방언접촉이 얼마나 활발했는가 하는것은 황해도방언의 특징적인 말소리현상이 평안남도의 방언들에 나타나고 평안도방언의 특징적인 말소리현상이 황해남북도의 방언들에 나타나고있다는 사실을 통해서도 알수 있다.

황해도방언의 특징적인 말소리현상의 하나는 홀모음 <一>가 단어의 첫머리에 올 때 일반적으로 <ㄱ>로 발음되는것이다. 실례로 <글→걸, 그림→거림, 들국화→덜국화>를 들수 있다.

문화어어휘 <그네>를 황해도에서는 <건네>라고 하고있다. 평안남도 남부지역에서는 <그네>를 <거니, 건네>라고 하고있다. 평안도 방언의 특징적인 말소리현상의 하나는 단어의 첫머리에 오는 <ㅇ>가 <ㅏ, ㅑ, ㅓ, ㅕ, ㅗ, ㅛ, ㅜ, ㅠ, ㅣ, ㅟ>앞에서 <ㄴ>로 발음되는것이다. 실례로 <옛날>을 <넛날>, <여름>을 <녀름>으로 발음하는 현상을 들수 있다. 문화어어휘 <이, 이발>을 평안도방언에서는 <니>라고 하고 황해북도 황주, 송림, 서흥군, 황해남도 재령군에서도 <니>라고 하고있으며 문화어어휘 <잉어>를 평안도방언에서는 <니에, 닝어>, 황해북도의 황주, 송림, 서흥군에서는 <닝아>라고 하고있다.

이상의 사실은 평안남도방언은 그것이 처한 자연지리적조건에 의하여 린접지역의 방언들과의 부단한 접촉을 가져왔으며 결국 황해남북도의 북부지역을 서북방언구획에 들어가게 하였다는것을 보여주고있다.

2. 방언조사에서 나서는 문제

방언조사를 진행하는데서는 많은 실무적인 문제가 제기된다.

지역선정, 조사방법, 조사대상자선택 등을 들수 있다. 그러나 이에 앞서 해야 할 일이 있다. 그것은 한마디로 조사계획이다.

앞에서 본 평안남도방언의 지리언어학적특성에 기초하여 평안남도방언에 대한 조사를 실리가 나게 과학적으로 진행하기 위해서는 조사계획을 올바르게 세워야 한다.

계획은 방언조사를 단번에 다 할수 없는것만큼 단계별로 세워야 한다.

단계별 계획을 올바르게 세우는 문제는 방언조사지역을 정확히 선정하고 사전준비를 잘 하도록 하는데서 기초적인 문제로 된다. 그것은 계획을 어떻게 세우는가에 따라서 방언조사지역과 사전준비가 달라지기때문이다. 만약 이번 조사를 방언들사이의 호상접촉을 보여주는 방언현상을 위주로 진행하겠다고 계획을 세웠다면 지역도 그에 따라 다른 도와 이웃하고있는 군들로 선정해야 할것이며 사전준비도 이 군들에 대한 방언지식, 역사, 지리, 풍습에 대한 자료습득위주로 해야 할것이다.

평안남도방언조사는 3단계로 갈라 진행하는것이 제일 합리적이라고 본다.

첫번째단계; 서북방언의 특성과 평안남도방언의 고유한 방언현상을 조사하는 단계.

둘째단계; 평안남도방언과 그를 이웃하고있는 린접방언들과의 방언접촉정형을 조사하는 단계.

셋째단계; 그에 기초하여 앞의 조사에서 미흡하다거나 새로운 방언현상이 있음직한 지역을 조사하는 단계.

오늘 토론회에서는 첫번째단계인 서북방언의 특성과 평안남도 방언의 고유한 방언현상을 조사하는 단계에서 제기되는 문제에서 지역선택과 사전준비에 대해서만 이야기하려고 한다.

우선 지역선정에 대하여 이야기하겠다.

지역선정을 어떻게 하는가 하는것은 소기의 목적을 달성하는가 못하는가를 좌우하는 기본열쇠이다. 평안남도의 자연지리적특성을 보면 산을 낀 지역이 있는가 하면 바다를 낀 지역, 지리적장벽이 없는것으로 하여 사회력사적으로 교제가 활발하게 진행된 지역도 있다.

평안남도방언의 조사지역선정은 이러한 자연지리적조건, 사회력사적조건을 고려하여 선정해야 한다.

방언조사지역은 여러가지 측면에서 선정할수가 있다.

첫째로 오래동안 고착되어 전해져오는 군을 조사지점으로 선정할수 있다.

오래동안 고착되어 전해져오는 군을 조사지점으로 선정해야 하는것은 우선 일반적으로 역사를 거슬러 올라갈수록 군은 다른 군에 속하여있는 사람들과의 교제를 약화시켜 언어적차이를 증대시키며 다음으로 역사적으로 내려오는 군이 그 당시의 정치, 경제, 문화, 자연지리적조건을 고려하여 설정된것과 관련하여 군안에서는 방언현상이 비교적 유지되고 보존될 가능성이 크기때문이다. 이에 부합되는 군들로서는 리조시기부터 오늘까지 변화없이 내려온 중화, 상원, 대동, 룡강, 평원, 성천, 순천, 안주, 속천, 개천, 덕천, 녕원, 은산, 양덕, 맹산, 증산 등을 들수 있다.

둘째로 다른 도의 군과 경계선을 이웃하지 않는 군들을 조사지점으로 선정할수 있다.

왜냐하면 다른 도의 군과 경계선을 이웃하고있는 군은 방언들의 호상접촉이 이루어 질수 있는 가능성을 주며 따라서 이런 지역은 그 지역에 고유한 방언현상을 수집하기 힘들게 할수 있기때문이다. 이에 부합되는 군들로서는 순천, 속천, 북창군 등을 들수 있다.

셋째로 자연지리적조건이 다른 지역집단주민들사이의 련계에 최대한 불리한 군을 조사지점으로 선정할수 있다.

자연지리적조건이 다른 지역주민집단들사이의 련계에 최대한 불리한 지역을 조사지점으로 선정해야 하는것은 그런 지역일수록 다른 주민집단들과의 교제와 접촉이 적고 따라서 다른 방언들의 영향을 덜 받을수 있기때문이다. 이에 부합되는 군들로서는 북창, 맹산, 양덕, 덕천, 녕원군 등을 들수 있다.

우에서 열거한 군들중에서 그중 몇개의 군만을 선택하여 방언조사를 진행하겠는가 아니면 우의 군전체를 다 조사지점으로 선정하여 방언조사를 진행하겠는가 하는것은 앞으로 더 논의해보아야 하는것만큼 더 구체화하지는 않으련다.

다음으로 사전준비에 대하여 이야기하겠다.

사전준비문제는 방언조사방법과 직접적으로 관련되어있다. 정확히 말하여 방언조사를 직접조사방법으로 진행하는것을 전제로 하는 경우 매우 중요한 문제로 제기된다. 따라서 사전준비를 잘 하는것은 방언조사의 정확성과 조사된 자료의 믿음성을 담보하는 기본조건으로 된다.

방언조사를 위한 사전준비는 크게 두가지 측면에서 해야 한다. 하나는 서북방언의 특성과 평안남도지역에서만 고유한 방언현상에 대한 사전료해를 충분히 하는것이다. 이에 대한 사전료해가 없이 방언조사에 착수한다면 방언조사과정에 나타날 수 있는 이상하거나 특이한 문법, 어음, 어휘적현상에 대하여 스쳐보낼수가 있다. 레하면 서북방언의 소유자가 서북방언의 특성에 저촉되는 방언으로 이야기한 대목에 대하여 방언조사자가 그것이 정확한가, 그런 현상이 왜 나타났는가를 즉시 포착하고 미타한 점에 대해서는 재조사해야겠으나 서북방언에 대하여 잘 모르는것으로 하여 스쳐보낼수 있다. 이것은 서북방언의 특성을 재확인하는데 지장을 준다. 따라서 방언조사의 정확성과 믿음성을 보장하기 위해서는 반드시 서북방언의 특성과 평안남도에만 고유한 방언현상에 대한 사전료해를 충분히 해야한다.

다른 하나는 해당 지역의 자연지리적대상과 생활세태, 이 지방의 특징적인 현상이나 지명변천 등을 사전에 아는것이다. 해당 지역에 대한 예비지식이 없이 방언조사에 착수한다면 소기의 성과를 거둘수 없을뿐아니라 경우에 따라서는 조사자체가 전혀 불가능한것으로 될수 있다. 방언소유자가 담화과정에 조금이라도 부자연스럽거나 어색한 느낌을 가지지 않도록 하자면 이 지방에 대한 예비지식을 착실히 품을 들여 갖추어야 한다.

북과 남, 해외의 언어학자들은 앞으로도 6.15북남공동선언의 기치를 계속 높이 추켜 들고 있는 힘과 지식을 합쳐 방언조사와 연구를 더욱 심화시켜 조선민족의 귀중한 언어유산을 보존하고 민족어의 단일성을 론증하며 민족어를 통일적으로 발전시키는데 적극 이바지하여야 할것이다.

민족어 방언 검색 시스템 개발

홍 윤 표(연세대학교 국어국문학과 교수)

1. 머리말

이 글은 2000년부터 2003년까지 4년간에 걸쳐 21세기 세종계획의 일환으로 이루어진 ‘민족어 방언 검색 시스템의 개발’(원이름은 ‘한국 방언 검색 시스템 개발’)에 대하여 그 목적 및 개발 방법과 과정 그리고 그 과정에서 발생하였던 문제점들을 제시하여서, 앞으로 남과 북의 방언 자료를 조사·수집·정리하여 이를 검색할 수 있는 프로그램을 개발할 때에 도움을 주고자 하여 쓴 것이다.

2. 목적 및 필요성

2.1. 목적

민족어 방언 검색 시스템 개발은 남과 북의 방언 및 세계 전 지역에서 사용되고 있는 우리말 방언을 컴퓨터로 검색할 수 있도록 하여, 21세기 지식 기반 사회에서 민족의 의사소통을 신속·정확하게 하며 또한 통일 이후에도 우리 민족 구성원들 간의 의사소통을 원만하게 함으로써 민족의 정보 공유를 유도하여 민족 문화 창달에 기여함을 그 목적으로 하여 계획되고 실현된 것이다.

우리말이 어느 지역이나 계층에서 독자적인 변화를 거쳐 사람들이 인식할 수 있는 정도의 상이한 체계를 이룰 때에 각각의 언어 체계를 우리말 방언이라고 한다. 우리말

은 전국적으로 완전히 동질적인 언어로 인식되지만, 실제로 있어서는 지역에 따라 다른 말로 이루어져 있는 것이다.

우리는 각 지역별로 다른 방언을 사용하고 있어서 남은 일반적으로 제주도 방언, 경상도 방언, 전라도 방언, 충청도 방언, 경기도 방언, 강원도 방언 등으로 통칭되어 오고 있다. 그리고 북은 함경도 방언, 평안도 방언, 황해도 방언 등으로 구분되어 오기도 하였다.

이러한 방언들의 차이로 인하여 발생하는 일이 지역적인 차이에 의한 것이든, 의사소통의 전달 방법이나 어휘의 차이에 의한 것이든 간에, 의사소통의 장애를 가져오며, 의사소통의 장애는 우리나라 문화 발전에 장애를 가져오게 된다. 왜냐하면 의사소통의 원활성은 곧 정보 전달의 정확성과 신속성을 의미하기 때문이다. 이러한 정보 전달의 원활성은 곧 문화 전달의 정확성과 신속성을 가져오며, 그것은 곧 문화의 발달로 이어진다.

그러므로 방언은 그 자체로 보존할 가치가 충분하면서도, 동시에 각 방언 사용자들의 원활한 의사소통을 위하여서는 각 방언들 간의 차이와 대조 관계를 밝혀 두어야 할 필요가 있는 것이다.

우리말의 방언은 남과 북의 주민은 물론, 중국과 러시아와 일본 등에 살고 있는 해외 동포의 방언까지 합하면 아주 다양하게 발달되어 있다. 한민족 언어 공동체를 이루기 위해서는 이러한 다양한 방언을 이해할 수 있도록 하지 않으면 안 된다. 이러한 도구 중 가장 효율적인 것이 각 방언형들을 컴퓨터로 검색할 수 있는 프로그램을 개발하는 일이다.

그러나 이러한 프로그램을 개발하는 일은 그리 수월한 일이 아니다. 그래서 이 일은 다음과 같이 단계적으로 시행할 수밖에 없었다.

- 1) 2000년 : 민족어 방언 검색 시스템 개발을 위한 기초 조사
- 2) 2001년 : 남의 방언 검색 시스템 개발
- 3) 2002년 : 북의 방언 검색 시스템 개발
- 4) 2003년 : 중국 및 기타 지역 방언 검색 시스템 개발

그리하여 2003년도에는 민족어의 모든 방언형을 검색할 수 있도록 종합화하였다.

2.2. 필요성

우리말은 지역마다 지역어 및 방언이 존재한다. 그리하여 각 지역 언어의 쓰임에 차이가 있다. 우리 민족이 전 세계에 지역적으로 넓게 분포하여 생활하면서 각각 독립된 언어생활을 영위한 결과이다. 현재 우리 민족은 크게는 주로 남과 북에 분포하여 생활하고 있다. 그리고 역사적인 이유로 말미암아 중국 동포들은 주로 중국의 동북 삼성(길림성, 요녕성, 흑룡강성)에 퍼져 약 200만 명이 그들만의 독특한 방언을 구사하며 살고 있다. 뿐만 아니라 우리 민족은 미국, 일본, 유럽 및 심지어는 아프리카에까지 퍼져서 살고 있다. 우리 민족의 전 세계 분포 현황을 보이면 다음과 같다.⁶⁾

지 역	사용자 수
남	45,985,289명
북	(1999년) 22,080,000명
중국	2,144,789명
중앙아시아	538,689명
일본	638,546명
북미	2,327,619명
유럽	113,442명
중남미	105,643명
아프리카	5,095명
중동	6,559명
아시아	196,401명
합 계	약 74,142,072명

결국 약 7천4백14만 2천 명이 우리말을 사용하고 있는 셈이어서, 전 세계에서 한 언어를 사용하는 사용자 수로 본다면 13위에 드는 것이다.

이들 우리 동포들은 그 역사성으로 인하여 현재의 우리말 사용 실태나 현상이 다르다. 미국에는 약 200만 명이 살고 있으나 일본 동포들과 마찬가지로 이주의 역사가 짧고 우리 민족끼리의 공동생활이 이루어지지 않아서 우리 언어를 순수히 지켜 가는 경

6) 북을 제외한 모든 통계 자료는 해외 동포 재단에서 2003년을 기준으로 조사한 것이다.

우는 드문 형편이다. 따라서 주로 우리말을 말하며 생활하는 사람들은 주로 남과 북의 주민, 그리고 중국 및 러시아의 동포들이라고 할 수 있다.

이들 민족을 하나의 한민족 공동체로 묶으려고 할 때에 가장 심각한 문제로 등장하는 것은 바로 언어이다. 왜냐하면 언어로 인한 의사소통의 장애가 일어나기 때문이다.

남의 방언은 그간의 지속적인 연구 결과와 대중 매체를 통한 부단한 소개로 방언 차이를 극복할 수 있는 길이 어느 정도 열리게 되었다고 하지만, 그것은 표면적인 현상일 뿐이다. 실제로 각 지역 출신자들이 다른 지역 출신자들과 대화를 나눌 때에, 북의 주민들에게는 아직도 생소한 것이며, 북의 방언은 아직까지 면밀한 조사·연구 여부의 정보를 알 수가 없고, 또 널리 소개된 적도 거의 없어서 남의 주민들에게 익숙하지 않다. 그리하여 남북 통일이 이루어졌을 때에 이 남북의 방언 차이는 남북 주민들 간의 의사소통에 커다란 장애가 될 수 있다.

남과 북의 언어 차이는 한 언어의 방언적 분화에 불과하다. 즉 남의 언어와 북의 언어는 서로 다른 언어가 아니라 우리말 안에서 지리적, 사회적 요인에 의하여 분화된 언어인 것이다. 따라서 남의 언어와 북의 언어는 이질성보다는 동질성이 더 강한 것이다. 그럼에도 불구하고 남북이 분단된 반세기 이후에 접하는 남북의 언어들이 서로 이질적인 것처럼 인식되는 것은, 남과 북의 우리말이 두 지역 간의 왕래가 불가능한 상태가 됨으로써, 지역 방언의 차이가 급속하게 이루어지게 되었고, 또한 남북간의 사회구조의 차이로 인하여 사회 방언의 차이가 극대화되었다고 인식되기 때문이다.

언어상의 차이로 인한 의사소통의 장애를 극소화하여 문화 발달의 저해를 최소화하는 것이 통일에 대비하는 가장 중요한 요체이다. 왜냐하면 언어나 방언 차이는 동시적 협력과 계기적 협력을 방해할 수 있기 때문이다. 그래서 현재까지 남과 북의 언어 차이나 언어생활의 차이에 대하여 많은 연구가 이루어져 왔다. 그러나 그 차이에 대한 연구는 주로 남의 표준어와 북의 문화어와의 차이를 통한 연구가 대부분이었다. 실제 남과 북의 주민들이 사용하는 실제적인 언어는 각 지역의 지역어나 방언이다. 따라서 진정으로 남북의 언어 차이로 인한 의사소통의 장애를 극복하기 위해서는 각 방언에 대한 연구와 조사가 가장 우선적으로 검토되어야 할 것이다.

그래서 지금까지 부분적으로 각 도의 방언이 사전으로 만들어지거나 자료집으로 편

찬되었다. 그러나 이들 자료를 하나로 통합된 방언 자료집이나 이들을 검색할 수 있는 시스템이 갖추어진 적이 없다.

이러한 의사소통의 원활성과 각 지역에서 특수하게 사용되는 방언형들로 인한 의사소통의 장애를 방지하기 위하여 각 방언형들을 검색하여 표준어와 대비시키는 작업이 필요하다.

우리는 통일이 된 후에 각 지역인들의 원만한 의사소통이 이루어지게 하기 위해서 남북 및 해외 동포들의 방언에 깊은 관심을 쏟아야 할 것이다. 왜냐하면 방언 차이로 인한 남북 및 해외 동포들의 이질화를 최소화하여야 하기 때문이다. 한민족 공동체의 동질화를 위해서라도 방언 연구를 통하여 서로 이해의 폭을 넓혀 가야 할 것이다.

뿐만 아니라 이 방언 검색 시스템의 개발은 부수적으로 학문 발전에 크게 기여할 것으로 기대된다. 왜냐하면 이 방언 검색 시스템 개발을 위하여 기존의 각 방언 자료집을 종합하게 될 것이며, 이를 토대로 광범한 방언사전 편찬이 이루어질 것이고, 또한 국어사전 편찬 및 보완을 위한 방언 자료를 확보하게 될 것이기 때문이다. 뿐만 아니라 국어의 각 어휘 사전 편찬을 위한 자료를 확보하는 학문적 성과를 이룰 수 있을 것이다.

3. 연구 동향

지금까지 많은 국어 정보화 작업의 결과물이 등장하고 있어서, 우리말을 배우는 프로그램은 비롯하여 국어 교육에 필요한 프로그램들이 마련되어 있지만, 민족어 방언을 검색하는 프로그램은 한번도 나온 적이 없다. 따라서 이 분야에 대한 연구는 이제까지 한번도 없었다. 그도 그럴 것이 이러한 방언 검색 시스템을 만들기 위해서는 우리나라의 방언에 대한 전반적인 조사나 연구가 수반되지 않으면 안 되는데, 실제로 그러한 노력이 전혀 없었기 때문이다.

일제 시대부터 지금까지 우리나라에서는 방언에 대한 깊은 관심을 가지고 많은 학자들이 방언 조사를 통하여 자료를 구축해 놓았고, 이것들이 종합화되어 각 지역의 방언사전이 출현하고 있으며, 요즘에는 남에서는 특히 지방 자치제가 실시되면서, 각 지역에서 그 지방에서 사용하는 언어 즉 방언에 대한 애착을 가지고 자료를 수집·정리

하여 오고 있다. 그래서 방언에 대한 조사와 연구는 그 방법이나 자료상에 문제가 있다고 하더라도 어느 정도는 성과를 거두어 왔다고 할 수 있다. 그러나 남의 방언 조사나 연구는 학문적으로나 양적으로나 질적으로 만족할 만한 수준에 있다고 할 수 없다. 이에 비해서 북의 방언이나 중국의 연변 지역의 방언에 대한 조사와 연구에 대한 정보는 남에는 그리 널리 알려져 있지 않다. 남에서 조사된 자료에 비하면 매우 소략한 편이라고 할 수 있다.

따라서 이 프로그램을 개발하는 과정에서 제일 먼저 이루어져야 할 일은 자료 조사라고 할 수 있다. 자료 조사는 직접 방언을 조사·수집하는 방법과 이미 조사해 놓은 방언 자료를 수집 정리하는 두 가지 방법이 있는데, 전자는 오랜 시간과 노력과 비용이 들어서 실현 가능성이 희박하여 결국 후자의 방법을 선택하게 되었다.

4. 진행 과정

이 컴퓨터 프로그램이 검색 프로그램으로서 기능을 다하기 위하여 반드시 거쳐야 할 과정은 다음과 같다.

- | | |
|-------------------|------------------|
| ① 방언 자료의 조사 | ② 방언 자료집의 수집 |
| ③ 각 방언 자료집의 구조 통일 | ④ 방언 자료의 선별 |
| ⑤ 방언 자료 정리 기준 마련 | ⑥ 방언 자료 입력 구조 별정 |
| ⑦ 프로그램 개발 | |

이 과정을 좀 더 자세히 설명하면 다음과 같다.

4.1. 방언 자료의 조사

남, 북, 기타 중국 및 세계의 우리말 방언 자료를 수집하였다. 그리하여 지금까지 간행된 방언 자료집 및 사전을 조사하고 그 신뢰성을 검토하여 방언 자료집을 선택하였다. 그리고 다음과 같은 방향에서 기초 조사를 착수하였다.

- ① 국어사전(국립국어연구원의 『표준국어대사전』, 한글 학회의 『우리말 큰사전』, 금성사의 『국어대사전』)에 등재되어 있는 방언형을 모두 추출하였다.

- ② 각 자료집에 등재되어 있는 방언형을 모두 추출하였다.
- ③ 각 지역에서 일반적으로 많이 쓰이는 방언형을 추출하였다.
- ④ 의미가 다른 방언형은 그 의미도 고려하여 추출하였다.
- ⑤ 남, 북, 중국 연변의 방언을 최대한 고려하였다.

이들 자료를 프로그램에서 사용할 수 있도록 입력하는 기본 입력 구조를 만들어야 하는데, 그것은 프로그램에서 검색할 수 있는 사항과 연관된다. 그 검색 사항을 제시하면 다음과 같다.

- ① 표준어를 검색하면 각 지역의 방언형이 제시되어야 한다.
- ② 방언형을 검색하면 표준어를 제시하여 주고, 더 원하면 각 지역의 방언형을 제시하여 주어야 한다.
- ③ 의미가 다른 방언형은 그 의미까지도 제시하여 주어야 한다.
- ④ 이들의 방언 분포를 프로그램으로 처리하여 방언 지도를 보여 줄 수 있도록 하여야 한다.

4.2. 방언 자료집의 수집

지금까지 조사 연구된 방언 자료집(남, 북, 중국 등)의 목록을 작성하였다. 그리고 이 중에서 전자 파일로 되어 있는 방언 자료집은 필자나 입력자의 허락을 받아 전자 파일로 수집하였는데 그것은 앞에 *표를 한 것이다.

- * 김병제(1980), 방언사전, 과학,백과사전출판사.
- * 김영배(1997), 평안 방언 연구(자료 편), 태학사.
- * 김영태(1975), 경상남도 방언 연구(1), 진명문화사.
- 金履挾(1981), 평북 방언사전, 한국정신문화연구원.
- * 김주석최명옥(2001), 경주 속담말 사전, 한국문화사.
- * 김태균(1986), 함북 방언사전, 경기대학교 출판부.
- 김형규(1974), 한국 방언 연구, 서울대 출판부.
- * 리운규심희섭안운(1992), 조선어 방언사전, 연변인민출판사.
- 석주명(1947), 제주도 방언집, 서울신문사.

- * 宣德五趙習金淳培(1990), 朝鮮語 方言 調査 報告, 연변인민출판사.
- 小倉進平(1927), 咸鏡北道方言の 研究, 朝鮮語 16.
- 小倉進平(1944), 朝鮮語道方言の 研究, 東京, 岩波書店.
- 순화조선어연구부(1935), 方言集, 순화조선어연구부.
- * 이기갑 외(1997), 전남 방언사전, 전라남도.
- 이돈주(1978), 전남 방언, 형설출판사.
- * 이상규(2000), 경북 방언사전, 태학사.
- 최학근(1962), 전라남도 방언 연구, 한국 연구원.
- 최학근(1978), 한국 방언사전, 현문사.
- * 한국정신문화연구원, 한국 방언 자료집
- * 한국정신문화연구원(1987), 한국 방언 자료집 3(충북 편)
- * 한국정신문화연구원(1987), 한국 방언 자료집 5(전북 편)
- * 한국정신문화연구원(1989), 한국 방언 자료집 7(경북 편)
- * 한국정신문화연구원(1990), 한국 방언 자료집 2(강원도 편)
- * 한국정신문화연구원(1990), 한국 방언 자료집 4(충남 편)
- * 한국정신문화연구원(1991), 한국 방언 자료집 6(전남 편)
- * 한국정신문화연구원(1993), 한국 방언 자료집 8(경남 편)
- * 한국정신문화연구원(1995), 한국 방언 자료집 1(경기도 편)
- * 한국정신문화연구원(1995), 한국 방언 자료집 9(제주 편)
- * 현평효 외(1995), 제주어 사전, 제주도.
- * 서울대학교(1997), 한국 방언사전)

7) 이 자료집은 국내외에서 간행된 방언사전과 방언 자료집에 수록된 방언형들을 표제어별, 지역별로 분류하여 만든 것이다. 이 자료집을 만들기 위해 이용된 방언사전류의 목록은 다음과 같다.

김병제(1980) 방언사전, 평양:과학, 백과사전출판사.

小倉進平(1944) 朝鮮語方言の研究 下, 東京:岩波書店.

韓國精神文化研究院(87 - 96) 韓國方言資料集 I - IX, 韓國精神文化研究院.

리운규·심희섭·안운(1992) 조선어방언사전, 연변:연변 인민출판사.

崔鶴根(1978) 韓國方言辭典, 玄文社.

河野六郎(1945) 朝鮮方言學試攷 -「缺」語攷.

방언에 관한 수많은 논문이 있으나 그 논문에서 제시한 단편적인 자료들은 조사하지 않았다. 다만 대표적인 저서를 선정하여 자료를 조사하였다. 아울러 북이나 중국, 카자흐스탄 등의 방언 자료들을 수집하기 위하여 저서나 연구 논문을 조사하였다.⁸⁾

King, J. R. P. & 연제훈(1992), 중앙아시아 한인들의 언어: 고려말, 한글, 217.

곽충구(1995), 강원도 북부 지역의 언어 분화, 동대논총 25.

곽충구(1997), 연변 지역의 함북 길주·명천 방언에 대한 조사 연구-음운·어휘·문법 조사 자료, 애산학보 20.

김영태(1997), 경남 방언과 지명 연구, 경남대학교출판부.

김영환(1982), 조선어방언학, 평양: 김일성종합대학출판사.

신홍예(1997), 중국 심양 지역의 평안북도 방언에 대한 음운론적 연구, 석사 학위 논문(동덕여대).

이기갑·김주원·최동주·연규동·이헌중(2000), 중앙아시아 한인들의 한국어 연구, 한글 247.

이병근·정인호(1999), 중국 조선어 방언 조사, 김시준 외 한반도와 중국 동북 3성의 역사 문화, 서울대학교출판부.

정용호(1988), 함경남도 방언연구, 평양: 교육도서출판사.

執筆組(1993), 中國朝鮮語實態調查報告, 遼寧民族出版社·民族出版社.

한영목(1999), 충남 방언의 연구와 자료, 이회문화사

한영목(2000), 충남 금삼 지역어 연구, 한국문화사.

한영순(1967), 조선어방언학, 평양: 김일성종합대학출판사.

한영순(1974), 조선어방언학, 평양: 김일성종합대학출판사.

한진건(2000), 룡진 방언 연구, 민족출판사.

황대화(1986), 동해안 방언 연구 - 함북, 함남, 강원도의 일부 방언을 중심으로, 평양: 김일성종합대학출판사.

8) 이 연구가 끝난 후에 다음의 책이 간행되었다. 따라서 그 자료는 이용할 수가 없었다. 즉 최명옥, 곽충구, 배주채, 전학석(2002), 함북 북부 지역어 연구, 태학사.

이 이외에도 북의 방언 자료를 수집하기 위하여 『한글』, 『말과글』, 『조선어연구』와 같은 학술 잡지 및 각종 연구 보고서와 논저에 부록으로 수록된 북의 방언, 남에서 간행된 북 지역의 道誌, 郡誌, 面誌, 邑誌, 島誌에 수록된 북의 방언, 사전과 자료집에 수록된 북의 방언을 수집하여 이를 정리·통합하여 프로그램화할 수 있도록 하였다. 북의 방언 자료를 수집한 자료들은 주로 다음과 같은 자료에서 추출하였는바, 그 구체적인 통계를 보이면 다음과 같다.

	한글	郡誌	단행본	말과글	평북 방언사전	함북 방언사전	평안 방언 연구	계
방언형 수	8,966	25,126	9,679	4,074	11,916	15,278	2,886	77,925 항목
표제어 수	7,402	24,973	7,994	3,777	10,247	5,401	1,447	61,241 항목

4.3. 각 방언 자료집의 구조 통일

전자 파일로 구성되어 있는 각 방언 자료집은 각각 그 구조가 달라서 이를 통일하는 작업이 필요하였다. 어느 사전은 표준어를 표제항으로 한 반면, 어느 사전은 방언형을 표제항으로 하였고, 대부분의 사전에는 용례가 들어 있지 않은데 비하여 최근에 간행된 방언사전들은 상당수가 용례가 들어가 있어서 모든 구조를 통일시키지 않으면 안 되었다.

4.4. 방언 자료의 선별 문제

이들 자료들을 모두 정리한 결과 상당히 많은 방언형들을 구축할 수 있었다. 그러나 이 항목들을 모두 검색 프로그램의 자료로 활용할 것인가 아니면 그중에서 일부만 선정할 것인가가 문제였다. 그리하여 다음과 같은 두 가지 방안을 놓고 검토하게 되었다.

- (가) 이 자료집에 등재되어 있는 모든 항목을 포함하는 방법
 - (나) 이 자료집 중에서 방언량이 많은 어휘 항목만 선정하는 방법
- (가)의 방안은 몇 가지 장단점을 지니고 있다.

① 위의 조사 자료 중에서 어느 자료집은 최근에 조사된 것이고, 어느 자료집은 조사된 시점이 너무 오래전이었어서, 통시적 관점에서 볼 때, 자료의 균형성이 문제가 된다.

② 어느 조사는 도 단위의 지역만 표시하였고, 어느 조사는 군 단위까지도 표시해 주고 있어서, 조사 지역의 균일성에 문제가 있다.

③ 한국정신문화연구원의 자료집을 제외하고는, 모든 지역을 균등하게 조사한 것이 아니라는 단점이 있다. 따라서 어느 표제항에 대한 방언형은 일부 지역의 방언형만 제시될 뿐, 전국적인 방언형들을 볼 수가 없다.

④ 자료 조사가 전문가에 의해 이루어진 경우도 있지만, 비전문가가 조사하여서 표기상의 통일성이 없는 경우가 흔하였다. 예컨대 ‘학교, 학꼬, 하꼬’ 등의 표기가 동일한 방언형의 표기인지 알 수가 없었다.

⑤ 小倉進平(1944), 河野六郎(1945)와 같은 해방 전 일본인에 의해 만들어진 방언사전의 항목들을 그대로 표제어로 삼은 자료집은 옛말이나 비표준어가 표제어로 제시되어 있어서, 기본 표제어를 결정하기가 쉽지 않았다.

⑥ 체언의 곡용형이나 용언의 활용형들이 그대로 표제어로 제시되어 있는 자료집이 많이 있었다. 이 경우 기본형만을 표제어로 제시할 경우 방언의 독특한 활용형(방언형)을 알 수 없게 되는 문제가 발생할 수 있었다.

⑦ 그러나 모든 자료의 성격과 출처를 제공해 준다면 이용자들이 이용하는 데에 큰 도움을 줄 것이며 또한 실제로 다시 조사할 수 있는 것이 아닌 상황이기 때문에, 역사적 자료로서의 가치를 인정할 수밖에 없었다.

(나)의 방안은 중요한 어휘 항목만을 제시하기 때문에, 다음과 같은 장단점이 있을 수 있다.

① 이용자들이 검색하고자 하는 다양한 정보를 제공해 주지 못한다.

② 어느 것이 기본적인 어휘 항목인지, 즉 방언량이 많은 것인지를 확정하기가 수월치 않다.

③ 그러나 어휘 항목을 줄이면, 기술이 수월하여지고, 통일성을 기하기가 쉽다.

4.5. 방언 자료 정리 기준

(가) 표준어 항목의 범위

프로그램에 탑재될 표준어 항목은 다음과 같이 한정하기로 하였다.

- ① 어휘 및 관용구에 한정하고 문법 형태소와 문장은 제외하기로 하였다.
- ② 항목 수는 가능한 한 많이 하기로 하고 조사된 항목은 모두 넣기로 하였다. 왜냐하면 빈도수가 낮은 것이 오히려 방언적 특색을 더 많이 가지고 있는 것일 수도 있기 때문이라는 생각 때문이다.

(나) 표제항과 방언형

- ① 표제항(올림말)은 기본형으로 한다.
- ② 방언형은 조사된 대로 활용형이든 곡용형이든 모두 그대로 올린다.
- ③ 이러한 사항은 일러두기에서 설명해 둔다.
- ④ 활용형이나 곡용형 중에서 이해하기 어려운 방언형들은 그 방언형 뒤에 ‘^ ^’ 기호를 이용하여 관련된 문법 정보 및 어휘 정보를 설명해 준다.
- ⑤ 표제항(표준어) 앞에는 ‘#’을, 방언형 앞에는 ‘*’를 넣어 구분한다.

예)

#가까이/부/

*가까이

*가깝이

*가께

*가깝이

*가차이

*가채

*가채이

*가챇이

*거꺼이

⑥ 표제항이나 방언형이 한자로 되어 있는 것은 한글로 바꾼다. 만약에 동음이의어의 문제가 있으면 괄호를 하고 그 뜻을 한자로 표시하여 준다. 한자는 기초 한자 1,800자 내의 것만 사용한다.

(다) 자료의 정리 기준

① 각 방언형의 표준어를 표시해 준다. 그러나 방언형에 대응되는 표준어가 없을 경우에는 방언형을 표제어로 올리고 해당 방언형을 반복해서 제시해 준다.

② 예문과 뜻풀이는 필요한 경우에만 하기로 하였다. 단 대응 표준어가 없는 경우에는 가능한 용례를 제시해 주도록 하였다.

③ 성조 표시나 음장 표시는 일단 자료집에 나타나 있는 그대로 제시해 주도록 하였다.

④ 검색어에서는 그 성조 표시를 삭제하였다. 이는 검색을 용이하게 하기 위해서이다.

⑤ 각 지역 방언 중에서 다른 지역에서 전혀 사용되지 않는 특이한 형태의 방언형이나 특이한 의미의 방언형들을 별도로 조사하였다. 원래 각 연구진들은 각 도의 방언을 각각 책임지게 되어 있어서 그 자료는 각도의 연구진들이 담당하기로 하였다.

(라) 방언형의 표기

① 방언형의 표기는 일단 ‘한글 맞춤법’에 따르되, 어원을 밝히기 어려운 것은 표음주의 표기를 원칙으로 한다. 분석 가능한 경우에는 분석한 형태로 표기한다.

예)

#홍두개 ... *다듬대 → *다듬때
... *방치대 → *방치때

② 동일한 음성형(방언형)임에도 불구하고 단지 표기만 달리 해 놓은 경우는 하나의 형태로 단일화한다.

예)

#거스름돈 ... *치개돈 *치개돈 *치갯돈 → ‘*치갯돈’으로 단일화
#가잡다 ... *간잡다 *갯잡다 *가잡다 → ‘*가잡다’로 단일화

③ 한글 맞춤법에 어긋나게 표기된 것은 한글 맞춤법에 준해서 수정한다.

예)

*갯티레 → *겉티레

④ 경어법 등급이 다른 경우, 어미를 분리하여 집필한다. 이때 가급적 ‘^단어 설명^’을 통해서 경어법의 등급과 관련한 정보를 제공할 수 있도록 한다.

예)

#가다 ... *잡지 *잡지비 *잡지오다 *가우다 *갓지러....

⑤ 음장 표시가 되어 있는 자료들은 그 방언형을 그대로 둔다. 즉 음장 표시를 삭제하지 않는다.

⑥ 비음 표시는 다음과 같이 한다.

예)

어머~이, 호레~이 (단, 물결 표시 부호는 컴퓨터 자판에 있는 부호를 이용한다. 이 사실은 일러두기에서 설명해 둔다).

⑦ 고저 표시

㉠ 고저 표시를 한 방언 자료집은 다음과 같이 수정하여 표시한다.

㉡ 모든 숫자와 ‘^’ 등의 부호는 그대로 둔다.

㉢ 이 부호들은 컴퓨터 자판에 있는 것을 이용한다.

㉣ 이 부호들은 각 음절의 오른쪽에 표시한다.

㉤ 경북 방언에서는

저조 : 0(무표)

저0조 : ^(유표1)

고조 : 2(유표2)

㉥ 경남 방언(3단)에서는

저조 : 0(무표)

중조 : 1(유표1)

고조 : 2(유표2)

㉦ 경남 방언(2단)에서는

저조 : 0(무표)

고조 : 2(유표)

등으로 표시한다.

⑧ ‘에’와 ‘애’가 중화된 표기로 사용한 [E]는 ‘에’로 표기하고 그 사실을 일러두기에
서 밝혀 둔다.

⑨ 제주도 방언에 나타나는 ‘으’의 표시는 ‘아(쌍아래아 아)’ 식으로 표시하도록 한다.

⑩ 충청도 방언 중 ‘귀신’의 이중 모음성 발음은 ‘구-이신’의 표기 방식으로 표시해
주고 이러한 사실은 일러두기에 밝혀 둔다.

(마) 행정 구역 표기 문제

① 지역명을 한자로 표시한 것이 있는데(특히 제주도 방언), 이것은 한글로 바꾸어
준다.

② 행정 구역이 바뀐 곳은 현재의 행정 구역명으로 바꾸어 주되, 그 지역명이 살아
있는 곳은 그대로 살려 두도록 한다.

③ 북의 방언 지도는 북의 도별 구역과 군별 구역으로 경계만 표시한 지도와 구체적
인 군명이 기입된 군별 방언 지도를 만든다.

④ 북의 행정 구역 표시는 남에서 사용하는 행정 구역명으로 한다. 따라서 북에서
광복 이후 개편한 행정 구역명은 남에서 사용하는 행정 구역명으로 모두 바뀌 준다.
이것은 북에서 개편한 행정 구역명이 남의 주민들에게는 낯설기 때문이었다.

⑤ 중국의 방언형 분포 지역 표시는 다음과 같이 하였다. 즉 분포 지역 표시가 입력
대상 자료에 따라 조사 지역이 중국 지명으로 되어 있는 경우도 있고(조선어실태조사
보고/함경도 동북부 지역 방언 연구 등), 남과 북의 지역명으로 되어 있는 경우(조선어
방언조사보고)도 있는데 입력할 때는 입력 대상 자료집에 표시되어 있는 지역명으로
입력하였다. 특히 한자의 간자 표기에 유의하였다. 그리하여 다음 조사 지역에 대한 지
명 표시는 다음과 같이 하였다.

「길림성 교하현 천북향 영진촌」부안	→ 「길림성 영진촌」부안
「길림성 교하현 천북향 영진촌」순창	→ 「길림성 영진촌」순창
「길림성 돈화현 현유향 입신촌」조치원	→ 「길림성 입신촌」조치원
「길림성 돈화현 현유향 입신촌」홍성	→ 「길림성 입신촌」홍성

「길림성 유하현 강가점향 경기둔」수원	→ 「길림성 경기둔」수원
「길림성 유하현 강가점향 경기둔」안성	→ 「길림성 경기둔」안성
「길림성 화룡현 용문향 병동촌」명천	→ 「길림성 병동촌」명천
「길림성 훈춘현 경신향 회룡봉촌」경흥	→ 「길림성 회룡봉촌」경흥
「심양시 동릉구 훈화점향 만읍촌」용천	→ 「심양시 만읍촌」용천
「심양시 동릉구 훈화점향 만읍촌」의주	→ 「심양시 만읍촌」의주
「요녕성 개현 서해농장」철산	→ 「요녕성 서해농장」철산
「요녕성 개현 서해농장」초산	→ 「요녕성 서해농장」초산
「요녕성 무순리석채향 리석채촌」맹산	→ 「요녕성 리석채촌」맹산
「요녕성 무순리석채향 리석채촌」선천	→ 「요녕성 리석채촌」선천
「흑룡강성 계동현 계림향 계림촌」단천	→ 「흑룡강성 계림촌」단천
「흑룡강성 계동현 계림향 계림촌」함흥	→ 「흑룡강성 계림촌」함흥
「흑룡강성 수화현 홍화향 홍화촌」경주	→ 「흑룡강성 홍화촌」경주
「흑룡강성 오상현 장산향 일승촌」영덕	→ 「흑룡강성 일승촌」영덕
「흑룡강성 태래현 사리오향 서광촌」울산	→ 「흑룡강성 서광촌」울산

(바) 기타

① 하나의 형태가 두 가지 이상의 품사로 쓰일 때에는 / , / 안에 두 품사의 약호를 쓰고 그 음절 사이에는 쉼표를 사용하여 표시해 준다. (예: /명, 부/)

② 보조 동사는 본동사의 표제항 아래에 방언형으로서 나열하되, 보조 동사임을 알 수 있도록 앞에 괄호를 치고 본동사의 활용형을 그대로 표시하여 준다.

예)

보다

* 봐라

* (먹어) 봐

③ 표제항이나 방언형에 형태소 경계를 표시하기 위하여 ‘-’으로 표시한 것이 있다. 예컨대 ‘조용-하다’는 ‘-’을 지우고 ‘조용하다’로 표기한다.

④ 단어의 뜻풀이는 ‘^ ^’를 사용하여 밝힌다. 이는 나중에 프로그램을 통하여 단어의 설명 부분에 삽입된다.

⑤ 대응 표준어가 없어 뜻풀이가 필요한 방언형은 시작과 끝에 ‘% @’ 기호를 이용하여 뜻풀이를 해 준다.

⑥ 예문의 앞에는 & 표시를, 그리고 뒷부분에는 \$ 표시를 한다.

⑦ 이 프로그램은 초기에 유니코드로 하지 않고 KSC 5601의 코드로 만들기 때문에 옛한글 처리가 어려우므로 다음과 같은 방식으로 정리하도록 한다. 특히 제주도 방언형에서 유의하여야 한다. 이것은 다음번에 모두 유니코드의 글자 형태로 바꿀 것이다.

예)

ㅏ : 가(아래아 가)

ㄴ : 날(아래아 날)

4.6. 방언 자료 입력 구조

방언 자료집의 입력 구조는 다음과 같이 하였다.

```
//제목
//작성자
//작성 날짜
//검색어
//표준어
//뜻풀이
//예문
//출전(이 출전은 방언 자료집을 의미함)
//관련 어휘(동일 표준어에 대한 방언형)
//방언형(방언형의 도별 분포)
//끝
```

① 품사는 표준어 옆에 써 넣기로 한다.

② 관용구의 품사 표시는 ‘구’로 표시하기로 한다.

이에 따라 입력된 모습은 다음과 같다.

//제목 북한 방언 검색기

//작성자 ○○○○

//날짜 2002. 7. 28

//점색어 가뵈다, 가궤다, 가꺠다, 가꺰다, 가뵐다, 가부엌다, 가븨다, 가븨다, 가븨다, 가븨다, 감삭하다, 감삭허다, 개꺠다, 개꺠다, 개꺰다, 개꺰다, 거꺠다, 베삼다, 하꺠다, 해꺠다, 해꺠다, 해꺰다, 현꺠다, 해꺠다

//표준어

||뜻풀이||

//예문

출전

//관련 어휘

//방언형

*가깝다<함남>[단천, 북청]<함북>[경성, 성진, 경원, 온성, 종성, 희령, 무산]

*가접다<강원><함남>

*가 굽다<평북>

*가뱌다<함복>[길주, 명천]

*가부엌다<평북>[태천]

*가뭄다<평북>[선천]

*가뭇다<평남>[순천, 영원]<평북>[위원, 초산, 벽동, 운산, 창성, 삭주, 귀성, 박천, 정주, 선천]

*가뵈다<황해>

*갑작하다<평북>[운산, 자성, 철산, 후창, 선천, 용천, 정주, 초산, 덕천, 벽동]<평남>
[영원, 대동, 순천, 개천, 신천]

*갑삭허다<평남><평북><황해>[심천]

*개잡다<함남>[영흥, 함흥, 함주]<함북>

*개점다<경남>[진주]<함남>

*게 굽다<강원><함남><함북>

*개급다<합납>

*거쩍다<함복>[학성]

*베삽다<평복><평복>

*하깝다<함복>

*해점다<전남>

*해갑다<함복>[경흥, 무산, 경원, 학성, 길주, 명천, 경성, 온성, 종성]<함남>

*해 끝다<평북><평북>

*헛가바서^ ‘가볍다’의 활용형(가벼워서). <합복>[중성, 희명]

*헌집다<함복>[경성]

*해접다<함남>[영흥]

11월

각 자료를 정리하는 구조는 다음과 같다.

- ① 표제어 앞에는 # 표를 붙인다.
- ② 표제어 뒤에는 칸을 띄지 않고 / / 표 안에 품사 표시를 해 준다.

품사 표시는 다음과 같이 한다.

- | | |
|-------------|-------------|
| ㉠ 명사 : /명/ | ㉡ 대명사 : /대/ |
| ㉢ 동사 : /동/ | ㉣ 형용사 : /형/ |
| ㉤ 관형사 : /관/ | ㉥ 부사 : /부/ |
| ㉦ 수사 : /수/ | ㉧ 관용구 : /구/ |
| ㉨ 접사 : /접사/ | ㉩ 어미 : /어미/ |
| ㉪ 조사 : /조/ | |

- ③ 표제어는 한 줄로 쓴다.
- ④ 표제어 아래에는 방언형들을 나열하는데, 그 순서는 다음과 같다.

- ㉠ 방언형, 도명, 소구획명의 순서로 쓴다.
- ㉡ 각 방언형 앞에는 *표를 붙인다.
- ㉢ 도명은 < > 표 안에 넣어야 한다.
- ㉣ 소구획명은 [] 안에 넣는다.
- ㉤ 소구획명이 여럿일 때에는 쉼표(.)를 찍고 한 칸을 띄어 쓴 후에 나열한다.

⑤ 방언 어휘에 대응하는 표준어를 찾을 수 없는 경우에는 방언형을 표제항으로 올리고, 방언형을 제시하는 자리에 방언형으로도 제시해 준다. 이는 검색을 쉽게 하기 위해서이다. 이때 반드시 뜻을 풀이를 해 준다.

예)

#노흠쓰다/형/%노여운 기분을 행동으로 나타내다.@

*노흠쓰다<평북>

⑥ 뜻풀이에서 의미 내용이 둘 이상일 경우에는 「1」, 「2」...와 같이 구분하여 해 준다.

⑦ 해당 방언형의 출처를 밝혀 준다.

프로그램에 의해서 만들어질 파일의 형태는 다음과 같다.

#가깝다 /형/

*가죽다	<경남>[합천, 함안, 통영, 거제, 밀양, 창원, 김해, 남해]
*가직다	<경남>[거창, 함양, 의령, 남해]
*가익다	<경남>[남해]
*가깝다	<경남>[거창]
*가적다	<경남>[산청, 합천, 진양, 의령]
*개적다	<경남>[하동]
*가참다	<경남>[창원, 남해, 울주, 의령]
*가깝다	<경남>[언양]
*개죽다	<경남>[양산, 창녕]
*가죽다	<경남>[고성]
*개직다	<경남>[사천]

5. 방언 자료집의 분석

5.1. 남의 방언 자료

각 방언 자료집에 나타난 표제항 및 방언량을 보면 다음과 같다.

	표준 국어사전	서울대 방언사전	제주어 사전	전남 방언사전	경남 방언사전	조선어 사전
총 항목 수	2,441	4,391	11,340	12,444	913	5,231
명사	1,623	2,932	6,751	5,478	710	3,622
대명사	19		22	70		18
수사	9	23	24	39	6	34
동사	363	790	1,942	2,775	101	843
형용사	201	343	804	1,471	70	383
부사	205	243	760	1,605	14	303
관형사	16	7	7	78		7
조사	2	7	131	98		
감탄사	17	18	70	119		6
의존 명사	11			104		
보조 동사	3			33		
보조 형용사				8		
접사		2	48	42		
어미			670	263	1	
선어말 어미			20			
句			85			
기타				331		13

5.2. 북의 방언 자료⁹⁾

품사	항목 수
명사	7,098
동사	2,235
부사	1,066
형용사	1,058
감탄사	50
대명사	40
수사	33
조사	72
어미	155
관형사	37
구(句)	470
접사	22
총계	12,336

	항목 수
ㄱ	2071
ㄴ	741
ㄷ	1038
ㄹ	48
ㅁ	1003
ㅂ	1297
ㅅ	1631
ㅇ	1676
ㅈ	1182
ㅊ	435
ㅋ	136
ㅌ	147
ㅍ	189
ㅎ	706
총계	12,300

5.3. 중국 및 해외 방언 자료¹⁰⁾

품사	항목 수
명사	3849
동사	1133
부사	377
형용사	586
감탄사	21
대명사	28
수사	32
조사	41
어미	105
관형사	23
구(句)	122
접사	9
총계	6326

	항목 수
ㄱ	1177
ㄴ	312
ㄷ	499
ㄹ	27
ㅁ	394
ㅂ	553
ㅅ	837
ㅇ	1035
ㅈ	597
ㅊ	179
ㅋ	77
ㅌ	120
ㅍ	130
ㅎ	379
총계	6316

9) 품사별 배열의 총계가 가나다순 배열의 총계보다 36개가 더 많은 것은 한 단어가 두 가지 이상의 품사 기능을 하는 경우, 예컨대 '수사'이면서 '관형사'이거나 또는 '동사'이면서 '형용사'인 경우가 있어서 통계상에 차이가 있는 것이다.

10) 품사별 배열의 총계가 가나다순 배열의 총계보다 10개가 더 많은 것은 한 단어가 두 가지 이상의 품사 기능을 하는 경우가 있어서 통계상에 차이가 있는 것이다.

5.4. 종합 통계

	남	북	중국(해외)
표제항 수(표준어 수)	19,432	12,266	6,313
방언형 수	157,876	38,027	26,254
한 표제항의 평균 방언형	8.12	3.10	4.16
지역 표시 수	259	86	94

6. 마무리

이상으로 민족어 방언 검색 시스템을 만드는 과정에 대해 약술하였다. 실제로 일반 국민들은 스스로 사용하고 있는 말이 방언의 영역에 속하는지, 아니면 표준어나 문화어에 속하는지를 잘 알지 못할 것이며, 또한 그러한 구분을 할 필요를 거의 느끼지 못할 때가 있다. 이러한 조건에서 민족어 방언 검색 시스템이 개발되면 언중들에게 표준어나 문화어와 함께 방언의 중요성을 제시해 줌으로써 우리말에 대한 자긍심을 일으킬 수 있을 것으로 생각한다. 뿐만 아니라 오늘날 표준어나 문화어에 가리어져서 그 중요성이 감소되었던 방언들도 훌륭한 우리 민족어로서 가꾸어야 한다는 생각에서, 방언이 단순한 우리 문화유산만이 아니라, '민족 공통어'로서 대우를 해 주어야 할 요소를 자각하게 된다. 그래서 이 작업을 통해 가장 직접적으로 필요성을 느낄 수 있었던 것은 이러한 방언의 중요성이라고 할 수 있다. 그래서 점차 사라져 가는 우리 문화유산을 보호하고 발전시켜야 할 책무를 느끼는 것이다.

국어학이 연구된 이래로 많은 방언 자료집이 도별로 또는 지방별로 만들어져 왔다. 뿐만 아니라 연구자나 조사자들에 의해서 부분적으로 각 지역어를 조사하여 자료로서 남아 있는 것도 많다. 하지만 이들 자료들은 각기 흩어져 있었던 관계로 하나로 통합하여 종합 자료로 이용하는 것이 불가능하였다. 책으로 간행하여도 엄청난 양이기 때문에 사용자들의 불편을 가져 올 수밖에 없었다. 그러나 '민족어 방언 검색 시스템'의 개발로 이들 자료들이 종합화되어 그 결과물이 학계의 연구 자료로 이용될 수 있을 것이며 또한, 문화계나 신문, 방송과 같은 대중 매체에서도 매우 요긴하게 사용될 수 있

을 것이다. 뿐만 아니라, 작가들이 글을 쓸 때에도 아주 긴요하게 사용될 수 있을 것이다.

이 작업은 우리말 사전 편찬에도 커다란 도움을 줄 것이다. 각 사전에는 방언형이 다 등재되어 있지 않다. 올림말로 올라 있는 방언형에 대해, 뚜렷한 자료 제시도 않은 채, ‘~의 방언’이라고만 기술되어 있는 경우를 흔히 보는데, 이 자료를 통해 방언에 대한 더 자세한 기술이 가능하리라고 생각한다.

방언 자료를 수집, 정리, 검토하는 과정에서 제기된 문제점과 개선점을 제시하도록 한다. 지금까지 조사된 방언 자료들이 조사자들의 조사 방법, 조사 지점과 체보자의 선정 기준, 조사 질문지의 작성에 통일성이 없었고, 전사 기준도 제각각 달라서 방언 검색 시스템을 구축하는데 어려움이 많았다. 따라서 방언 자료의 활용도를 높이기 위하여 다음과 같은 점이 고려되어야 할 것이다.

① 조사 지점이 전국적이어야 하며, 더 나아가서는 조사 지점이 전 세계의 우리 민족이 사는 지역 전체가 되어야 한다.

② 일정한 원칙에 따라 조사 지점이 결정되어야 한다. 그렇지 않고서는 전국적인 방언형의 분포를 파악하기 힘들다.

③ 조사 항목이 통일되어 있어야 한다. 이것은 통일된 질문지가 있어야 함을 의미한다. 조사자는 조사자별로 질문지를 준비하여 조사하거나, 또는 남에서처럼 한국정신문화연구원에서 만든 『한국 방언 조사 질문지』를 이용하거나 하는데, 이들 질문지들을 참고로 하여 새로운 조사 질문지를 작성할 필요가 있다고 생각한다. 왜냐하면 한국정신문화연구원에서 만든 방언 조사 질문지는 그 목적이 주로 음운 조사에 치중되어 있었기 때문에, 어휘 지도를 작성할 때에는 만족스럽지 못한 경우가 많았다. 북에서 만든 조사 질문지의 내용이 궁금한데, 아직 그러한 정보에 접하지 못해 안타깝다.

④ 전사 방식이 통일되어 있어야 한다. 특히 한글 전사의 경우에는 이 문제에 유의하여야 할 것이다. 음운 표기의 전사 방식을 택하는 경우에도 종성 표기나 어간 말 자음군의 표기, 모음의 표기 등에 통일성을 기하여야 할 것이다.

⑤ 음장, 고저, 억양 등의 표기 방법도 통일되어 있어야 하며, 조사할 때에 이 요소들에 대한 조사 방침이 미리 결정되어 있어야 할 것이다. 그래야 그 전사에 통일성이 보

일 것이기 때문이다.

⑥ 특별한 음운을 가지고 있는 방언에 대한 조사에서 특별한 기호, 예컨대 아래 아자(이나 …)의 사용 등에 대한 결정이 있어야 할 것이다.

⑦ 방언 자료에는 주로 표준어에 대응하는 방언형을 조사한 것이 대부분인데, 표준어에 대응되는 단어가 없이, 특정한 한 지역에서만 사용되는 방언형들도 중점적으로 조사될 필요가 있다. 특히 『제주어 사전』을 검토해 보면 이러한 필요성을 절감하게 될 것이다.

⑧ 제보자로부터 얻은 방언형에 대한 정확한 의미 정보를 파악하기 위하여 용례가 제시되어야 할 것이다. 의미 파악이 수월하다고 파악되는 것은 방언형만 기술하여도 무방하지만, 특별한 의미(대응되는 표준어와는 다른 의미)를 가지고 있다고 인식되는 방언형에는 반드시 그 용례를 밝혀 의미 파악이 수월하도록 하여야 할 것이다.

⑨ 조사자는 한 방언형의 역사적 변이 형태들을 알 필요가 있다. 그렇지 않으면 특이한 방언형에 대한 정밀한 조사를 하기 어려울 것이다.

⑩ 해외의 우리말 방언에 관심을 가져야 할 것이다. 특히 중국의 우리말 방언은 특별한 의미를 갖는다. 중국 동북3성의 방언권은 일반적으로 요녕성의 평안도 방언권, 길림성의 함경도 방언권, 흑룡강성의 경상도 방언권 등으로 구분되고 있다. 그러나 실제로 이 지역의 방언이 과연 한반도 내의 평안도, 함경도, 경상도 방언권의 안에 포함되는 것인지, 아니면 다른 방언권에 속하는 것인지, 포함된다면 어떠한 소방언권에 소속되는지에 대한 검토가 필요한 때가 되었다.

⑪ 앞으로 방언 검색 시스템은 방언 자료들이 문자로만 전사되어, 이용자들에게 제공하는 것에서 탈피하여, 실제의 발음을 그대로 들을 수 있는 디지털 형태로도 제공되어야 할 것이며, 동시에 그 발음 장면까지도 볼 수 있도록 고안되어야 할 것이다.

⑫ 위와 같이 조사 연구된 자료들은 앞으로의 방언 조사 연구의 훌륭한 기초 자료로 제공되어 요긴하게 이용될 수 있을 것이며, ‘한민족 방언사전’이 등장할 수 있는 단초가 되기를 기대한다.

제2분과 말뚥치의 구축과 활용



빈 면

균형코퍼스에 대한 새로운 이해와 그의 제작에서 나서는 몇가지 문제

권 종 성(사회과학원 언어학연구소 실장)

균형코퍼스는 해당 언어의 특성을 전면적으로 파악할수 있게 하고 언어정보처리나 사전편찬 등 여러 분야에 광범히 리용할수 있게 한다. 이런것으로 하여 현 시기 그의 제작은 대단히 중요시되고있으며 큰 의의를 가진다.

본 논문은 균형코퍼스에 대한 새로운 이해와 그의 제작에서 나서는 몇가지 문제에 대하여 서술하려고 한다.

코퍼스는 본문코퍼스와 음성코퍼스로 크게 나눌수 있는데 여기서는 본문코퍼스에 대하여서만 이야기하려고 한다.

1. 균형코퍼스의 정의

균형코퍼스의 제작을 위해서는 우선 균형코퍼스란 무엇인가 하는것부터 알아야 한다. 여기서는 균형코퍼스를 지난 시기 이야기하던것과는 다르게 리해하려고 하는것만큼 그에 대한 정의부터 먼저 주려고 한다.

균형코퍼스는 무엇보다도 분야별로 균등하게 배분되어 묶어지는것이 아니다. 이것은 균형코퍼스는 골고루가 아니라는것을 뜻한다. 레하면 코퍼스에 넣을 언어분야를 크게 문학예술, 사회정치, 신문보도, 과학기술의 4개 분야로 크게 분류하고 문학예술을 다시 소설, 희곡, 시, 민족고전, 번역소설 등으로 중분류하였다고 하자. 이때 대분류가

4개 분야로 되었다고 하여 각각 25%씩의 량으로, 또 문학예술이 5개 분야로 중분류되었다고 하여 각각 20%씩의 량으로 배분하자는것이 아니다.

균형코퍼스는 매 분야별로 그 쓰임의 잦음도와 중요성, 해당언어에서 그가 차지하는 위치 등을 고려하여 량이 배분된것이다. 광복전에 과학기술분야가 신문보도분야나 문학예술분야보다도 사람들속에서 적게 쓰이고 또 출판물의 량도 적게 나왔다면 광복전 과학기술분야는 신문보도분야나 문학예술분야보다도 적은 량을 배분하여야 한다. 광복후에 사회정치분야가 우리 생활에서 매우 중요한 위치를 차지하게 된것만큼 사회정치분야는 광복전에 비하여 광복후에는 더 큰 비중으로 배분하여야 한다.

균형코퍼스는 다음으로 그 민족어를 대표할수 있는 《민족어의 축소판》으로 되는 것이다.

코퍼스는 리용목적에 따라 그 대상을 달리하여 작성되는것도 있다. 레하면 언어의 역사를 연구하기 위하여 작성되는 코퍼스는 력사적자료를 시대적으로 정리하여 컴퓨터처리를 진행함으로써 이루어진다. 그러나 이러한 코퍼스는 력사언어자료 하나만 놓고보면 균형문제를 이야기할수 있을지 모르나 민족어전반을 놓고보면 어느 한 분야에 지나지 않는다. 따라서 이렇게 작성된 코퍼스는 만족어전반을 대표할수 없으며 균형코퍼스로 될수 없다.

균형코퍼스는 해당 민족어전반에 대하여 균형이 잡힌 코퍼스를 의미한다. 민족어전반에 대한 균형이 잡힌 코퍼스는 그 민족어의 축소판으로 되는것이다. 원래 아무리 코퍼스를 크게 제작한다고 하여도 민족어의 모든것을 다 컴퓨터에 입력하여 자료기지를 만들수는 없다. 그러므로 그 민족어를 대표할수 있게 균형을 맞춰 수집정리한 언어자료를 컴퓨터처리하여 코퍼스를 제작하게 되는데 그때 제작된 코퍼스는 《민족어의 축소판》으로 되는것이다.

균형코퍼스는 다음으로 해당 민족어의 대규모언어모임이다.

대규모언어모임이라는 의미를 어떻게 리해할것인가. 다시 말하여 어느만한 량으로 되여야 대규모언어모임으로 되는가? 그 기준은 해당 민족어의 전모를 엿볼수 있을만한 량이라고 말할수 있다. 따라서 민족어전반을 엿볼수 있을만한 량으로 되려면 백만이나 몇백만이면 부족하고 적어도 몇억정도의 어휘는 포함되여야 한다. 이런 의미에서

놓고 볼 때 지난 시기 이야기하던 균형코퍼스와 오늘날 말하게 되는 균형코퍼스는 그 내용과 폭에서 일정하게 차이가 나게 된다는것을 알수 있다.

이상 말한것에 기초하여 균형코퍼스를 정의한다면 다음과 같이 말할수 있다. 균형코퍼스는 해당 민족어를 대표할수 있는 균형이 잡힌 대규모자료모임이다.

2. 균형코퍼스의 성격과 그 제작

- 균형코퍼스의 성격

균형코퍼스는 그가 차지하는 위치로부터 민족어의 연구나 언어정보처리, 사전편찬 등을 비롯하여 여러 분야에 광범위하게 리용될수 있는 성격을 가진다.

코퍼스는 그것이 어디에 쓰이는것인가에 따라서 그 성격이나 구성방법 그리고 자료의 내용이나 규모 등이 달라질수 있다.

균형코퍼스는 그 리용분야가 매우 다양한것으로써 그 어떤 특정한 목적에 리용되는 것이 아니다. 균형코퍼스는 해당 민족어를 대표할수 있는 균형이 잡힌 대규모언어모임인것만큼 해당 민족어의 모든 분야가 비교적 풍부하게 반영되어있다. 따라서 사용자는 이 균형코퍼스를 응용하여 여러가지 목적에 맞게 개작하고 리용할수 있다. 레하면 사전편찬을 위한 목적으로 쓰려고 한다면 균형코퍼스 전체를 리용하면 될것이고 언어력사의 어휘나 문법, 그 변천과정을 고찰하고싶다면 균형코퍼스에 포함되어있는 언어력사를 반영할수 있는 부분을 취하되 만약 거기에 반영된것이 좀 부족한 감이 있으면 더 보충하여 리용하면 된다. 신문보도어휘에 대한 실태와 그 변화에 대한것이라면 균형코퍼스에서 그에 대한 자료를 선택하여 자기 연구목적에 맞게 리용하면 된다. 이것은 언어학적 시소러스를 제작하면 그에 기초하여 정보검색시소러스나 기계번역용 시소러스를 만들어 사용하게 되는것과 마찬가지로이다.

- 균형코퍼스의 제작

균형코퍼스를 제작하려면 여러가지 문제들이 있을수 있는데 여기서는 그 몇가지만을 말하려고 한다.

균형코퍼스를 제작하려면 우선 그 총적인 설계와 세밀한 요강이 작성되어야 한다.

설계는 자료의 선택과 그 규모, 자료의 입력과 그 방식, 소요인원, 비용 그리고 단계와 해수, 제작도구와 기술개발, 태그정보의 준비 등 균형코퍼스제작에서 나서는 모든 문제들을 예견하여 세워져야 한다. 총설계는 제작하는 과정에 일부 부족점이 나타나면 수정보충하여 완성해나갈수 있다.

균형코퍼스제작은 여러 사람들이 동시에 그리고 장시간에 걸쳐 진행하는것만큼 하나의 통일요강에 기초하여 진행하여야 한다.

다음으로 균형코퍼스제작을 위하여서는 선택된 자료(총설계에 의하여 취한 자료)를 입력하여야 한다. 자료의 입력은 반드시 그 순차성을 지키면서 진행해야 하는것은 아니다. 그리고 입력방식은 키보드에 의한 타이핑방법이나 OCR에 의한 방법으로 진행할수 있다. 그러되 자료의 구입이나 여러가지 입력환경조건으로 하여 현대자료를 먼저 입력하고 그 다음 력사자료(력사자료입력을 원만히 진행할수 있는 환경이 마련된 다음)입력을 진행할수 있다. 그러나 총체적으로 균형코퍼스의 요구에 맞도록 진행하여야 한다.

다음으로 균형코퍼스는 일단 균형코퍼스로서의 면모를 갖추게 된 다음에도 그 량을 계속 늘일수 있도록 해야 한다. 균형코퍼스는 닫긴 코퍼스가 아니라 열린 코퍼스로 제작하여야 한다. 그렇게 할 때만이 균형코퍼스로서의 가치를 계속 유지할수 있다. 민족어는 계속 변화발전하게 된다. 그런데 앞에서 말한바와 같이 균형코퍼스는 민족어의 축소판으로 되어야 하는데 만약 닫긴 코퍼스로 제작하여 어느때까지의 언어로 한정시켜놓으면 그후의 언어가 반영되지 못하게 되므로 균형코퍼스로서의 자기의 가치를 상실하게 된다.

다음으로 균형코퍼스는 매 분야별로 균형을 맞추면서 보다 큰 분야에서의 균형이

잡힌 코퍼스로 되게 하는 방법으로 제작하여야 한다.

균형코퍼스와 관련한 필자의 의견을 몇가지 이야기하였다. 여기서 필자가 주장하는 균형코퍼스는 철저히 그 민족어를 대표할수 있는 《민족어의 축소판》으로 되어야 한다는것이다. 이런 의미에서 놓고본다면 지난 시기 균형코퍼스로 제작되었던 100만이나 수백만 단어정도의 량을 가진 코퍼스나 또 그 어떤 분야를 수록한 코퍼스는 비록 그자체로서 그 어떤 분야에서의 균형을 잡은 코퍼스라고 하더라도 균형코퍼스라고 말할수 없다. 그러면 지난 시기 제작하였던 백만이나 몇백만 단어정도의 량을 가진 코퍼스나 그 어떤 분야로 한정시킨 코퍼스는 무엇이라고 불러야 하겠는가 하는것이다. 이러한 코퍼스는 필자의 생각으로서는 균형코퍼스에 상대하여 《비균형코퍼스》라고 말할 수는 없고 《분야별 코퍼스》 혹은 《부문별 코퍼스》라고 말해야 한다고 본다. 왜냐하면 이러한 코퍼스는 비록 리용목적에 따라서 자기 나름으로서의 균형을 이야기는 하였으나 민족어전반에서 놓고 볼 때에는 결코 균형이 잡힌 코퍼스로 될 수 없으며 그리고 그 어떤 부문이나 분야로 한정시킨 코퍼스이기때문이다.

균형코퍼스의 제작은 민족사적인 의의를 가지는 매우 중요한 사업이므로 그 설계로부터 시작하여 완성에 이르기까지(균형코퍼스의 완성은 항상 상대적이다) 근기있게 완강하게 진행해나가야 할것이다.

민족어 말뭉치 지도 작성

- 21세기 세종계획 북한 말뭉치를 중심으로 -

소 강 춘(전주대학교 국어교육학과 교수)

1. 서론

이 글은 한국 문화관광부 21세기 세종계획 특수 자료 구축 분과에서 구축하고 있는 북한 및 해외 한국어 말뭉치 구축 사업의 결과를 중심으로 그동안 한국에서 구축한 조선의 말뭉치 구축 현황을 파악하는 데 목적이 있다.

말뭉치(corpus)는 광의로 '언어를 연구하는 각 분야에서 필요로 하는 연구 재료로서, 언어의 본질적인 모습을 총체적으로 드러내 보여 줄 수 있는 자료의 집합'을 뜻하지만, 협의로는 '일정 규모 이상의 크기를 갖추고 내용적으로 다양성과 균형성이 확보된 자료의 집합체'라는 뜻으로 사용된다(서상규 외, 1999:27).

이 정의를 토대로 적절한 말뭉치를 구축하기 위해서는 그 언어의 모습을 총체적으로 보여 주고 있는가 하는 언어의 대표성, 규모의 적절성, 내용의 다양성과 균형성 등의 조건이 고려되어야 함을 알 수 있다.

이러한 조건에 맞는 말뭉치를 구축하기 위해서는 면밀한 설계와 설계에 따른 정확한 구축, 그리고 구축된 자료에 대한 실제 분석을 통해 문제점을 파악해 다시 수정하는 과정이 반복적으로 진행되어야 한다.

이 글은 이와 같이 말뭉치 구축을 위한 이론적 연구가 아니고, 21세기 세종계획의 2단계까지 구축된 '북한 및 해외 한국어 말뭉치'에서¹⁾ '북한 말뭉치' 구축 현황을 파악

1) 이 용어는 21세기 세종계획에서 사용하고 있는 말뭉치의 이름이기 때문에 고유 명사화해서 사용하도록 하겠다.

하는 데로 범위를 한정하도록 하겠다.

2. 21세기 세종계획의 현대 국어 말뭉치 구축 지침

21세기 세종계획의 기초 언어 자료 베이스 분야의 목표는 4개 영역 - 현대 국어 말뭉치, 북한/해외 한국어 말뭉치, 옛 문헌/방언/구비 문학 말뭉치, 대역(한국어-외국어) 말뭉치 - 으로 구분하여 아래와 같은 단계별 목표를 정하고 있다.

[표 1] 21세기 세종계획 단계별 세부 내용

구 분	1단계 (1998-2000)	2단계 (2001-2003)	3단계 (2004-2007)	합 계
현대 국어 말뭉치	9,000만 (+7,500만)	1억 500만	1억 6,000만	4억 3,000만 어절
현대 국어 분석 말뭉치	1,100만 (+200만)	2,200만	4,400만	7,900만 어절
북한/해외 한국어 말뭉치	1,500만	1,500만	1,000만	4,000만 어절
옛 문헌/방언/구비 문학 말뭉치	900만	850만	850만	2,600만 어절
대역(한국어-외국어) 말뭉치	200만	400만	400만	1,000만 어절

이 계획에 근거하여 1차 연도 구축 목표를 정했고, 구축 지침을 개발했다.²⁾ 현대 국어 말뭉치 구축에 필요한 전체 구성 개관과 분야별 제한 조건 규정을 제시하면 아래와 같다.³⁾

(1) 전체 구성 개관

- 1910년 이후의 자료를 대상으로 하되, 최근 것 위주로 구축하였다.

2) 말뭉치를 구축하는 작업은 원하는 말뭉치를 설계하고, 설계에 따라 부분적으로 말뭉치를 구축하며, 그 결과를 분석하여 문제점이 발견되면 다시 설계에 반영하는 과정으로 진행되기 때문에 여기서 작성된 구축 지침은 초기 단계의 설계 부분에 해당한다.

3) 1998년 21세기 세종계획 기초 자료 구축 분과 보고서 pp. 92-3의 내용을 재인용한 것이다.

- 텍스트 분량 : 단일 텍스트일 경우 가능한 한 8만 어절 내외로 제한하였으며, 단일 텍스트가 아닐 경우 원칙적으로 10만 어절 이내로 조정하였다.

(2) 분야별 제한 조건 규정

- 가능한 한 다음의 구성 비율을 유지하도록 했다.
- 상황에 따라 구성 비율의 조정이 가능하나, 각 분야별로 규정된 분량에 대해 최저 70%, 최고 130%의 범위를 유지하도록 했다.

신문	20%
잡지	10%
책-정보	35%
책-상상	20%
기타/보충용	5%
구어/준구어	10%
총	100%

(2)의 분야별 제한 조건 규정의 세부 사항은 아래와 같이 규정되어 있다.

(3) 신문

- 다음의 내용별 구성 비율을 준수했다.
- 상황에 따라 구성 비율의 조정이 가능하나, 각 분야별로 규정된 분량에 대해 최저 70%, 최고 130%의 범위를 유지하도록 했다.

사설/칼럼	30%
종합(정치)/사회/경제/외신/북한	30%
문화/매체/생활/과학	30%
스포츠	5%
기타(해설/일지/명단/인물 등)	5%

(4) 잡지

- 잡지의 다양성을 고려, 한 분야에 치우치지 않도록 적절히 구성했다.

(5) 책 - 정보

- 단일 텍스트일 경우 가능한 한 8만 어절 내외로 제한했다. 단 한 저자의 저작물이라도 단일 텍스트가 아닌 경우 원칙적으로 10만까지 허용하며, 여러 사람의 텍스트를 모은 저작물의 경우도 원칙적으로 10만까지 허용했다.
- 다음의 내용별 구성 비율을 준수했다.
- 상황에 따라 구성 비율의 조정이 가능하나, 각 분야별로 규정된 분량에 대해 최저 70%, 최고 130%의 범위를 유지하도록 했다.

총류	15%
교육 자료	10%
체험 기술적 텍스트	15%
인문	20%
사회	15%
자연	10%
예술/취미/생활	15%

(6) 책 - 상상적 텍스트

- 장편일 경우 가능한 한 8만 어절 내외로 제한했다. 단 단편 모음집일 경우에는 원칙적으로 10만까지 허용했다.
- 운문 자료는 제외했다.
- 다음의 내용별 구성 비율을 준수했다.
- 상황에 따라 구성 비율의 조정이 가능하나, 각 분야별로 규정된 분량에 대해 최저 70%, 최고 130%의 범위를 유지하도록 했다.

장편 소설	50%
중·단편 소설	40%
동화	10%

이상은 2단계 사업이 끝난 지금까지 가능하다면 지키려고 노력한 조건들이다. ‘현대 한국어 말뭉치’ 구축 분과와는 달리 특수 자료 구축 분과에 속한 ‘북한 및 해외 한국어

말뭉치' 구축 사업은 비록 여러 가지 면에서 특수한 상황에 처해 있기는 하지만 공식적으로 조선, 중국, 구소련 등에서 사용되고 있는 자료를 구축한다는 면에서 한국 말뭉치 구축과 차이가 없다.⁴⁾ 따라서 필자는 가능한 한 이 조건을 지키려고 노력하였다.

3. '세종 북한 말뭉치' 구축 현황

이 장에서는 그동안 구축했던 조선 문어 자료 말뭉치 구축 현황과 문제점에 대해 논의한다.

[표 2] 세종 북한 말뭉치 구축 현황

구분\년도	80년대	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	총 합계	비율 (%)
신 문 32.1									116209	431992	123497	296323	356487	763413		2,087,921	32.1
잡 지 18.7	종합							152477		14498	25664	80741	39957			313,337	4.8
	문학							84879	55501		44443	207536	131419			523,778	8.0
	예술							76291	35383			154655				266,329	4.1
	인문												66144			66,144	1.0
	사회												47746			47,746	0.7
	자연											8936				8,936	0.1
	총류															0	0
책 35.5	사상	67467		21469	136806	76601		76648	77366	182542						638,899	9.8
	교육 자료	50546				15306								489176		555,028	8.5
	체험 기술적			16890	56277	68524		35363								177,054	2.7
	인문				45331	124275	109023	15461	34260							328,350	5.1
	사회		44268	126530		67394	47592	46851	6558							339,193	5.2
	자연				55813		15403	39066	37412		94957					242,651	3.7
	예술/취미/생활				33822											33,822	0.5
	장편 소설											277821				277,821	4.3
	중·단편 소설	33882		27018	61038	138158	37301	194203			38967	33882				564,449	8.8
	동화							24684			10165	2409				37,258	0.6
기 타																	
계	151895	44268	191907	389087	474952	224625	432276	469243	389635	446490	337693	1062303	641753	1252589		6,508,716	100

4) 예산과 인력이 허락된다면 모든 자료를 동등하게 구축하는 것이 바람직한 방향이다. 이는 민족어의 통합이라는 문화적 측면에서도 중요하지만, 정보화된 자료의 교류라는 산업적 측면에서도 아주 중요한 일이다.

3.1. 대분류에 따른 논의

[표 2]의 '세종 북한 말뭉치' 내용 중에서 대분류된 내용과 세종계획 처음에 세웠던 (2)의 분야별 제한 조건 규정의 내용을 비교하면 아래와 같다.

[표 3] 분야별 제한 규정과 세종 북한 말뭉치 실제의 대비

구 분	규 정	조선 실제	비 고
신문	20 %	32.1 %	+12.1 %
잡지	10 %	18.7 %	+8.7 %
책-정보	35 %	35.5 %	+0.5 %
책-상상	20 %	13.7 %	-6.3 %
기타/보충용	5 %		-5.0 %
구어/준구어	10 %		-10.0 %

[표 3]의 증감 내용을 보면 세종 북한 말뭉치에서 신문과 잡지가 많이 구축되었고, 기타와 구어 자료는 거의 구축되지 않았다.⁵⁾ 구어 자료는 한국에서 순수 구어 말뭉치를 구축하고 있기 때문에 조선의 구어 말뭉치는 처음부터 구축 대상에서 제외되었다. 신문 자료가 많이 구축된 것은 남북 교류의 어려움 때문에 현실적으로 다양한 자료의 구입이 어려운 실정에서 가장 보편적으로 당시 언어 양상을 보여 주는 자료가 신문이라는 판단에 기인한 것이다.

책-상상 자료는 6.3%가 부족하고, 잡지 자료는 8.7%가 많이 구축되었다. 이는 조선 종합 문예지인 '조선문학, 청년문학, 아동문학, 통일문학'에서 장르별 구분 없이 책 전체를 입력했기 때문이다. 이 종합지에는 아래와 같이 하위 장르로 분류될 수 있는 내용들이 포함되어 있다.

5) '거의'라는 용어를 사용한 것은 구어/준구어 항목이 전혀 구축되지 않은 것으로 되어 있지만 후술할 문학, 예술 잡지에 포함된 희곡이 상당량 있어, 준구어는 어느 정도 구축되었다고 보아야 하기 때문이다.

[표 4] 조선문학(2000, 5)

	전체	단편	시	수필	전설	산문
글자	144581	62376	14060	9500	32559	26086
낱말	32734	13996	3752	2190	7123	5673
줄	5300	1929	1442	266	943	726
문단	3436	1143	1442	108	479	270
쪽	124	44	33	7	22	18
원고지	898.6	363.4	153.6	53.5	185.1	143.4
점유율		43.1%	9.7%	6.6%	22.6%	18.0

[표 5] 청년문학(1997. 3)

	전체	단편	시	수필	희곡	산문
글자	97367	31085	23134	10164	11285	21698
낱말	22022	6745	5807	2349	2494	4627
줄	4623	938	2526	323	421	587
문단	3884	522	2526	175	328	336
쪽	103	22	58	8	10	14
원고지	704.5	180.5	265.5	60.3	72.4	126.1
점유율		31.9%	23.8%	10.4%	11.6	22.3

[표 6] 아동문학(2001. 6)

	전체	단편	시	수필	전설	우화	속담
글자	56204	21906	12968	9102	5265	6298	665
낱말	11620	4667	2437	1925	1114	1321	156
줄	2665	667	1325	274	176	201	28
문단	2233	444	1325	177	131	137	25
쪽	61	16	31	7	4	5	1
원고지	388.6	126.3	136.6	52.4	31.4	37.1	4.8
점유율		49.0%	23.1%	16.2%	9.3%	11.2%	1.2%

[표 7] 통일문학(1999. 3)

	전체	장편	단편	시	희곡	전설	산문
글자	202639	32537	63760	30933	25083	6094	44232
낱말	44372	6741	13728	7581	5557	1351	9414
줄	7908	922	1908	3306	858	192	1288
문단	6086	424	1022	3306	612	114	614
쪽	176	21	44	76	20	5	31
원고지	1337.4	181.3	365.8	348.5	153.5	25.3	252.9
점유율		16.1%	31.5%	15.2%	12.4%	3.0%	21.8%

[표 4-7]은 각 문예지의 평균이 아니고 가장 전형을 보이는 것으로 생각되는 한 권의 내용만을 제시한 것이다.⁶⁾ 각 표에서 ‘단편’은 중편 소설과 단편 소설을 포함하고, ‘전설’은 전설과 사화를 포함하고, ‘산문’에는 평론, 작가 소개, 작품 소개, 강단, 기타 산문적인 내용 모두를 포함하고 있다. 따라서 이 내용으로 전체 문예지의 장르별 분량을 정확하게 예측할 수는 없지만, 편집자들이 장르별로 배당하는 분량은 어느 정도 추측이 가능하다.⁷⁾

정보-상상의 자료 부족 부분은 [표 4-7]의 문학 잡지에 포함된 장편, 단편, 시, 우화, 속담 등이 충당해 준다. 위 표의 장편, 단편, 시를 합한 평균 61.1%와 전설 11.6%는 책-상상으로 계산되어야 하고, 아동문학의 우화 11.2%와 속담 1.2%도 상상의 동화로 계산되어야 한다. 그리고 수필은 책-정보-체험 기술로 계산되어야 한다.

이렇게 전환이 이루어진다면 잡지-문학 523,337 어절의 72.7%인 380,786 어절이 책-상상으로 전환되고, 아동문학 153,261의 12.4%인 19,004 어절이 책-상상으로 전환된다. 결국 책-상상은 1,279,318 어절로 19.7%가 되어 원래 계획했던 양보다 0.3%가 모자라게 된다.

6) 이 표는 문서 편집기 한글에서 제공하는 문서 통계 정보를 이용한 것이다.

7) 장기적으로 본다면 마크업된 정보에 따라 어느 정도 파악은 가능하지만, 각 잡지나 신문에 포함된 하위 장르에 따른 정확한 분류가 필요하다.

3.2. 소분류에 따른 논의

먼저 '세종 북한 말뭉치'의 신문 자료는 소분류되지 않았다. 이는 한 달 분량의 신문을 전체 입력했기 때문이다.

둘째로 잡지는 한국 말뭉치와는 달리 '종합, 문학, 예술, 인문, 사회, 자연'으로 구분했다. '종합, 문학, 예술'의 경우는 종합지적 성격이 강하지만, '인문, 사회, 자연'의 경우는 주로 해당 분야의 전문 학술지를 대상으로 하고 있다.⁸⁾ 인문의 경우는 문학, 예술 분야의 종합지에 상당량의 인문 분야 글들이 있어 문제가 없다. 그러나 자연 분야의 자료가 부족한 것은 자료 입력의 어려움 때문이었다. 앞으로는 이 분야를 많이 보충하려 한다.

셋째로 책-정보에 '사상' 분야를 추가했다. 이는 조선에서 이 분야에 대한 많은 글들이 나오고 있고, 그 내용 면에서도 충분히 독립적인 지위를 확보할 수 있다는 판단 때문이다. 총류는 거의 구축이 안 되었고, 자연 분야는 잡지의 경우와 동일한 이유 때문이다. 예술 분야는 예술 분야의 종합지 속에 이 분야를 충족시킬 수 있는 부분이 있다고 판단된다. 교육 분야의 2002년 자료는 금년도 사업으로 추진될 내용을 제시한 것이다.

넷째로 책-상상에서 장편 소설이 단편이나 중편 소설보다 적게 구축된 것은 다양한 작가의 다양한 작품을 구축하기 위한 것이었다.

4. 결론 및 제언

말뭉치 구축은 1단계로 체계적인 설계가 이루어져야 하고, 2단계로 설계에 따라 실제로 말뭉치가 구축되고, 3단계로 구축된 자료를 이용한 언어학 연구와 자연어 처리 관련 분야의 산업에 적용함으로써 그 문제점을 찾아내고, 그 문제점은 다시 말뭉치 구축의 첫 단계부터 새롭게 수정할 수 있는 기회를 부여하는 형태로 진행되어야 한다. 21세기 세종계획에서 구축된 말뭉치는 이제 3단계 작업의 결과가 나오기 시작하면서,

8) 이는 한국이나 조선이나 이들 분야에 다양한 종합지적 성격의 잡지가 없기 때문이기도 하고, 필자가 잡지를 구할 수 있는 길이 한정되어 있기 때문이기도 하다.

그 문제점을 분석하여 첫 단계부터 수정 작업을 진행해야 하는 단계에 진입했다고 할 수 있다.

2단계 6차 연도 사업이 완료되고, 3단계 7차 연도 사업이 시작된 현 시점에서 [표 2]의 세종 북한 말뭉치 구축 현황에서 볼 수 있는 바와 같이 약 650만 어절의 원시 말뭉치가 원래 계획과 큰 차이 없이 구축되었다. 그러나 이 말뭉치 전체를 활용한 연구는 아직 활발히 이루어지지 못한 실정이다. 현재까지 이 말뭉치를 이용한 연구 결과는 필자가 ‘한국, 조선, 중국의 국어 교과서에 나타난 어휘적 특징’이라는 주제로 3국 교과서의 어휘적 특징을 살펴본 논문과 남한-북한 및 북한-남한 정서법 변환기를 만들기 위해서 필자가 구축한 대역어 사전, 필자가 SoriTagger를 이용하여 주석 말뭉치를 구축하는 과정에 만들어진 조선어 주석 학습용 사전,⁹⁾ 김진규(2004)에서 북한어 형태소 분석기를 개발하고, 평가하는 데 사용되었다.¹⁰⁾

말뭉치 구축의 문제는 다양한 자료의 구입이 어렵다는 점이다. 이는 앞으로 남북 교류가 확대되면 자연스럽게 해결될 수 있을 것이라고 생각된다. 그러나 그 이전에 남북 연구자들이 서로의 필요에 의해 구축된 말뭉치가 있다고 한다면 교환하는 방법도 모색할 단계가 아닌가 생각한다.

<참고 문헌>

- 강범모(1997), 『전자 텍스트 부호화 개설: TEI 라이트』, 고려대학교 민족문화연구소.
 김영택 외(2001), 『자연 언어 처리』, 생능출판사.
 김진규(2004), ‘어절 형성 전이도와 HMM을 이용한 북한 문화어 품사 태깅 시스템’, 연세대 석사 논문.
 김홍규·강범모(1997), 『한글 사용 빈도의 분석』, 고려대학교 민족문화연구소.
 문영호 외(1993), 『조선어빈도수사전』, 평양:과학백과사전종합출판사.

9) 2004년 6월 2일 현재로 195,532어절이 구축되어 있다.

10) 형태소 분석기를 개발하고 평가하기 위해서는 정제된 주석 말뭉치가 필요하다. 현재 세종 북한 말뭉치 속에는 약 74만 어절의 주석 말뭉치가 포함되어 있다.

사회과학원(1992), 『조선말대사전』, 평양:사회과학출판사.

서상규·한영균(1999), 『국어 정보학 입문』, 태학사.

소강춘(2002), ‘정보 처리 프로그램에 대하여-SynKDP를 중심으로-’, 한국어와 정보화, 태학사.

소강춘(2004), ‘한국, 조선, 중국 국어 교과서에 나타난 어휘 특징에 대한 정보학적 연구’, 국제고려학회 평양학술대회 발표 원고.

소강춘·김진규(2001), ‘통합형 한글 자료 처리기(SynKDP)에 대하여’, 국어사 자료 연구 2호.

소강춘·김진규·박진양(2001), ‘북한 자료를 이용한 국어 정보 처리의 실제’, 이광호 교수 회갑 기념 논총.

연변 사회과학원(1992), 『조선말사전』, 연변인민출판사.

홍윤표 외(1997), 『21세기 세종계획(안)-국어 정보화 중장기 발전 계획』, 문화체육부 보고서.

Hausser, R. 1999, Foundations of Computational Linguistics, Springer.

Ratnaparkhi, A. 1996. “A Maximum Entropy Part-Of-Speech Tagger”. Proceedings of the Conference on Empirical Methods in Natural Language Processing, 133-142.

[부록] 세종 북한 말뭉치(연차별 구축 현황)

[1998년 : 406,581]

파 일 명	용 량	책 이 름	분 류
1NH84001.HWP	67,467	김일성 저작집 28(1973.1~1973.12)	정보-사상
1NE95000.HWP	32,222	단편소설집 봄날의 미소	상상-단편
1NE94001.HWP	20,733	단편소설집, 고향의 모습	상상-단편
1NE93001.HWP	31,953	문학작품집(1990-1991)	상상-단편
1NH96001.HWP	16,298	위대한 령도자 김정일 동지의 사상리론 법학 2	정보-사상
1NH96002.HWP	9,736	위대한 령도자 김정일 동지의 사상리론 경제학 3	정보-사상
1NH96011.HWP	17,510	위대한 령도자 김정일 동지의 사상리론 철학 1	정보-사상
1NH94001.HWP	6,356	이비인후과수술학II	정보-자연
1NH96006.HWP	6,558	인구학 개론	정보-사회
1NH94001.HWP	9,047	조선과학기술발전사(해방후편1)	정보-자연
1NH95000.HWP	11,961	조선교육사6	정보-사회
1NCA1A20.HWP	37,921	북한 국어 교과서	교과서
1NCA2A10.HWP		북한 국어 교과서	교과서
1NCA3A10.HWP		북한 국어 교과서	교과서
1NCA4A10.HWP		북한 국어 교과서	교과서
1NCA4A20.HWP		북한 국어 교과서	교과서
1NCA0Z10.HWP	14,459	북한 조선어 교과서	교과서
1NCA1Z10.HWP		북한 조선어 교과서	교과서
1NCA2Z10.HWP		북한 조선어 교과서	교과서
1NCA3Z10.HWP		북한 조선어 교과서	교과서
1NCA4Z10.HWP		북한 조선어 교과서	교과서
1NCA5Z10.HWP		북한 조선어 교과서	교과서
1NAV9798.HWP	116,209	노동신문 18 97.5.2 ~ 98.1.31	신문

[1999년 : 1,310,759]

과 일 명	용 량	책 이 름	분 류
2NA98V12.HWP	431,992	로동신문	신문
2NB96010.HWP	29,230	천리마 1	잡지-종합
2NB96011.HWP	25,991	조선영화	잡지-예술
2NB96020.HWP	25,516	조선영화 2	잡지-예술
2NB96040.HWP	32,421	천리마 4	잡지-종합
2NB96050.HWP	30,457	천리마 5	잡지-종합
2NB96060.HWP	32,644	천리마 6	잡지-종합
2NB97060.HWP	14,898	조선예술 6	잡지-예술
2NB97070.HWP	20,485	조선예술 7	잡지-예술
2NE91000.HWP	27,018	새로운 연대	상상-단편
2NE92000.HWP	37,547	은정	상상-단편
2NE93000.HWP	35,270	쌍바위	상상-단편
2NE93001.HWP	31,118	백두산의 눈	상상-단편
2NE93002.HWP	39,817	사랑의 로정	상상-단편
2NE95000.HWP	32,294	하늘과 땅	상상-단편
2NE95001.HWP	37,066	삶의 신념	상상-단편
2NF92000.HWP	23,491	인민의 태양	상상-단편
2NG91000.HWP	16,890	적후의 다섯소년	상상-단편
2NG92000.HWP	56,277	영광의 날에	상상-장편
2NG93000.HWP	68,524	만고의 령장	상상-장편
2NG95000.HWP	35,363	푸른계절	상상-단편
2NH91000.HWP	21,469	주체사상에 대하여	정보-사상
2NH92000.HWP	6,254	혁명적당건설의 근본문제에 대하여	정보-사상
2NH92001.HWP	33,822	음악예술론	정보-예술
2NH93001.HWP	57,192	철학론문집 20	잡지-인문
2NH94000.HWP	28,609	재미있는 민속놀이	정보-인문
2NH96000.HWP	33,822	위대한 령도자 김정일동지의 사상리론, 문예학 2	정보-사상
2NH96001.HWP	37,412	면역생물학	정보-자연
2NH97000.HWP	2,445	만경대혁명학원은 주체의 혈통을 이어나갈 핵심 공간양성기지이다	정보-사상
2NH97001.HWP	2,235	주체철학은 독창적인 혁명철학이다	정보-사상
2NH97002.HWP	3,120	위대한 수령 김일성동지의 조국통일유훈을 철저 히 관철하자	정보-사상

[2000년 : 1,301,684]

파 일 명	용 량	책 이 름	분 류
5NA00V01.HWP	295323	로동신문 2000년 1월분(1-21일)	신문
3NA99V01.HWP	123497	로동신문 99년 1월분(1-9일)	신문
3NB96A03.HWP	29912	천리마 96.3	잡지-종합
3NB96C03.HWP	24784	조선영화 96.3	잡지-예술
3NB96F02.HWP	21831	청년문학 96.2	잡지-문학
3NB96J03.HWP	16469	아동문학 96.3	잡지-문학
3NB96J04.HWP	14828	아동문학 96.4	잡지-문학
3NB96J05.HWP	15786	아동문학 96.5	잡지-문학
3NB96J07.HWP	15965	아동문학 96.7	잡지-문학
3NB97F03.HWP	24599	청년문학 97.3	잡지-문학
3NB97F71.HWP	30902	조선문학97.7	잡지-문학
3NE89001.HWP	33882	절정	상상-단편
3NH90001.HWP	44268	조선정치제도사	정보-사회
3NH91001.HWP	65384	조선교육사	정보-사회
3NH91002.HWP	61146	조선수군사	정보-사회
3NH92004.HWP	55813	치료수기	정보-자연
3NH92002.HWP	57562	세기와 더불어(1)	정보-사상
3NH92003.HWP	72990	세기와 더불어(2)	정보-사상
3NH93004.HWP	31068	희극영화와 웃음	정보-인문
3NH93002.HWP	36015	조선현대아동소설 연구(해방후편)	정보-인문
3NH93003.HWP	67394	자연보호사업을 강화할데 대하여	정보-사회
3NH94004.HWP	20603	조선구전문학개요	정보-인문
3NH94002.HWP	47592	김일성주석과 민족대단결	정보-사회
3NH94003.HWP	59811	조선인민의 반침략투쟁사	정보-인문
3NH96002.HWP	34260	일제의 조선침략사	정보-인문

[2001년 : 1,414,648]

과 일 명	용 량	책 이 름	분 류
4NA01V01.HWP	260,427	로동신문 17일	신문
4NA01W01.HWP	27,753	통일신보 3일	신문
4NA01X01.HWP	77,307	민주조선 7일	신문
4NB00A07.HWP	12,978	금수강산(2000) 07	잡지-종합
4NB00A08.HWP	13,042	금수강산(2000) 08	잡지-종합
4NB00A12.HWP	13,878	금수강산(2000) 12	잡지-종합
4NB00D01.HWP	13,546	조선여성(2000) 01	잡지-종합
4NB00D02.HWP	13,798	조선여성(2000) 02	잡지-종합
4NB00D03.HWP	13,499	조선여성(2000) 03	잡지-종합
4NB00F01.HWP	33,264	조선문학(2000) 01	잡지-문학
4NB00F02.HWP	33,436	조선문학(2000) 02	잡지-문학
4NB00F04.HWP	31,201	조선문학(2000) 04	잡지-문학
4NB00G01.HWP	8,936	수의출산(2000) 01	정보-자연
4NB00I02.HWP	16,393	아동문학(2000) 02	잡지-문학
4NB00I05.HWP	18,051	아동문학(2000) 05	잡지-문학
4NB0F011.HWP	23,033	조선예술(2000) 01	잡지-예술
4NB0F021.HWP	22,879	조선예술(2000) 02	잡지-예술
4NB0F041.HWP	23,894	조선예술(2000) 04	잡지-예술
4NB0F051.HWP	24,064	조선예술(2000) 05	잡지-예술
4NB0F061.HWP	24,338	조선예술(2000) 06	잡지-예술
4NB0F071.HWP	25,447	조선예술(2000) 07	잡지-예술
4NB98A03.HWP	14,498	금수강산(1998) 03	잡지-종합
4NB99A03.HWP	13,930	금수강산(1999) 03	잡지-종합
4NB99A04.HWP	11,734	금수강산(1999) 04	잡지-종합
4NE00000.HWP	94,244	한나의 메아리 제1부(장편소설)	상상-장편
4NE00001.HWP	93,804	터전(장편소설)	상상-장편
4NE94000.HWP	16,568	위인의 초상	상상-단편
4NF00000.HWP	10,724	정일봉, 리준식 시집	상상-단편
4NH00440.HWP	42,112	통일문학 44	잡지-문학
4NH92J00.HWP	45,331	주체의 미학과 조형미	정보-인문
4NH93J02.HWP	76,601	김정일선집 02	정보-사상
4NH95000.HWP	39,066	조선포자식물 6-1(조류편 3)	정보-자연
4NH95001.HWP	92,621	헝스피어희곡선 2	상상-단편
4NH97J09.HWP	75,234	김정일선집 09	정보-사상
4NH99430.HWP	44,443	통일문학	잡지-문학
4NI00000.HWP	2,409	큰 대문 열리는 소리(그림동요집)	상상-동화
4NI99000.HWP	10,165	조선민족의 원시조 단군(그림책)	상상-동화

2002년 :

[원시 말뭉치 : 1,013,213]

파 일 명	용 량	책 이 름	분 류
5NA02L01.HWP	31,953	문학신문	신문
5NA02P01.HWP	236,069	평양신문	신문
5NA02V01.HWP	149,673	로동신문	신문
5NB0F205.HWP	33,079	조선문학 5	잡지-문학
5NB1A110.HWP	14,962	금수강산 10	잡지-종합
5NB1E102.HWP	24,056	경제연구 2	정보-사회
5NB1E103.HWP	23,690	경제연구 3	정보-사회
5NB1E247.HWP	24,995	김일성종합대학학보	잡지-종합
5NB1F106.HWP	16,747	아동문학 6	잡지-문학
5NB1F206.HWP	36,692	조선문학 6	잡지-문학
5NE00B01.HWP	23,158	고향을지켜	상상-단편
5NE95B01.HWP	24,684	로보트승리호	상상-단편
5NH95001.HWP	34,890	민족의 운명과 김정일령도자	정보-사회
5NH95002.HWP	76,648	김일성선집	정보-사상
5NH97001.HWP	99,508	김정일선집 12	정보-사상
5NH97003.HWP	15,461	문학예술에바친	정보-인문
5NB1F107.HWP	17,191	아동문학 7	잡지-문학
5NB1F212.HWP	34,800	조선문학 12	잡지-문학
5NB9G103.HWP	41,282	생활과건강장수 3	정보-자연
5NB9G104.HWP	53,675	생활과건강장수 4	정보-자연

[주석 말뭉치 : 256,312]

파 일 명	용 량	책 이 름	분 류
4NA01V01.HWP	24,074	로동신문	
4NB98A03.HWP	14,498	금수강산 3	
4NB99A03.HWP	13,930	금수강산 3	
4NH99430.HWP	44,443	통일문학 43	
NB96F020.HWP	21,831	청년문학 2	
NB96J030.HWP	16,469	아동문학 3	
NB97F071.HWP	30,902	조선문학 7	
NB960100.HWP	29,230	천리마 1	
NB960110.HWP	25,991	조선영화	
NB970700.HWP	20,485	조선예술 7	
NCA4A200.HWP	37,921	조선어 특강교재 4학년	

2003년 :

[원시 말뭉치 : 566,646]

파 일 명	용 량	책 이 름	분 류
6NA02V01.HWP	345,716	로동신문 2002.1.11-31	신문
6NB01D05.HWP	13,255	조선녀성 5	잡지-종합
6NB01D06.HWP	12,789	조선녀성 6	잡지-종합
6NB01F02.HWP	40,364	조선어문 2	잡지-인문
6NB01G03.HWP	25,780	력사과학 3	잡지-인문
6NE00B01.HWP	89,773	붉은 산줄기	상상-장편
6NE99B01.HWP	38,967	금수산 기념궁전..	상상-단편

[주석 말뭉치 : 394,486]

파 일 명	용 량	책 이 름	분 류
6TNA1X01.HWP	85,317	민주조선	
6TNB0A08.HWP	10,186	금수강산 08	
6TNB0D03.HWP	14,981	조선녀성 03	
6TNB0F02.HWP	39,018	조선문학 02	
6TNB1F12.HWP	40,608	조선문학 12	
6TNCA0Z1.HWP	9,140	조선어 교과서 문장기초	
6TNCA1A2.HWP	3,520	국어 교과서 1학년	
6TNCA1Z1.HWP	889	조선어 교과서 1	
6TNCA2A1.HWP	7,899	국어 교과서 2학년	
6TNCA2Z1.HWP	1,015	조선어 교과서 2	
6TNCA3A1.HWP	11,102	국어 교과서 3학년	
6TNCA3Z1.HWP	1,781	조선어 교과서 3	
6TNCA4A1.HWP	11,786	국어 교과서 4학년	
6TNCA4A2.HWP	12,293	국어 교과서(특강교재)	
6TNCA4Z1.HWP	2,468	조선어 교과서 4	
6TNCA5Z1.HWP	3,959	조선어 교과서 5	
6TNE0000.HWP	100,865	한나의 메아리	
6TNE0001.HWP	10,000	터전(약100,000 중 일부)	
6TNE0B01.HWP	27,659	고향을 지켜	

말뭉치 자료 정보 부착과 남북 문법의 통합 방안

권 재 일(서울대학교 언어학과 교수)

kwonjil@snu.ac.kr

1. 머리말

우리말의 일반 언어학 연구뿐만 아니라 전산 언어학 연구에 있어 말뭉치(코퍼스) 구축의 필요성과 중요성은 새삼스럽게 강조할 필요는 없을 것이다. 최근에 대규모의 말뭉치 구축 사업이 본격적으로 이루어지고 있는 것과, 남북 언어학자가 민족어의 연구와 발전을 위해 공동의 말뭉치 구축 사업에 참여하고 함께 논의하는 것은 대단히 바람직한 일이라 하겠다.

말뭉치 구축 사업과 관련하여 전제되어야 할 과제들은 많이 있다. 먼저 자료의 객관성을 보장하는, 자료의 정확성과 대표성을 유지하는 방법론을 개발하는 것이 중요하다. 또한 자료의 범용성을 보장하는 일도 중요하다. 이를 위하여 말뭉치 자료를 선택하여 입력하는 일, 논리 정보와 언어 정보를 부착하는 일, 나아가서 구축된 자료를 유지·보수하면서 보급하는 일, 모두 중요한 과제라고 생각한다. 그 가운데 말뭉치를 구축할 때 언어 정보, 즉 문법 정보를 부착하는 방법론 역시 중요한 과제라고 생각한다. 따라서 오늘 발표에서는 말뭉치 자료에 문법 정보를 부착하기 위해 필요한 남북 문법의 표준화 방안에 대하여 생각해 보고자 한다. 즉, 오늘 발표는 그간 남과 북에서 이루어진 우리말 문법 연구의 기술 방법을 대조하여 서로 같은 점과 다른 점을 밝히고, 이를 바탕으로 남북의 문법 표준화를 위한 바람직한 방안을 제시하는 것을 목적으로 한다. 이것은 궁극적으로 남북 언어의 통합을 지향하고자 하는 하나의 시도라고 하겠다.

그간 남북의 학술 교류가 활발해진 이래 남북 언어의 같은 점과 다른 점에 대한 논의가 꾸준히 이어져 왔다. 차이는 언어 구조에서 찾을 수 있을 뿐만 아니라, 언어를 기술하는 방법에서도 찾을 수 있다. 그러나 하나의 언어를 서로 다른 방법으로 기술하는 것은 극복되어야 할 과제이다. 이는 언어 연구라는 측면에서도 그러하거나와 통일을 앞두고 있는 이즈음에 통일을 대비하는 측면에서도 그러하다.

이를 위해 먼저 그간 남북에서 발간된 이론 문법서와 최근 논저를 대상으로 말뭉치 자료에 문법 정보를 부착하기 위해 필요한 문법 사항, 주로 품사론과 형태론을 대조하여 서로 다른 점을 찾아 앞으로 문법 기술을 표준화하기 위한 발판을 마련하고자 한다. 그러므로 오늘 제시하는 것은 어디까지나 논의를 위한 발판이지, 합의 단계에 도달한 표준 방안은 아닐 것이다.

2. 남북 문법의 품사론 대조

남-문법에서는 단어를 조어론과 품사론으로 나누어 연구하는데, 조어론에서는 단어 구성 방법에 대해, 품사론에서는 각 품사의 성격과 아울러 조사와 어미의 문법 기능을 중심으로 연구한다. 그러나 북-문법에서는 단어를 단어조성론, 품사론, 형태론, 셋으로 나누어 연구한다. 남-문법의 조사와 어미, 파생 접사에 해당하는 문법 형태를 모두 “토”로 설정하고 이를 품사론에서 따로 떼 내어 형태론에서 연구한다. 그리고 토의 첨가 여부를 품사 체계의 주요 기준으로 삼고 있다.

2.1. 품사 분류

먼저 남북의 우리말 품사 체계를 대조해 보기로 한다. 남-문법은 평면적으로 체언, 관계언, 용언, 수식언, 독립언이라는 큰 부류를 설정하고, 하위 범주로 9 품사를 설정한다. 그러나 북-문법은 토의 첨가 여부에 따라 두 가지, 곧 토가 첨가되어 형태 변화 체계를 갖춘 부류와 그렇지 못한 부류로 가르치고, (1) 형태 변화 체계를 갖춘 부류는 체언적 단어 부류와 용언적 단어 부류로 나누며, (2) 형태 변화 체계를 갖추지 못한 부류는 수식어적 단어 부류와 독립어적 단어 부류로 가른다. 이러한 네 가지 부류를 다시 명

사, 대명사, 수사; 동사, 형용사; 관형사, 부사; 감동사로 나누어 모두 8개 품사를 설정한다(고신숙 1987:57, 정순기-리금일 2001:397).

1. 남-문법의 품사 체계

- (1) 체언: 명사, 대명사, 수사
- (2) 관계언: 조사
- (3) 용언: 동사, 형용사
- (4) 수식언: 관형사, 부사
- (5) 독립언 : 감탄사

2. 북-문법의 품사 체계

- (1) 형태 변화 체계를 갖춘 단어 부류
 - (1-1) 체언적 단어 부류: 명사, 대명사, 수사
 - (1-2) 용언적 단어 부류: 동사, 형용사
- (2) 형태 변화 체계를 갖추지 않은 단어 부류
 - (2-1) 수식어적 단어 부류: 관형사, 부사
 - (2-2) 독립어적 단어 부류: 감동사

이를 통해 보면, 두 분류 사이에 차이가 거의 없음을 알 수 있다. 다만, 북-문법에서는 남-문법의 관계언, 즉 조사를 설정하지 않고, 토라는 범주를 따로 세운다. 그리고 남-문법의 감탄사를 감동사라 부른다. 그러나 정순기-리금일(2001:48)에서는 감동사와 함께 감탄사라는 용어도 제시하고 있어, 용어의 통일은 그리 문제가 되지 않는다.

2.2. 문법 형태의 성격

앞에서 살펴본 바와 같이 북-문법의 품사 분류에서는 남-문법의 조사를 설정하지 않고, 남-문법의 조사와 용언의 어미를 묶어 토라는 범주를 따로 세운다. 이 점이 남북 문법 체계의 큰 차이라고 하겠다. 이러한 우리말 문법 형태의 품사 귀속 문제는 우리

말 문법 연구사의 오랜 논의의 결과이다.

체언과 관련한 문법 형태, 용언과 관련한 문법 형태를 각각 A, B라 할 때, (1) A, B 모두를 단어로 보는 관점, (2) A, B 모두를 단어로 보지 않는 관점, (3) A만 단어로 보고 B를 단어로 보지 않는 관점이 있다. 북-문법의 토가 (2)-유형이며, 남-문법의 조사와 어미가 (3)-유형이다. 그러나 양쪽 주장 모두 확고한 논리를 가지고 있으며, 한편으로는 한계도 지니고 있다. 이 문제는 앞으로 남북 문법 표준화의 가장 큰 과제라고 생각한다. 물론 과거나 현재 남-문법에서도 이들을 (2)-유형으로 기술한 바 있어, 앞으로 남북 모두가 적극적인 관심을 가지고 논의를 전개할 필요가 있다고 하겠다. 다만, 정보 부착의 편의성을 감안한다면, “토”라는 큰 범주를 두고 그 아래 “조사:체언토”와 “어미:용언토”를 나누어 설정하는 것이 한 방법이 될 것이다.

2.3. 명사, 대명사, 수사

남북 문법에서 명사의 하위 범주는 다음과 같다.

남: (1) 보통 명사/고유 명사

(2) 자립 명사/의존 명사

북: (1) 보통명사/고유명사

(2) 활동체명사/비활동체명사

(3) 완전명사/불완전명사, 단위명사, 어근적명사

북-문법의 명사에서 활동 능력을 가지는 대상이 활동체명사이며, 그렇지 않은 것이 비활동체명사이다. 활동체명사에는 ‘에게, 한테’가 결합하고, 비활동체명사에는 ‘에’가 결합하는 차이가 있다. 남-문법에서는 이를 의미 자질 [유정성]의 있음과 없음에 따라 유정 명사/무정 명사로 구분하나, 명사의 분류 체계에 포함하지는 않고 있다. 명사의 의미 자질까지 정보로 부착할 경우에는 이 범주도 포함시킬 수 있을 것이다.

남-문법의 자립 명사와 의존 명사는 북-문법의 완전명사와 불완전명사에 해당한다. 용어의 차이는 있지만, 개념은 일치한다. 특히 북-문법은 불완전명사의 하위 범주로 단

위명사를 따로 설정한다. 단위명사에 대한 정보를 따로 부착할 필요는 당연히 있다고 생각한다. 그만큼 독자적인 문법 특성을 지니기 때문이다. 어근적명사 역시 정보를 부착할 필요가 있다고 본다. ‘국제, 원시, 용감, 충실’과 같이 자립명사에서 다른 품사로 넘어가거나 명사의 잔재 기능을 가지고 복합어 구성에 관여하는 어근적명사는 독특한 기능을 수행하고 있기 때문이다. 남-문법의 전자 사전 기술에서는 ‘어근’을 이미 중시하고 있다.

남북 문법에서 대명사의 하위 범주는 다음과 같다.

남: (1) 인칭 대명사

(2) 지시 대명사

북: (1) 사람대명사(인칭대명사)

(2) 가리킴대명사(지시대명사)

(3) 물음대명사(미정-의문대명사)

북-문법에서 기술하는 물음대명사(김동찬 2002:56의 용어로는 미정-의문대명사)에는 ‘누구, 아무, 무엇, 얼마, 몇, 어디, 언제’ 등이 있다. 이들에 대한 정보를 부착하는 것 역시 필요하다고 생각한다.

남북 문법에서 수사의 하위 범주는 다음과 같다.

남: 양수사/서수사

북: (1) 구성상 특성: 고유어수사/한자말수사, 단순수사/합성수사

(2) 의미상 성격: 수량수사/순서수사

북-문법에서 기술하는 구성상 특성에 의한 정보를 부착할 필요성은 없어 보인다. 왜냐하면, 우리말 수사는 문맥에 따라 고유어와 한자 말의 선택이 분명하다는 특성이 있지만, 고유어/한자어 구분은 수사에만 국한된 것이 아니기 때문이다. 그리고 단순수사와 합성수사는 조어법:단어조성법의 문제이기 때문이다.

2.4. 동사, 형용사

남북 문법에서 동사의 하위 범주는 다음과 같다.

남: (1) 자동사/타동사, 양용 동사(능격 동사, 중립 동사)

(2) 자립 동사/보조 동사

북: (1) 자동사/타동사

일반 자동사/타동사

특수 자동사/타동사

(2) 자립적동사/보조적동사

북-문법에서 따로 특수적인 자동사와 타동사를 설정하는데, 여기에서 불완전한 자동사, 불완전한 타동사, 그리고 이중적 유형에 있다. 이중적 유형은 남-문법의 양용 동사(능격 동사, 중립 동사)를 말한다. 그러므로 이중적 유형은 양용 동사 개념에 포함시키고, 불완전 자동사, 타동사에 대한 정보는 따로 제시하는 것이 좋을 것이다.

1. 불완전한 자동사

(감기를) 앓다, (다리를) 쉬다

2. 불완전한 타동사

을/에 대하여 생각하다/설명하다, 와/를 만나다

3. 자동사/타동사의 이중적 유형

집단을/집단이 움직였다, 설명을 그쳤다/비가 그쳤다

남북 문법에서 형용사의 하위 범주는 다음과 같다.

남: (1) 정상 형용사

(2) 지시 형용사

북: (명시적인 하위 범주 없음)

북-문법에서는 성질, 상태 등으로 형용사의 의미를 제시하지만, 특별히 우리말 형용사는 구조-문법적으로 따로 분류하지 않는다고 하였다(정순기-리금일 2001:38). 그러나 성상 형용사, 지시 형용사에 대한 정보는 제시하는 것이 좋겠다.

2.5. 관형사, 부사

남북 문법에서 관형사의 하위 범주는 다음과 같다.

- 남: (1) 성상 관형사
 (2) 지시 관형사
 (3) 수량 관형사
 북: (1) 성질관형사
 (2) 분량관형사

남북 문법의 관형사에는 성상:성질, 수량:분량이라는 용어의 차이가 있는 한편, 품사 분류상에 한 가지 특징적인 차이가 있다. 남-문법의 수량 관형사 일부와 지시 관형사를 북-문법에서는 각각 수사와 대명사로 분류한다. 남-문법에서는 ‘한 사람, 두 사람’의 ‘한, 두’를 수량 관형사라 하여 관형사로 규정하고 있으나, 북-문법에서는 ‘한 마리, 두 척, 구 개월, 아홉 상자’에서 ‘한, 두, 구, 아홉’을 순서수사라 하여 수사로 규정하고 있다. 또한 ‘이 사람, 그 집, 저 책’의 ‘이, 그, 저’를 남-문법에서는 지시 관형사로 규정하고 있으나, 북-문법에서는 가리킴대명사(지시대명사)로 규정하고 있다. 이들은 의미를 기준으로 보면 수사나 대명사임에 틀림없으나, 기능을 기준으로 보면 관형사이다. 체언은 다른 체언 앞에서 항상 관형어:규정어의 성격을 가지기 때문에 이들을 따로 관형사로 설정하기보다는 북-문법의 기술에 따라 각각 순서수사와 가리킴대명사(지시대명사)로 보는 것이 타당할 것이다.

남북 문법에서 부사의 하위 범주는 다음과 같다.

- 남: (1) 성분 부사
 성상 부사, 상징 부사(의성 부사, 의태 부사), 지시 부사, 부정 부사

(2) 문장 부사

양태 부사, 접속 부사

북: (1) 일반부사

행동부사, 성질부사(상태부사), 양태부사(문장부사)

(2) 상징부사(의성어, 의태어)

(3) 보조적부사

부정부사, 의존부사

(4) 접속부사(이음부사)

남북 문법의 부사는 비록 큰 틀은 차이가 있지만, 하위 범주는 거의 일치한다. 남-문법에서는 성분을 수식하느냐, 문장을 수식하느냐에 따라 둘로 나누었으나, 북-문법은 평면적으로 여러 부류로 나누었다. 성상 부사:행동부사/성질부사, 상징부사, 부정 부사, 양태부사, 접속부사는 그대로 남북이 일치한다. 그러나 남-문법의 지시 부사는 북-문법에 없지만 따로 제시할 필요가 있으며, '아들은 아버지와 함께 떠났다'의 '함께'와 같은 북-문법의 의존부사 역시 따로 제시할 필요가 있다고 생각한다.

2.6. 감탄사:감동사

남북 문법에서 감탄사/감동사의 하위 범주에 대해서는 모두 명시적으로 제시하지는 않지만, 남-문법에서는 (1) 감정 감탄사, (2) 의지 감탄사로 분류하기도 하고, 북-문법에서는 (1) 말하는 사람의 감정을 나타내는 감동사, (2) 말하는 사람의 의지나 요구를 나타내는 감동사, (3) 상대방의 이야기에 대한 태도를 나타내는 감동사로 분류하기도 한다. 그러나 감탄사:감동사의 하위 범주 정보를 따로 제시할 필요는 없다고 생각한다.

3. 남북 문법의 형태론 대조

남-문법에서 기술한 “조사”와 “어미”, 북-문법에서 기술한 “토”를 바탕으로, 남북 문법의 문법 범주 체계를 대조하여 살펴보기로 하겠다. 이들 형태론적 방법으로 문법 범

주를 실현하는 수단을 대조하면 다음과 같다.

<북-문법>	<남-문법>	(예)
격토	격조사	이/가, 의, 을, 에, 로, 와
도움토	(보조사)	은, 도, 만
복수토	(접미사)	들
맺음토	종결 어미	다, 느냐, 어라, 자
이음토	접속 어미	고, 으나, 면, 으니까
규정토	관형사형 어미	는, 은, 을, 던
상황토	부사형 어미	게, 도록, 듯이
상토	파생 접사	이/히/리/기
존경토	선어말 어미	시
시간토	선어말 어미	는, 었, 겠
강조토	(어미의 일부)	서, 썬, 다가, 까, 끔
바꿈토	명사형 어미	음, 기
	서술격 조사	이(다)

3.1. 격 - 격 조사:격토

남북 문법의 격 조사:격토에 의한 격의 하위 범주는 다음과 같다(학교 문법 2002:96, 리근영 1985:116-117, 정순기-리금일 2001:53-58).

남: 주격, 서술격, 목적격, 보격, 관형격, 부사격, 호격

북: 주격, 대격, 속격, 여격, 위격, 조격, 구격, 호격, 절대격

남북 문법의 격을 대조하면, 우선 용어가 남-문법은 주격, 목적격, 부사격처럼 통사적 격 명칭인데 비해 북-문법은 대격, 속격, 여격처럼 형태적 격 명칭이다. 이들은 각각 주격(남):주격(북), 목적격:대격, 관형격:속격, 부사격:여격/위격/조격/구격, 호격:호격에 대응한다. 남-문법학자 가운데는 북-문법처럼 형태적 격 명칭으로 기술하는 예가 있으므로, 형태적 격 명칭으로 통일하는 것이 어렵지 않을 것이다.

남-문법의 부사격이 북-문법에서는 여격/위격/조격/구격 등으로 세분되어 있다. 남-문법에서도 부사격을 위치격/방편격/비교격 등으로 세분하는데, 이들을 세분하는 것은 정보 부착에 유의할 것이다. 그리고 북-문법에서 설정하는 절대격을 남-문법에서는

설정하지 않으며, 북-문법에서는 남-문법의 보격을 따로 설정하지 않는데, 절대적은 영-형태로 실현되어 정보 부착에 관여하지 않기 때문에, 그리고 보격은 그 타당성에 문제가 있기 때문에 설정하지 않는 것이 좋겠다.

남-문법의 서술격은 '이다'인데, 이것 역시 격으로 인정하기 어렵다. 북-문법에서는 바꿈토(용언화토)로 기술한다. 우리말의 '이다'는 문법적 특성에 논의의 여지가 많은 것으로, 잠정적으로 '이다'에는 특별한 표지를 부여하는 것이 좋겠다.

3.2. 보조사:도움토

남-문법에서는 보조사를 특정한 어휘 의미를 가지는 것으로 기술하는 반면, 북-문법은 도움토가 '포함, 제한, 지정, 구획, 강조, 망라, 양보, 고사' 등과 같은 문법 범주를 실현한다고 기술한다. 그러나 이들을 문법 기능을 수행하는 개념으로 보기보다는 의미 특성을 실현하는 것으로 보는 것이 더 타당하다고 생각한다. 따라서 세분된 의미/기능 정보를 부착하기보다는 보조사:도움토라는 단일 범주를 부착하는 것이 더 나을 것이다.

3.3. 수 - 복수 접미사:복수토

복수를 표시하는 형태 '들'에 대해 북-문법에서는 복수토로 기술하고, 남-문법에서는 복수 접미사로 기술한다. 그 성격을 어떻게 보든지 복수 정보는 표시를 하는 것이 좋겠다.

3.4. 문장 종결법:범범주 - 종결 어미:맺음토

남-문법의 문장 종결법은, 언어 내용 전달 과정에서 청자에 대하여 화자가 가지는 태도 또는 의향을 실현하는 문법 범주이다. 문장을 끝맺는 어미로 실현되기 때문에 마침법, 문장 종결법으로 기술되어 왔으며, 문법 기능으로서 의향을 실현하기 때문에 의향법으로도 기술되어 왔으며, 한편으로는 서법, 문체법으로도 기술되어 왔다.

1930년대 이래 지금까지 우리말 문법 연구의 최대 성과인 최현배의 『우리말본』(1937/1971:262-280)에서는 '마침법(종지법)'이라는 범주를 설정하고 배품끝(서술형),

물음꼴(의문형), 시킴꼴(명령형), 피임꼴(청유형)로 하위 범주를 설정하였다. 허웅(1975:486-, 1983:225-)에서도 '마침법(의향법)'이라는 범주를 설정하고 서술법, 물음법, 시킴법, 피임법 네 가지의 하위 범주를 설정했다. 남기심(1973:52-, 2001:385-)에서 역시 의향서법에는 평서형, 의문형, 명령형, 청유형의 네 가지가 있을 뿐이라고 하였다. 발표자 역시 권재일(1992:제5장)에서, (1) 청자에 대하여 요구함이 있음/없음, (2) 행동 수행이 있음/없음이라는 두 기준을 바탕으로 문장 종결법의 기본 하위 범주를 서술법, 의문법, 명령법, 청유법 등 넷으로 체계화한 바 있다.

그러나 이러한 기본 하위 범주를 네 가지로 설정하는 체계에 동의하지 않는 다른 연구도 꾸준히 이어오고 있다. 대표적으로 고영근(1974:120-)을 들 수 있다. 무의지적 서법이 표시되는 설명법, 의문법, 감탄법과 의지적 서법이 표시되는 명령법, 허락법, 공동법, 약속법, 경계법을 제시하였다. 다분히 의미적인 측면에서 설정할 수 있는 최대의 하위 범주를 제시한 셈이다. 그 이후 고영근(1993)에서는 “평술문, 의문문, 명령문, 제안문, 감탄문”을 설정하였다. 서정수(1994:250-)에서는 문말서법을 설정하고 서술법, 약속법, 의문법, 명령법, 청유법 등 하위 범주를 다섯으로 정하였다. 학교 문법(1985:110, 1996:83, 2002:171-3)에서는 문장 종결법을 설정하고서, 평면적으로 평서법, 감탄법, 의문법, 명령법, 청유법 등 다섯을 세워 체계화하여 교육하고 있다.

북-문법에서 용언의 맺음형이 실현하는 범주 가운데 “형태론적으로 양태성을 표현하는 범주로서 행동(또는 상태)과 현실과의 관계에 대한 이야기하는 사람의 태도를 나타내는 문법범주”를 범범주라 하는데(리근영 1985:199-), 범범주의 하위 범주로 “알림, 물음, 추김(권유), 시킴(명령)”을 기본으로 설정하였다(리근영 1985:199-). 그리고 “약속”과 “감탄”은 범범주의 한 갈래로 설정하기 어렵다고 하면서 그 근거를 제시한 바 있다(리근영 1985:204-5). 김동찬(2002:177-)에서도 서술 목적에 따라 문장을 직설법(알림법), 의문법(물음법), 권유법(추김법), 명령법(시킴법)으로 하위 체계화하고, 정순기-리금일(2001:89-)에서도 범범주를 알림법(직설법), 물음법(의문법), 추김법(권유법), 시킴법(명령법)으로 설정하였다.

남북 문법에서 보면 하위 범주를 어떻게 설정하는가 하는 것이 문제이다. 다음 네 범주에 대해서는 용어 외에는 남북 문법에서 차이가 없다. 서술법:알림문(서술문)/직

설법(알림법), 의문법:물음문(의문문), 명령법:시킴문(명령문), 청유법:추김문(권유문). 문제가 되는 하위 범주는 이른바 감탄:느낌과 약속이다. 약속과 느낌에 대한 북-문법의 설명은 분명하다. 약속은 오직 두 개의 형태만이 있어 문법 형태의 체계가 서지 않는다고 하였다. 느낌 역시 그러할 뿐만 아니라, 느낌은 진술자의 주관적, 평가적 계기가 반영되는 문장이기 때문에 범범주와는 성격이 다른 문장 유형으로 규정한다.

그러나 문법 현상은 문법 특징으로 설명하는 것이 바람직하다고 본다. 그리고 형태 범주와 통사 범주(문장 범주)가 일치한다면 문법이 더욱 더 체계적일 것이라 생각한다. 이에 대한 가장 분명한 현상은 우리말의 직접 인용문:바로옮김법이 간접 인용절:풀어옮김법으로 안길 때 나타난다. 직접 인용문이 어떠한 종결 어미로 맺었더라도 간접 인용절이 될 때에는 4 개의 어미로만 중화되어 나타난다. 이에 대해서 일찍이 남기심(1973:52-)에서 언급한 바 있는데, 서술문(이른바 평서, 약속, 감탄 모두)이 안길 때에는 ‘-다’만, 의문문이 안길 때에는 ‘-으냐’만, 명령문이 안길 때에는 ‘-으라’만, 청유문이 안길 때에는 ‘-자’만 허용된다. 따라서 문장 종결법:범범주의 하위 범주는 다음과 같이 표준화할 수 있을 것이다.

1. 서술법(알림법)
2. 의문법(물음법)
3. 명령법(시킴법)
4. 청유법(추김법)

이와 아울러 청자 높임법:말차림법의 등급 표시도 함께 고려해야 하는데, 남북 문법 모두 같다. 남-문법의 청자 높임의 등급은 ‘하십시오, 하오, 하게, 해라’와 ‘해요, 해’를 설정하고 있고, 북-문법의 말차림법의 등급은 ‘높임<하십시오, 해요>, 같음<하오, 하게, 해>, 낮춤<해라>’으로 설정하고 있다(리근영 1985:209-213, 정순기-리금일 2001:91-93).

3.5. 접속:이음 — 접속 어미(연결 어미):이음토

북-문법에서 이음토로 실현되는 연계 체계는 그 성격에 따라 둘로 나눈다(리근영

1985:183-5). 별림과 매임이 있는데, 이것은 남-문법의 대등과 종속에 해당한다. 별림에는 ‘합침, 맞세움, 가림’ 관계가 있는데, 이것은 남-문법의 ‘나열, 대조, 선택’ 관계에 해당한다. 매임에는 ‘조건, 양보, 원인, 순차, 수단, 방식, 의도, 목적’ 관계가 있는데, 이것 역시 남-문법의 종속 관계와 같다. 따라서 접속:이음의 하위 체계는 일부 용어를 제외하고는 일치하고 있어 표준화에 어려운 문제는 없다.

한편 남-문법의 이른바 보조적 연결 어미도 접속 어미에 포함하여 기술하면 될 것이다.

3.6. 내포 : 체언형바꿈/규정/상황

— 명사형/관형사형/부사형 : 체언형바꿈토/규정토/상황토

남-문법에서 내포절 구성에 명사절, 관형사절, 부사절을 설정한다. 여기에는 각각 명사형 어미, 관형사형 어미, 부사형 어미가 관여한다. 북-문법에서는 이들의 범주를 각기 따로따로 설정한다.

남-문법의 명사형 어미는 북-문법의 체언형 바꿈토에 해당한다. 체언형은 용언의 말줄기에 바꿈토 ‘음’, ‘기’가 결합하여 실현되는 것을 말한다(오다-음/오기, 높다-높음/높기). 한편, 북-문법에서는 바꿈토를 통한 용언형도 설정한다. 용언형은 남-문법의 서술격 조사(혹은, 지정사) ‘이다’에 대응한다.

남-문법의 관형사형 어미는 북-문법의 규정토/없음토에 해당한다. ‘과거, 과거지속, 현재, 미래’를 뜻하는 토로서 남-문법의 어미와 다를 바 없다. 또한 남-문법의 부사형 어미는 북-문법의 상황토/꾸밈토에 해당한다.

한편, 인용절 구성에 관여하는 인용 표지 ‘-고, -라고’ 등에 대한 표준화도 필요하다. 남-학교 문법에서는 인용을 나타내는 부사격 조사의 한 가지로 기술한다. 북-문법에서는 이를 이음토의 한 가지로 기술한다. 어떤 경우이든, 인용 표지 정보는 부착해야 할 것이다.

이와 같이 서로 다른 내포절 구성의 체계를 어떻게 표준화할 것인지는 더 많은 검토가 필요하다. 복합문을 구성하는 방식에 초점을 맞추어 접속과 내포로 나누고 이들을

기능에 따라 다시 명사절, 관형사절, 부사절로 나눈 남-문법의 체계도 타당하며, 토의 분류 체계에 따라 규정토, 상황토, 바꿈토로 설정한 북-문법의 체계도 의의가 있다.

3.7. 높임:존경, 시제:시간 — 선어말 어미:끼움토에 의한 범주

남-문법에서 주체 높임의 선어말 어미 ‘시’를 북-문법에서는 끼움토인 존경토로 기술한다. 남-문법에서 시제의 선어말 어미 ‘었, 겠’ 등을 북-문법에서는 끼움토인 시간토로 기술한다. 정보 표지에서 “높임”, “시제”(편의상, 과거, 현재, 미래, 회상:과거지속으로 나누어)로 부착하면 될 것이다.

3.8. 사동법:사역, 피동법:피동 — 파생 접미사:상토

남-문법에서는 형태적 사동법과 피동법에는 각각 사동 접사와 피동 접사가 관여한다. 이들 접사는 파생 접사 ‘이/히/리/기’ 등이다. 그런데 이들을 북-문법에서는 상토로 규정한다. 따라서 상토에 의해서 사역상과 피동상을 실현한다고 본다.

3.9. 강조

북-문법에서 강조토로 규정하는 ‘서, 써, 다, 다가, 까’ 등은 남-문법에서 따로 설정하지 않는다. ‘에서, 으로서, 으로써, 에다가, 어다가, 니까’와 같이 한 문법 형태의 일부로 보아, 더 이상 나누지 않는다. 역사적으로는 분리할 수 있겠지만, 현대 우리말에서는 하나의 문법 형태로 녹아 붙은 것으로 보는 것이다.

3.10. 접사

조어법:단어조성론의 영역에 포함될 내용이기기는 하지만, 문법 정보로서 접사를 설정할 필요가 있다. 접사에 접두사와 접미사를 두는 것은 남북이 같다. 각각 명사화 접사, 동사화 접사, 형용사화 접사, 부사화 접사 등을 설정할 수 있을 것이다.

4. 맺음말

말뭉치 구축 사업과 관련하여 남북이 함께 풀어야 할 과제는 많이 있다. 그 가운데 말뭉치를 구축할 때 문법 정보를 어떻게 부착할 것인가 하는 방법론 역시 중요한 과제라고 생각한다. 그래서 오늘 발표에서는 말뭉치 자료에 문법 정보를 부착할 때 필요한 남북 문법의 표준화를 위한 바람직한 방안을 제시하고자 하였다. 이것은 궁극적으로 남북 언어의 통합을 지향하고자 하는 하나의 시도라고 하겠다.

이를 위해 남북의 최근 논저를 대상으로 말뭉치 자료의 문법 정보 부착을 위해 필요한 문법 사항을 대조하여 차이점을 찾아 문법 기술의 표준화를 시도하였다. 이러한 시도를 발판으로 삼아 이제 앞으로 더 많은 논의와 실제 작업을 통해 가능한 사항부터 하나씩하나씩 합의해 가는 노력을 아끼지 말아야 할 것이다.

<참고 문헌>

<북-문법서>

- 고신숙 1987, 「조선어리론문법(품사론)」, 과학,백과사전출판사.
- 김동찬 1987, 「조선어리론문법(단어조성론)」, 고등교육출판사.
- 김동찬 2002, 「조선어실용문법」(1999년 저작), 도서출판 박이정.
- 김용구 1986, 「조선어리론문법(문장론)」, 과학,백과사전출판사.
- 리근영 1985, 「조선어리론문법(형태론)」, 과학,백과사전출판사.
- 정순기-리금일 2001, 「조선어문법편람」, 도서출판 박이정.

<남-문법서>

- 문교부 1985, 「고등학교 문법」, 대한교과서주식회사.
- 교육부 1996, 「고등학교 문법」, 대한교과서주식회사.
- 교육인적자원부 2002, 「고등학교 문법」, (주)두산.

- 고영근 1974 현대 국어의 종결 어미에 대한 구조적 연구, 「어학 연구」 10-1.
- 고영근 1993, 「우리말의 총체 기술과 문법 체계」, 일지사
- 권재일 1992, 「한국어 통사론」, 민음사.
- 남기심 1973, 「국어 완형 보문법 연구」, 계명대학 한국학연구소.
- 남기심 2001, 「현대 국어 통사론」, 태학사.
- 남기심-고영근 1985/1995, 「표준 국어 문법론」, 탑출판사.
- 서정수 1994, 「국어 문법」, 뿌리깊은나무.
- 최현배 1937/1971, 「우리말본」, 정음사.
- 허 응 1975, 「우리 옛말본」, 샘문화사.
- 허 응 1983, 「국어학」, 샘문화사.
- 허 응 1995, 「20세기 우리말의 형태론」, 샘문화사.
- 허 응 1999, 「20세기 우리말의 통어론」, 샘문화사.

코퍼스를 리용한 다의성처리방법에 대하여

김 철 호(사회과학원 언어학연구소 연구사 학사)

문장안에서의 단어들사이의 의미관계는 일반적으로 의존관계에서 생각할수 있겠지만 의존관계가 있지 않거나 거리가 상당히 먼 단어들사이에도 그 어떤 의미관계가 있다고 볼수 있다. 이것을 심리학에서는 련상 혹은 련상추리(추론)이라고 하는데 여기서는 련상관계라고 할수 있겠다.

론문에서는 문장안에서의 단어들사이의 련상관계에 의한 의미적조합성에 기초하여 다의적단어들이 문장에서 어느 의미로 쓰이였는가 하는 실질의미를 얻을수 있다는것을 제기하면서 이것을 코퍼스를 리용하여 해결하기 위한 방법을 한가지 소개한다.

다음의 문장을 생각해 보자.

양과 염소들을 모두 우리에 몰아넣었다.

여기서 《양》, 《염소》, 《우리》는 문장속에서의 의존관계는 없지만 《양》과 《염소》는 다같이 풀을 먹는 소과에 속하는 짐짐승이며 한편 짐승우리와 같은곳에 넣어 기른다는 측면에서 련상관계에 놓여있다. 즉 련관어들이다.

이 련상관계는 《양》과 《염소》에 대하여 다음과 같은 결론을 내릴수 있게 한다.

염소:

첫째로 《염소》는 화학주기계의 제7족에 있는 염소(Cl)가 아니라는것(조선말대사전에서는 동음이의어로서 어깨번호로 구별하고있다. 그러나 쓰임에서는 동음이의어나 다의

어나 형태상 차이없는것으로 하여 다같이 다의어로 취급한다).

둘째로 조선말대사전에서의 ①의 의미로서 소과에 속하는 집짐승의 뜻이라는것.

양:

첫째로 《서양》의 의미를 나타내는 양(洋)이 아니라는것.

둘째로 조선말대사전에서 ①의 의미로서 소과에 속하는 집짐승의 뜻이라는것.

컴퓨터에 의하여 우와 같은 결론을 내리자면 컴퓨터에 《양》, 《염소》, 《우리》가 련상관계에 놓여있다는 련상관계의 정보가 사전에 지식원천으로 내장되어있어야 한다.

이러한 련상관계의 정보를 일정한 형태로 컴퓨터에 넣어두고 그것을 리용하여 다의성이나 의미적매성 등을 처리하려는 시도는 이미전부터 연구되고 진척되어왔다. 그러나 그 연구들의 대부분이 다 어떤 협소한, 한정된 범위에서만 련상관계를 밝히고 그것에 기초하여 다의성처리를 위한 모델을 만드는것이 중심적이였다. 최근에와서는 컴퓨터의 능력강화에 따라 대규모적인 련상관계를 형성하려는 움직임이 보인다. 문제는 이 대규모적인 련상관계를 수동적으로 작성한다는것이 대단히 품이 드는 일이라는데 있다. 이로부터 인간사전(조선말대사전과 같은 일반리용자사전)의 주석의미에 대한 비교층들, 사전네트워크에 의한 의미교차점검출 등의 방법들을 보면 현재 자연속에 존재하는 지식원천을 그대로 리용하려고 하고있다. 여기서는 이미 얻어진 코퍼스(의미정보를 부가하지 않은 코퍼스)를 지식원천으로 삼고있다.

△ 련상관계의 정보를 어떤 형태로 정하겠는가

련상관계의 정보형태를 보기에 앞서 우선 다의어 W에 대하여 그 실질의미를 어떻게 밝히는가를 보자.

앞에서 서술한바와 같이 문장속의 단어의 의미는 그 주위의 단어들과의 련상관계에 의한 의미적조합성에 의하여 알수 있다. 따라서 다의어 W가 일정한 의미로 사용될 때 어떤 단어들과 공기하기 쉬운가를 사전에 조사해서 그 공기하기 쉬운 단어들을 다의어 W의 공기어로서 컴퓨터에 넣어두고있다가 입력문에 다의어 W가 나타나면 그 주

위의 단어들을 모두 컴퓨터안의 다의어 W의 공기어들과 비교하여 W의 실질의미를 알아낼수 있다.

그러면 매 단어의 매 의미에 따라 공기어를 만들어야 하는가, 또 그 공기어들을 코퍼스로부터 자동적으로 뽑아낼수 있겠는가.

이건은 매우 힘든 문제이다.

이 난점은 단어의 다의성을 사전과 같은 주석이 아니라 다음과 같은 특별한 형태로 정의하여 극복할수 있다.

모든 단어들을 의미적으로 몇개의 묶음으로 분류한다. 다의적인 단어라면 그 단어는 여러 묶음에 중복되어 속하게 된다. 예를 들어 묶음을 사람, 집짐승, 도구, 음악, 가구, 지역 등으로 분류하고 거기에 해당하는 의미로서 단어들을 소속시키면 《양》은 《서양》을 나타내는 의미로서 지역에도 들어가고 《양(羊)》의 의미로서 집짐승에도 들어가며 《어질고 순한 사람》의 의미로서 사람에도 들어가게 된다.

이러한 자료는 시소러스의 일정한 준위를 가로 잘라서 그 매 웃마디를 묶음으로 대신하고 마디의 하위에 위치하는 모든 단어들을 그 묶음의 소속으로 하여 얻을수 있다.

다의성을 이렇게 정의한다면 결국 다의성의 처리는 복수개의 묶음에 속하는 단어(다의어)가 문장속에 나타날 때 거기서 어느 묶음의 의미로 쓰이였는가를 결정하는것으로 된다. 그리고 다의성 처리에 필요로 되는 자료는 매 단어의 매 의미에 따르는 공기어가 아니라 매 묶음에 대한 공기어로 된다.

연상관계의 정보형태를 이렇게 하면 의미정보가 없는 코퍼스를 가지고도 능히 다의성처리를 위한 정보를 얻을수 있는 가능성이 생긴다.

△ 코퍼스로부터의 연상관계정보 작성

먼저 코퍼스에서 매 묶음별로 모든 단어의 출현위치를 조사한다. 다음 그 위치에서 앞뒤로 일정한 범위의 단어들을 선정하여 공기어로 한다. 선정하는 범위는 수십부터 백정도안팎에서 적당히 할수 있다(반복되는 공기어는 피한다).

이것으로서 매 묶음에 따라 모든 단어에 대한 공기어가 얻어진다. 이 자료에 기초해

서 입력문에 다의어가 나타나면 그 앞뒤의 단어들을 다의어의 공기어들과 묶음별로 비교하여 적절한 묶음을 결정한다. 즉 그 묶음의 의미가 다의어의 실질의미로 된다.

그러나 이상의 정보만 가지고서는 다의성처리를 원만히 진행할수 없다. 그것은 매 묶음에 따르는 공기어속에 적절치 못한 단어도 끼여들어있어 다의성처리를 약하게 혹은 그 이상으로 방해하기때문이다. 예를 들면 《양》이라는 단어는 지역의 묶음, 집짐승의 묶음, 사람의 묶음에 소속되어있는데 코퍼스로부터 집짐승묶음속의 양에 대한 공기어를 자동선정한다면 그것은 결국 집짐승의 의미로 쓰일 때의 공기어도, 지역의 의미로 쓰일 때의 공기어도 함께 들어가게 하는 현상을 초래한다. 다의성처리를 박로 하자면 이 잡음적인 요소의 영향을 가능한껏 감소시켜야 한다.

이를 위하여 묶음 i 에 따르는 매 공기어 W 에 대하여 다음과 같은 무게판정을 진행한다.

$$\log \frac{P(W|\text{묶음 } i)}{P(W)} = \log \frac{(\text{묶음 } i \text{의 공기어안의 } W \text{의 출현회수}) / (\text{묶음 } i \text{안의 공기어총수})}{(\text{코퍼스안의 } W \text{의 출현회수}) / (\text{코퍼스안의 단어총수})}$$

이 식은 단어 W 가 코퍼스전체에 비하여 어느만한 무게를 가지고 묶음 i 의 공기어로 나타나는가를 표시하는데 그 값이 크면 클수록 묶음 i 에서는 중요한 공기어로 된다고 생각한다.

이러한 무게판정을 진행하면 적절한 공기어들의 무게는 적절치 못한 공기어들의 무게보다 훨씬 크게 될것이다.

그 이유는 다음과 같다.

묶음 i 에 속하는 공기어들은 일반적으로 그 묶음안의 단어들과 많이 관계하는 것으로 해서 출현회수가 많을것이다. 반대로 적절치못한 잡음적인 공기어들은 그 단어와만 공기한다는 제한이 붙는다. 이것은 적절치 못한 공기어는 많아도 그 개개의 출현회수는 작을것이라는 것을 보여준다. 예를 들어 집짐승의 묶음에 속하는 어느 한 다의어의 적절한 공기어는 《집짐승의 의미와 많이 관계한다》라는 조건속에 그 묶음의 여러 단어의 공기어로 될것이며 출현회수가 커진다. 반대로 적절치 못한 공기어는 그 다의

어에 국한될것이다. 결국 출현회수는 작아지며 무게값도 작다.

무게판정의 결과 매 묶음에 속하는 공기어들은 적절하든 적절치 못하든지간에 모두 무게판정값이 부여된다. 다의성처리는 이 무게판정값에 의하여 결정된다. 입력문속의 어떤 다의성을 가진 단어 W_n (복수묶음에 속하는 단어)가 나타나는 경우 그 문장의 앞 뒤 단어들 $W_{n-k}, \dots, W_{n-1}, W_{n+1}, \dots, W_{n+1}$ 을 다의어 W_n 의 매 묶음별로 따르는 공기어들과 비교하여 공기어들의 무게값의 총합을 계산한다. 이때 무게총합이 최대로 되는 묶음의 의미가 다의어 W_n 의 실질의미로 된다.

이 방법은 입력문속의 다의성만이 아니라 코퍼스 그 자체의 다의성도 해결할수 있는데 결국은 완전한 의미코퍼스는 못돼도 그 비슷한 코퍼스는 만들수 있을것이다.

이 다의성처리방법은 여러번의 실험을 통하여 완성되며 정밀해질 것이다.

말뭉치 활용 도구의 개발

이 승 재(국립국어연구원 학예연구관)

컴퓨터를 이용하여 언어 연구를 하기 시작하면서 무엇보다도 많이 만들게 된 것이 말뭉치(코퍼스)이다. 특히 다양한 용례를 필요로 하는 사전 편찬 분야에서는 말뭉치를 활용하여 용례를 추출하는 작업을 매우 중시하게 되었다. 그래서 1990년대 초부터 다양한 사전이 기획되면서 많은 양의 말뭉치도 함께 만들어지기 시작하였다¹¹⁾.

1990년대 초부터 만들어지기 시작한 한국어 말뭉치는 사전을 편찬하거나 언어 정보를 이용하려는 기관마다 개별적으로 만들었기 때문에 그 형식이 모두 다를 수밖에 없었고 그 말뭉치를 활용할 수 있는 도구(프로그램) 또한 다를 수밖에 없었다. 그러다보니 말뭉치를 활용하려고 하는 이용자는 말뭉치의 파일 형식, 문자 코드, 표지 부착 형식의 종류 등에 관한 매우 세부적인 정보들을 모두 알고 있어야만 말뭉치를 제대로 이용할 수 있게 되었다. 1998년 '21세기 세종계획'의 시작과 더불어 말뭉치 형식의 표준화가 이루어지면서 말뭉치는 각 장르를 망라하여 본격적으로 구축되기 시작하였는데 여기에서도 역시 이러한 문제들이 말끔하게 해결되지는 못 하였다.

말뭉치의 활용도를 높이려면 말뭉치를 이용하는 사람들이 손쉽게 말뭉치를 활용하여 자신의 목적에 맞는 결과를 얻을 수 있어야 한다. 그러기 위해서는 말뭉치의 세부적인 사항들을 모두 포용할 수 있는 활용 도구가 만들어져야 한다. 그러나 말뭉치를 이용자가 자유롭게 활용할 수 있는 프로그램이나 시스템은 아직까지도 많지 않은 것이 사실이다. 그래서 본고에서는 말뭉치를 보다 손쉽게 활용할 수 있는 프로그램을 만들기 위한 문제점을 살펴보고 나아가서 말뭉치를 활용할 수 있는 바람직한 프로그램

11) 이 때 말뭉치를 만들었던 대표적인 곳은 연세대, 고려대, 카이스트(kaist), 국립국어연구원이었고 이 말뭉치들의 대부분은 1998년 '21세기 세종계획'의 시작과 함께 '세종 말뭉치'로 통합되었다.

을 만들기 위한 여러 조건들을 단계적으로 알아보도록 하겠다.

1. 말뭉치(코퍼스) 활용의 문제점

일반 사용자가 말뭉치를 활용하려고 할 때 문제가 되는 것은 말뭉치 파일의 형식, 사용된 문자 코드, 사용된 부가 표지(태그)가 통일되어 있지 않다는 것이다. 효율적으로 말뭉치를 활용하는 도구를 만들려면 이러한 세 가지 요소가 유기적으로 연결되어 있어야 한다. 그러나 아직까지는 이러한 요소들이 유기적으로 연결되어 개발된 도구를 찾기 어려운 실정이다. 왜냐하면 파일 형식, 문자 코드, 부가 표지 등의 형식이 아직까지도 표준화되지 않았기 때문이다. 그러다 보니 대부분의 말뭉치 활용 도구들은 해당 프로그램에서만 지원하는 일부 제한된 말뭉치만 활용할 수 있게 되어 그 활용도가 매우 떨어지게 되었다.

말뭉치를 만들려면 한국어를 적절한 편집기에 의해 입력할 수 있어야 한다. 한국어에는 현대 한글, 옛 한글, 한자 등이 섞여 있다. 그런데 현재 우리가 사용하고 있는 문자 코드 중에는 이들 문자를 모두 지원하는 코드도 있고 부분적으로 지원하고 있는 코드도 있다. 한국어 말뭉치를 만들기 위하여 사용되었던 문자 코드를 나열하면 아래와 같다.

(1) 한국어 말뭉치 구축에 사용된 문자 코드

- ① 2바이트 완성형 한글 코드(KS X 1001)
- ② 유니코드(KS X 1005-1): 옛 한글 포함
- ③ 마이크로소프트의 통합형 한글 코드(윈도 코드)
- ④ 한글의 2바이트 조합형 한글 코드
- ⑤ 일본, 중국 등 외국의 문자 코드(JIS, GB 등): 병렬 말뭉치에서 사용

2바이트 완성형 한글 코드는 현재까지 말뭉치를 만들기 위하여 가장 많이 사용되었던 문자 코드이다. 이 코드는 KS 표준으로 되어 있었고 대부분의 프로그램에서 이 코

드를 지원했기 때문에 말뭉치나 말뭉치 활용 도구를 만들기 위하여 가장 많이 사용하게 되었던 코드이다. 그러나 이 코드는 옛 한글을 전혀 반영할 수 없다는 약점을 지니고 있어 세종 말뭉치 등에서는 한글의 2바이트 조합형 한글 코드와 유니코드를 사용하게 되었다. 그리고 한국어와 외국어를 함께 입력해야 하는 병렬 말뭉치에서는 해당 국가의 문자 코드를 사용하지 않을 수 없어 일본의 JIS 코드나 중국의 GB 코드를 사용하게 되었으나 유니코드가 점차 개선되면서 유니코드로 통합될 수 있는 가능성을 지니게 되었다. 마이크로소프트 윈도우에서 지원하고 있는 통합형 한글 코드는 표준 코드는 아니지만 윈도우 시스템에 적용되어 있기 때문에 많은 사람들이 무의식적으로 사용할 수밖에 없는 코드인데 말뭉치를 만들기 위한 바람직한 코드라고는 할 수 없다.

다음으로는 이러한 문자 코드를 사용하여 만든 파일의 형식이 문제인데 국제적인 호환성을 고려하면 텍스트 파일(아스키 파일) 형식이 가장 좋겠지만 이 또한 문자 코드와 관련이 되어 있어 여러가지 다양한 파일 형식이 사용되었다. 지금까지 말뭉치를 만들기 위하여 사용된 파일 형식을 살펴보면 아래와 같다.

(2) 한국어 말뭉치 구축에 사용된 파일 형식

- ① 텍스트 파일(아스키 파일)(*.TXT)
- ② 한글 파일(HWP)
- ③ 한글 2바이트 파일(*.2B)
- ④ 마이크로소프트 워드 파일(*.DOC)

텍스트 파일은 전 세계적으로 호환성이 보장된 파일의 형식이다. 그런데 한국어의 경우 옛 한글을 지원하는 표준적인 코드가 없는 상태여서 표준 코드를 반영한 텍스트 파일 형태로 말뭉치를 만들기가 조금 곤란한 것이 사실이다. 최근에 유니코드 영역에 옛 한글을 정의하여 사용하는 방식이 등장하여 유니코드를 반영한 텍스트 파일 형태로 말뭉치를 만들 수 있는 길이 열렸으나 이 역시 아직 표준 방식이라고 할 수 없어 불완전한 상태이다. 이러한 사정 때문에 옛 한글을 대부분 지원할 수 있는 한글의 2바이트 문자 코드를 사용하여 만들어진 말뭉치가 많은데 세종 말뭉치가 대표적인 예라

고 할 수 있다. 한글의 2바이트 문자 코드를 사용한 파일은 한글의 문서 파일 형태(hwp 파일 형태)를 사용하기도 하였지만 일부 프로그램(감작새 등)에서는 한글의 문서 형식을 포함하지 않은 2바이트 파일 형태(2b 파일 형태)를 사용하기도 하였다. 그리고 마이크로소프트의 한글 워드 파일(DOC 파일)도 일부 말뭉치에서 사용되었다.

이처럼 다양한 문자 코드와 파일 형식을 사용한 말뭉치는 부가 정보를 표시하는 방식에 있어서도 다양한 양상을 보이는데 그 내용을 정리하면 아래와 같다.

(3) 한국어 말뭉치 구축에 사용된 부가 표지(태그)의 형식

- ① 표지 없음: 단순한 텍스트 파일로만 말뭉치를 구축하고 어떠한 표지도 붙이지 않은 경우
- ② HGREP, HDB 형식: HGREP과 HDB 프로그램에서 사용할 수 있는 표지를 붙인 경우
- ③ SGML 형식: TEI 등
- ④ XML 형식

초기의 말뭉치는 아무런 표지 없이 단순 텍스트 파일로 만들어진 경우가 많았다. 그러나 곧 부가 정보의 필요성이 부각되어 서지 정보나 쪽 번호, 출전 정보 등을 간단한 기호로 표시하는 방법이 사용되었다. 국립국어연구원에서 만든 hgrep이나 hdb 프로그램에서는 서지 정보, 쪽 번호, 제목 표시 등을 비롯하여 필드 구분자까지 임의의 기호를 설정하여 표시할 수 있게 하였으며 국립국어연구원에서 구축한 대부분의 말뭉치는 이 프로그램들이 지원하는 기호를 사용하여 만들어졌다. 그 후 말뭉치 표지의 표준화 문제가 부각되면서 국제적으로 호환성을 보장하는 SGML 방식의 언어를 사용한 TEI 표지가 세종 말뭉치에서 본격적으로 사용되기 시작하였다. 최근에는 표지의 형식 자체를 XML 방식으로 바꾸어 사용하는 예도 발견되고 있다.

지금까지 살펴본 문자 코드, 파일 형식, 표지 부착 형식 등을 최근까지 만들어진 말뭉치와 비교하여 살펴보면 아래와 같다.

(4) 현행 말뭉치 구축 현황

말뭉치 파일	파일 구조	문자 코드	표지 부착 형식
연세 말뭉치	텍스트 파일	KS 코드	독립 표지
국어 정보 처리 기반 말뭉치	텍스트 파일	KS 코드	간이 SGML
국립국어연구원 말뭉치	한글 파일	한글 2바이트 코드	HGREP 프로 그램 지원 형식
세종 말뭉치 (1998년 ~ 2003년)	한글 파일	한글 2바이트 코드	SGML(TEI)
세종 말뭉치 (2004년부터)	한글 파일 텍스트 파일	유니코드 유니코드	SGML(TEI)

말뭉치를 만드는 이유는 말뭉치 안에 들어 있는 다양한 언어 정보를 추출하기 위해서이다. 그러므로 말뭉치의 언어 정보를 효과적으로 이용하기 위해서는 말뭉치를 만들 때 표준적인 형식으로 된 다양한 부가 정보를 표시해야 한다. 그래서 인문학 쪽에서는 SGML 언어를 이용하여 인문학 문서의 표준으로 고안된 TEI(Text Encoding Initiative) 언어를 사용하여 표지를 부착하는 경우가 많은데 SGML 방식이나 XML 방식의 표지를 부착한 문서는 국제적 표준을 따라가기 쉽다는 점에서 권장할 만하다. 하지만 이들 언어는 상당히 개방된 형식의 언어이기 때문에 표지가 쓰인 정확한 지침(문서 형식, DTD(data type declaration))이 없으면 이를 효율적으로 활용할 수 없다. 또한 이러한 지침에 따른 적절한 편집 도구나 검색 도구가 개발되지 않으면 실제 구축에서 활용에 이르기까지 상당한 부담을 안고 작업을 해야 하기 때문에 원문과 동일하게 입력하여 만든 말뭉치보다 활용도가 떨어질 수도 있다¹²⁾.

말뭉치에 부가 정보를 표시하는 방식은 그것을 지원할 수 있는 편집 도구의 영향을 받게 되고 이는 문서 저장 형식에도 영향을 미치게 된다. 다양한 부가 정보를 표시하는 것이 말뭉치 활용을 위해서는 좋지만 복잡한 부가 정보를 표시하려면 그것을 쉽게 편집할 수 있는 편집기가 아울러 개발되어야 한다. 그러나 SGML이나 XML 문서는 앞서도 지적하였듯이 매우 개방된 형식의 문서이기 때문에 정형화된 편집기가 존재하기 힘들다. 그래서 말뭉치를 만드는 사람들은 대부분 일반 문서 편집기를 사용하여 부가 정보를 표시하게 되는데 이 과정에서 많은 형식 오류가 발생할 수 있고 문서를 저장할 때 특정 문서 편집기의 형식을 사용하게 될 가능성이 있다.

종합하면 말뭉치를 활용할 수 있는 프로그램은 옛 한글이나 확장 한자까지 지원되는 폭넓은 문자 코드 체계를 지원해야 하고 다양한 문서 형식을 효과적으로 지원할 수 있어야 하며 말뭉치에 표시된 부가 정보를 효율적으로 활용할 수 있는 기능을 갖추어야 한다.

2. 말뭉치 활용 도구의 개발

2.1. 말뭉치 활용 도구의 요건

말뭉치 활용 도구를 개발하기 위해서는 말뭉치 활용 도구의 기본 구조를 어떻게 가져갈 것인가 하는 것을 결정해야 한다. 말뭉치 활용 도구는 PC 기반의 프로그램으로 만들 수도 있고 웹 기반의 프로그램으로 만들 수도 있다. 검색 조건이 복잡할 경우에는 PC에서 검색하는 것이 서버에 부담을 주지 않는 방법이 될 수 있으나 말뭉치의 양이 많아지게 되면 말뭉치를 PC보다 서버에 두고 관리하는 것이 훨씬 좋은 방법이 될 수 있다.

-
- 12) 실제로 XML 표지(tag)를 붙여 만든 대부분의 말뭉치들이 일부 사용자를 제외하고는 표지를 모두 제거한 후 사용한다는 점을 감안할 때 문서 형식에 대한 지침의 정확성이나 편집 및 검색 도구의 개발이 XML 문서의 활용에 있어 매우 중요하다는 것은 두말할 필요가 없을 것이다.

말뭉치 활용 도구의 첫번째 요건은 무엇보다도 그 프로그램이 지원하는 말뭉치의 문자 코드와 파일 형식, 그리고 표지 부착 형식을 유기적으로 연결하여 사용자가 별다른 조작 없이도 말뭉치를 쉽게 활용할 수 있도록 해야 한다는 것이다. 사용하고자 하는 말뭉치의 문자 코드가 프로그램에서 지원하는 것과 다르다면 이를 프로그램 내부에서 변환해줄 수 있는 모듈을 장착하여 사용자는 다양한 형식의 말뭉치 파일을 프로그램에서 선택하기만 하면 곧바로 말뭉치를 활용할 수 있도록 해야 한다. 말뭉치 활용 도구를 사용하기 위해서 부가적인 프로그램을 별도로 사용해야 한다면 이는 사용자에게 매우 부담을 주는 일일 수밖에 없다.

활용하고자 하는 말뭉치 형식이 지원되면 그 다음에는 검색을 하고자 하는 말뭉치를 어떻게 효과적으로 지정할 것인가 하는 문제가 남게 된다. 지금까지 만들어진 도구들은 대체적으로 특정 디렉터리를 선택하여 그 안에 있는 파일을 대상으로 검색을 하는 것이 일반적인 경향이었는데 장르별, 시대별로 검색을 하고자 한다면 해당 조건을 만족하는 말뭉치만을 뽑아낼 수 있는 기능이 필요하게 된다. 이미 많은 양의 말뭉치를 확보한 상태에서 검색할 대상 말뭉치를 골라야 한다면 효과적으로 대상 말뭉치만을 골라내어 집단화할 수 있는 기능이 필수적이라고 할 수 있다. 특히 말뭉치는 가공 정도에 따라 원시 말뭉치, 형태 분석 말뭉치, 구문 분석 말뭉치 등으로 나눌 수 있는데 이들 자료의 표지 부착 형식과 이들 표지를 활용할 수 있는 기능은 각각 별도로 설계해야 하므로 검색 대상 파일을 찾아내는 기능과 아울러 원시 말뭉치, 형태 분석 말뭉치, 구문 분석 말뭉치 등을 별도로 검색할 수 있는 화면을 설계해야 한다.

말뭉치 활용 도구의 성능은 무엇보다도 찾고자 하는 요소를 얼마나 정확하고 빠르게 찾아주느냐 하는 것에 달려 있다. 말뭉치 검색은 말뭉치 파일에 담겨 있는 어절 수 만큼의 문장을 만드는 일이나 마찬가지로이기 때문에 상당히 많은 양의 문장에서 검색 결과를 찾은 후 또다른 많은 양의 문장을 만들어내야 한다. 그렇기 때문에 말뭉치 검색을 할 때에는 반드시 먼저 색인을 만든 후 색인 검색을 하는 것이 좋은데 자소 검색을 하거나 검색 조건이 복잡할 때에는 이에 해당하는 색인을 만들어 놓기가 어렵기 때문에 색인을 쓰지 않는 직접 검색을 할 수밖에 없다. 그러나 직접 검색을 하게 되면 원하는 내용을 찾는 시간이 현저하게 느려진다. 따라서 말뭉치 활용 도구를 설계할 때에

는 단순 검색과 상세 검색 메뉴를 분리하여 색인 검색과 직접 검색을 적절히 조합해야 한다.

말뭉치 활용 도구는 말뭉치를 활용하여 사용자가 얻고자 하는 정보를 추출하기 위한 도구이다. 그 중에서 가장 많이 쓰는 기능이 말뭉치에서 용례를 추출하는 기능이다. 용례를 추출할 때에는 추출한 용례를 어떤 방식으로 사용자에게 보여주느냐 하는 것이 중요한데 보통은 검색 표제어가 해당 용례 문장 안에만 들어가 있는 KWIC(keyword in context) 방식과 검색 표제어가 별도로 밖으로 나와 있는 KWOC(keyword out of context) 방식으로 검색된 결과를 보여 주는 경우가 많다. 그래서 검색 결과는 KIWC 방식과 KWOC 방식으로 모두 출력할 수 있어야 한다. 그리고 검색 표제어의 성격에 따라 검색 표제어의 앞 부분을 많이 참조해야 할 경우도 있고 검색 표제어의 뒷 부분을 많이 참조해야 할 경우도 있기 때문에 용례 문장 안에서의 검색 표제어의 위치 및 좌우 출력 어절 수를 조정할 수 있는 기능이 있어야 한다.

말뭉치 활용 도구에서 가장 중요한 것은 검색 기능에 관련된 것인데 여기에는 몇 가지 고려해야 할 사항이 있다.

(5) 말뭉치 활용 도구의 검색 기능

- ① 검색 단위의 설정: (연접된) 자소 단위, 음절 단위, 어절 단위, 분석 형태(태그) 단위
- ② 검색 조건: 사직연산(불린) 검색, 정규식 검색, 특수 기호 검색 지원
- ③ 용례 연산(재검색) 기능 지원
- ④ 검색 조건식의 저장 기능
- ⑤ 통계 기능: 자소별, 음절별, 어절별, 분석 태그별(품사 등) 통계 정보 제공

말뭉치 활용 도구에서 검색할 수 있는 검색 단위는 매우 중요하다. 국어의 음운 규칙이나 기타 현상을 관찰하고자 할 때에는 연접되는 자소들의 연쇄를 검색해야 할 경우가 많은데 이러한 때에는 자소 단위로 검색 조건을 주어야 한다. 그런데 완성형 기반의 문자 코드를 사용하는 유니코드나 KS 완성형 한글 코드로는 이러한 기능을 구현

하기가 매우 힘들다. 그러나 자소 검색 기능은 반드시 필요한 기능 중의 하나이다. 그 이외에 음절, 어절, 분석 형태 단위의 검색은 별 무리없이 진행될 수 있는 요소들이다. 검색 조건은 일반적으로 지원하는 사칙연산 검색과 찾고자 하는 형태를 정규식의 형태로 입력받아서 이를 검색할 수 있는 기능, 그리고 특수한 의미를 가진 특수 기호 등을 제외하고 검색하는 기능들을 구현할 수 있어야 한다. 그리고 사용자가 편리하게 이용하기 위해서는 한 번 검색한 조건을 저장해 두었다가 비슷한 유형의 검색이 있을 때에 이를 다시 불러내어 쓸 수 있는 기능이 필요하다. 통계 정보는 자소별, 음절별, 어절별 통계를 내기 위하여 필요하고 특히 분석 태그가 어휘 단위로 되어 있을 때에는 태그만을 분석하여 어휘 빈도 정보도 낼 수 있도록 설계되어야 한다.

지금까지 살펴본 말뭉치 활용 도구가 갖추어야 할 조건들을 다시 한 번 세부적으로 살펴보면 아래와 같다.

(6) 말뭉치 활용 시스템의 구성 요건

- ① 유니코드의 사용: 표준적이고 호환성이 보장되는 시스템을 만들려면 유니코드를 사용해야 하는데 유니코드를 사용한다 하더라도 실제 시스템을 만들 때에는 이를 어떤 방식으로 인코딩해서 사용할 것인가를 결정해야 한다. 현재 데이터베이스 시스템이나 검색 엔진 등에서 가장 폭넓게 지원하는 인코딩 방식은 UTF-8 인코딩 방식¹³⁾인데 한글이나 한자를 위해서는 유니코드를 모두 16비트로 인코딩하는 UTF-16 방식이 좋으나 유닉스 운영체제나 기타 응용 프로그램에서 쓸 수 없는 경우가 많아 UTF-32 인코딩 방식을 사용하지 않는 이상 UTF-8 인코딩 방식을 사용할 수밖에 없다¹⁴⁾.

13) UTF-8 인코딩 방식은 아스키 문자는 그대로 1바이트를 사용하면서 한글이나 한자 등을 사용할 때에는 3바이트를 사용하는 방식으로 이 방식을 사용하게 되면 한글이나 한자 자료를 처리할 때 저장 용량이 늘어나게 되고 색인을 만들게 되어도 색인의 크기가 커지는 단점이 있다. 하지만 대부분의 외국계 프로그램이 UTF-8만을 지원하는 경우가 많고 유닉스 운영 체제에서는 UTF-16을 사용할 수 없는 문제가 있어 현재는 UTF-8이 가장 범용적인 유니코드 인코딩 방식으로 자리잡고 있다.

14) 아직까지는 유니코드의 옛 한글 규격 및 확장 한자의 규격이 고정되지 않은 상태이고 유니코드의 한글 처리 방식이 완성형을 기본으로 하고 있기 때문에 유니코드로 도구를 만든다

② 말뭉치의 등록

- 말뭉치 파일을 등록할 때에는 말뭉치를 세밀하게 분류하여 등록해야 원하는 용도로 말뭉치를 집단화하여 검색할 수 있게 된다. 개인용 컴퓨터에서 실행되는 프로그램으로 말뭉치 활용 시스템을 만들었을 경우 사용자가 검색하고자 하는 대상 파일들을 쉽게 조직화할 수 있으려면 사용자가 특정 디렉터리를 만들어 그곳에 검색하고자 하는 말뭉치 파일을 모아 놓고 검색하도록 하면 된다. 그러나 웹 기반의 검색 시스템을 만들게 되면 분야별, 시대별, 파일별, 등록날짜별로 사용자가 검색하고자 하는 파일만 쉽게 모아서 검색 대상을 확정할 수 있도록 해야 한다. 그렇지 않으면 사용자는 매번 모든 말뭉치를 대상으로 검색을 해야 하는 수고를 감수해야 한다.

- 말뭉치 파일의 관리를 위해서는 말뭉치에 입력된 다양한 머리말(헤더) 정보를 정확하게 반영할 수 있는 체계가 만들어져야 한다. 특히 세종 말뭉치와 같이 TEI 표지가 붙어 있는 말뭉치를 등록할 경우에는 TEI 표지의 정보를 적극적으로 반영할 수 있는 데이터베이스 구조를 만들어 말뭉치를 등록해야 한다.

③ 등록된 말뭉치 목록의 조회: 등록된 말뭉치 파일의 목록과 서지 정보를 다양한 분류 기준에 의하여 조회할 수 있어야 한다. 등록된 목록이 효과적으로 검색되지 않으면 등록된 말뭉치의 현황을 잘 알 수 없을 뿐만 아니라 파일을 체계적으로 관리하기도 매우 불편하다.

④ 검색 방법의 지정: 글잡이 프로그램이 직접 탐색 프로그램과 색인 탐색 프로그램으로 나뉘어져 있듯이 말뭉치 활용 시스템도 두 부분으로 나뉘어질 필요가 있다. 직접 탐색을 하게 되면 색인을 사용하지 않기 때문에 검색 속도는 매우 느려질 수 있지만 검색 조건을 매우 다양하게 줄 수가 있어 제한된 말뭉치의 검색에는 색인 탐색 방법보다 더 나은 결과를 가져올 수

하더라도 HGREP이나 캄썹새 같이 자유로운 자소 검색을 구현하기가 쉽지 않다. HGREP 프로그램의 경우에도 일부 완성형 옛 한글은 자소 검색 기능을 사용할 수 없게 되어 있다.

있다. 색인을 만들게 되면 색인을 만들 때 불필요한 숫자, 기호 등을 제거하기 때문에 숫자나 특정 기호를 이용한 검색을 할 수 없게 되는 경우가 많다. 그래서 정밀한 검색을 위해서는 시간이 걸리더라도 색인을 이용하지 않는 직접 탐색 방법을 이용한다. 다만 직접 탐색 방법은 서버에 무리한 부담을 줄 가능성이 많으므로 말뭉치 활용 도구를 PC용 프로그램으로 만들거나 서버용 프로그램으로 만들더라도 사용자 수나 검색 대상 파일을 적절히 제한하는 방법을 사용해야 한다.

⑤ 검색 조건 및 출력 형식의 설정

- 한국어는 교착어이기 때문에 형태소에 관련된 검색이 매우 중요하고 형태소는 음절 단위로 기능하기도 하지만 자소 단위로 기능할 때도 많기 때문에 초성, 중성, 종성을 각각 지정하여 검색할 수 있는 자소 단위 검색 기능이 필요하다. 그리고 사칙 연산 기능(불린)을 이용한 검색도 가능해야 한다. 특히 부정 조건을 주어 '중성이 'ㄱ'으로 끝나지 않으면서 다음 단어의 초성이 'ㄱ'으로 시작하는 단어'와 같은 조건을 줄 수 있는 기능이 필요하다. 그리고 이를 위하여 초성, 중성, 종성에 무슨 글자가 오든지 상관없다는 기호도 하나 필요하다.
- 특수 기호를 일반 기호로 검색할 수 있어야 한다. 말뭉치에 부착한 표지 중에서 표지의 시작을 알리는 '<'와 같은 기호는 상당히 많이 나타나게 되는데 이러한 기호들은 특수한 의미를 가지기 때문에 검색 프로그램에서 일반 기호로 처리하지 않고 특수한 의미를 가지는 기호로 처리하게 된다. 그런데 이러한 기호는 표지로 나타날 수도 있지만 본문 중에서 일반 기호로 사용될 수도 있다. 그러나 검색 조건에서 '<'를 찾게 되면 이 기호는 특수한 의미를 갖는 기호이기 때문에 찾지 않고 그냥 넘어가게 된다. 따라서 특수한 의미를 가지는 기호를 특수한 의미가 아니라 일반적인 의미의 기호로 검색을 하려면 유닉스의 '\ '와 같이 어떤 기호를 일반적인 입력 기호로 인식하여 검색하라는 조건을 줄 수 있는 기호가 필요하다. HGREP 프로그램의 경우 '&, *, ?' 등은 각각 페이지 연결, 아무

문자, 한 문자 등을 의미하는 기호들인데 본문에 나타난 '?'기호를 찾으려면 검색 조건에서 '\?'와 같이 표시해 주어야 본문에 쓰인 '?' 기호를 찾을 수 있다.

- 검색 조건과 출력 형식은 많이 사용하는 형식을 3~4개 정도 골라 표준 형식으로 만들어 놓고 대부분의 경우 이 형식을 선택한 후 검색하고자 하는 문자열만을 바꾸어 검색할 수 있는 방법을 마련하면 사용자가 비교적 부담을 가지지 않고 말뭉치를 검색할 수 있다. 그러나 검색 조건을 다양하게 만들고자 하는 사용자는 사용자 정의 조건을 이용하여 복잡한 검색 조건식을 만들 수 있고 이것을 저장해 두었다가 나중에 비슷한 유형의 검색 조건을 만들 때 사용할 수 있도록 하면 다양한 검색 조건식을 모두 지원할 수 있게 된다. 또한 검색된 용례를 출력할 때에는 검색 포제어를 중심으로 하여 KWIC이나 KWOC 방식으로 결과를 출력할 수 있어야 한다.
- 검색된 용례 정보에 원전의 위치를 알 수 있는 정보가 있거나 원전의 해당 부분을 보여줄 수 있어야 한다. 사전 편찬을 위한 용례로 사용할 경우 원전 대조 작업은 필수적인 작업인데 말뭉치 파일에 원전의 페이지를 나타내는 표시가 없으면 그 용례를 나중에 원전과 대조할 수 없어 앞뒤 문맥을 참고할 수 없게 된다. HGREP 프로그램의 경우에는 원시 말뭉치에 '<파일 이름, 페이지>'와 같은 표지를 넣어 이 문제를 해결하고 있고 글잡이 프로그램에서는 검색된 용례를 마우스로 누르면 화면 아래에 해당 용례가 입력되어 있는 원전 파일이 나타나는 방식으로 이 문제를 해결하고 있다. 그런데 세종 말뭉치 중 대부분의 파일에는 페이지를 나타내는 <pb n=...> 등의 표지가 붙어 있지 않은 경우가 많아 원전 확인을 어렵게 하고 있다.

⑥ 검색 결과

- 검색된 결과는 이 안에서 다시 재검색이 가능해야 하고 사용자의 용도에 맞게 재가공이 가능해야 한다. 검색된 용례를 별도의 파일로 저장한다거

- 나 검색된 상태에서 검색 결과를 효과적으로 줄일 수 있는 기능이 있어야 한다. 검색된 용례를 화면상에서만 보고 말 것이 아니라면 반드시 별도의 파일로 저장해야 하고 화면상에서 불필요한 부분을 없애고 결과의 양을 줄여서 저장하기 위해서는 어느 정도의 편집도 가능해야 한다.
- 검색된 결과는 사용자가 보기 편하도록 효과적으로 정렬될 수 있어야 하고 한국어의 특성을 반영한 순방향 정렬과 역방향 정렬도 자소 단위, 음절 단위로 이루어질 수 있어야 한다.

2.2. 말뭉치 활용 도구의 종류 및 현황

말뭉치를 활용하는 도구, 특히 용례를 추출하는 도구는 직접 검색인지 색인 검색인지, 그리고 어떤 종류의 말뭉치를 대상으로 하는지에 따라 다음의 몇 가지 유형으로 나눌 수 있다.

(7) 말뭉치 활용 도구의 종류

- ① 원시 말뭉치에서 특정 문자열을 단순 검색하는 도구: HGREP, 글잡이(직접), 감작새
- ② 원시 말뭉치를 입력으로 받아들여 이를 형태소 분석한 후 색인을 만들고 여기에서 특정 문자열을 검색하는 도구: 글잡이(색인)
- ③ 분석 말뭉치에서 특정 문자열을 검색하는 도구: 말쑈, 글잡이(색인)

지금까지 개발된 도구는 대부분 직접 검색을 위주로 한 도구였다. 그러나 말뭉치의 양이 많아지면서 글잡이를 시작으로 하여 많은 프로그램들이 색인을 이용한 빠른 검색 기능을 구현하기 시작하였다. 지금까지 개발된 여러 프로그램들의 장단점을 살펴보면 아래와 같다.

(8) 말뭉치 활용 도구의 현황

도구	장 점	단 점
HGREP.EXE	- 한글에서 지원하는 모든 문자 사용 가능 - 검색 조건이 다양	- 사용하기가 어려움 - 속도가 느림 - 한글 파일만을 지원
말씀	- 연세 말뭉치(형태소 분석 말뭉치) 검색 가능	- 외부에서 사용 못 함 - 연세 말뭉치만을 지원
글잡이(직접)	- 검색된 용례의 원문을 자유롭게 볼 수 있음	- KS 한글 코드만을 지원
글잡이(색인)	- 원시 말뭉치를 형태 분석하여 검색하는 것이 가능 - 검색 속도가 빠름	- 색인의 크기가 크고 색인을 만드는 시간이 오래 걸림 - KS 한글 코드만을 지원
깜짝새	- 옛 한글까지 검색 가능 - 다양한 자소 검색 가능	- 한글 파일만을 지원 - 용량에 따라 검색 속도가 느림

위의 표에서 보듯이 지금까지 개발된 많은 도구들은 나름대로의 장단점을 모두 가지고 있는 프로그램들이다. 그러나 이들 프로그램들이 지원하는 말뭉치의 형식은 프로그램마다 조금씩 다르다.

(9) 말뭉치와 말뭉치 활용 도구

문자 코드	글자 수	KS 규격	국제 규격	적용 시스템	관련 도구
2바이트 완성형	한글 2,350 한자 4,888	KS C 5601-1987 ⇒ KS X 1001	ISO-2022	한글 MS-DOS , 윈도 95	
2바이트 완성형	한글 2,350 한자 4,888	KS C 5601-1992 ⇒ KS X 1001, 1003	ISO-2022	윈도 95, 윈도 98에 서 변형하여 적용	말뭉, 글잡이
2바이트 조합형	한글 11,172 한자 4,888	KS C 5601-1992	×	KS 규격 제정 이 전에 대부분 사용	글 잡 이 에 서 내 부 적 으 로 사용
한글 2바이트 조합형	65,536 영역 에서 자유 롭게 사용	×	×	한글에서 독자적으 로 사용	HGREP.EXE, 세종 말뭉치
유니코드	한글 11,172자 포함	KS C 5700 ⇒ KS X 1005	ISO 10646-1	윈도 98에서 부분 적으로 지원, 윈도 2000 이상에서 지 원	한일 병렬 말 뭉치
유니코드 + 확장 옛 한글 (사용자 정의 방식)		KS C 5700 ⇒ KS X 1005	ISO 10646-1	MS-WORD 2000에서 지원	깜짝새, 세종 말뭉치
유니코드 + 확장 옛 한글 (조합 방식)		KS C 5700 ⇒ KS X 1005	ISO 10646-1	MS-WORD 2002에서 지원	

이처럼 다양한 형식으로 만들어져 있는 말뭉치와 말뭉치 지원 도구를 사용자 입장에서 간단한 도구로 정리하려면 말뭉치를 일정한 문자 코드를 사용하는 형식으로 만들어 중앙 컴퓨터에 저장하고 이를 자유롭게 검색할 수 있는 시스템을 만드는 것이 좋을 것이다.

한국어 말뭉치를 구축하고 그것을 활용할 수 있는 도구를 만들기 위해서는 위 (6)에

서 살펴본 여섯 가지의 기본 사항을 모두 갖추고 있어야 한다. 말뭉치는 언어 정보를 효율적으로 다루기 위하여 만드는 것이지만 그것을 최대한 활용할 수 있는 도구가 만들어지지 않는다면 사상누각에 불과할 수 있다. 따라서 말뭉치 활용 도구는 말뭉치에 부착되어 있는 정보를 최대한 활용할 수 있어야 하고 말뭉치를 활용하여 할 수 있는 다른 작업이나 다른 프로그램(어휘 분석 도구, 사전 편찬 도구, 통계 처리 도구 등)과 연동될 수 있어야 한다. 그런가 하면 사용자가 손쉽게 말뭉치를 활용할 수 있도록 말뭉치의 다양한 파일 형식을 지원해야 하고 다양하게 검색 대상 말뭉치를 지정할 수 있어야 하며 검색 조건도 다양하게 지정할 수 있어야 한다. 그리하여 말뭉치 활용 도구는 말뭉치를 활용한 국어 정보 은행을 구축할 수 있는 도구로 발전해야 할 것이다.

〈참고 문헌〉

- 장범모(2002), 「텍스트 부호화와 문서 표준: TEI/SGML」, 『한국어와 정보화』, 태학사
- 국립국어연구원(1995), 『한글 코드에 관한 연구』(연구 책임자: 홍윤표)
- 김홍규(1999), 「21세기의 정보 환경과 한글 옛글자」, 21세기의 정보 환경과 한글 옛글자 심포지움 발표 자료
- 문영호(1996), 「국제표준글자부호계에 등록할 조선글자체계의 자모순서문제에 대하여」, '96 Korean 컴퓨터처리 국제학술대회
- 문화관광부(2000), 21세기 세종계획 국어 기초 자료 구축 분과 보고서 - 용례 추출 도구의 개발(연구 책임자: 임해창)
- 서상규(2002), 「한국어 말뭉치의 구축과 과제」, 『한국어와 정보화』, 태학사
- 성석수(1999), 「Microsoft Word 2000에서의 한글 구현」, 21세기의 정보 환경과 한글 옛글자 심포지움 발표 자료
- 소강춘·김진규(2001), 「통합형 한글 자료 처리기(SynKDP)에 대하여」, 국어사 자료 연구 제2호, 국어사자료학회
- 신병훈·김성재·안상규(2002), 「마이크로소프트 Word 2002에서의 옛한글 구현」, 『국어 연구 자료 구축 1』, 국립국어연구원

- 연규동 · 박진호 · 최운호(2003), 『인문학을 위한 컴퓨터』, 태학사
- 이승재(1999), 「정보화 시대의 국어 생활」, 국어문화학교, 국립국어연구원
- 이승재(2002), 「한국어와 정보 처리 환경」, 『한국어와 정보화』, 태학사
- 이승재(2003), 「한국어 정보 자료의 구축과 활용 방안 연구」, 서강대학교 박사학위 논문
- 정내권(1999), 「유니코드 확장 방안」, 21세기의 정보 환경과 한글 옛글자 심포지움 발표 자료
- 최병수(1996), 「컴퓨터 부호 체계에 넣어야 할 옛글자에 대한 연구」, '96 Korean 컴퓨터처리 국제학술대회 논문집
- 홍윤표 · 이기문 · 김충희 · 김흥규 · 박성종 · 정인상 · 김병선 · 김홍배 · 이민재(1991), 「한글 옛글자의 컴퓨터 처리 방안에 대한 연구」, 한국어 전산학 창간호, 한국어전산학회
- Kobayashi, Tatsuo(2001), "Universal Character Set" - A infrastructure for languages in cyberspace, International Symposium on "Language in Cyberspace", Korean National Commission for UNESCO
- Sinclair, J. M.(1991), Corpus, Concordance, Collocation, Oxford, Oxford University Press
- The Unicode Consortium(2000), The Unicode Standard, Version 3.0, Addison Wesley Longman

2004년 학술 회의 관련 사진





학술 회의 참가자 전원(호텔 앞에서)



만찬 때의 모습(왼쪽부터 전하선 교수 남기신 원장 무여송 소장)



인사말(남기심 원장)



인사말(최응구 교수)



기조 연설(문영호 소장)



기조 연설(남기심 원장)



기조 연설(전학석 교수)



말뭉치 분과의 발표 모습



방언 분과의 발표 모습



발표장의 북쪽 참가자들



발표장의 남쪽 참가자들



학술 회의 참가자 전원(발표장에서)



방언 분과의 토론 모습





◇ 담당 연구원 : 이 승 재 (李承宰)
(이문규범연구부 학예연구관)

국립국어연구원 2004-3-8

제3차 남북 국제 학술 회의 논문집

2004년 8월 29일 인쇄

2004년 9월 1일 발행

발 행 인 남기심

발 행 처 국립국어연구원

서울특별시 강서구 방화3동 827번지

전 화 (02) 2669-9712

전 송 (02) 2669-9777

홈페이지 <http://www.korean.go.kr>

전자우편 kacademy@korean.go.kr

인 쇄 크리홍보 주식회사

(전화: 02-737-5377)
