

우리글 정보처리 남북한 공동 호흡

민족어 통일과 언어정보산업 표준에 관한 학술 모임 개최

남북이 분단되기 전까지 우리나라에는 훈민정음에 바탕을 둔 우리글이 한 가지뿐이었다. 그러나 지금은 우리글에 대한 명칭조차 다르다. 남한에서는 '한글'이 우리글이지만 북한에서는 '조선어'로 통용된다. 자음의 배열도 다르다. 북한에서는 'ㄷ' 다음에 'ㅈ'이, 'ㅎ' 뒤에는 'ㅇ'이 위치한다. 이러한 차이는 컴퓨터가 등장하면서 더욱 심각해졌다. 즉 각 글자의 컴퓨터 코드가 다르고 자모순이 다르기 때문에 남북한에서 각각 개발한 소프트웨어의 호환성이 없다. 또한 자판의 배열, 컴퓨터용어의 차이도 남북의 정보교류협력에 걸림돌이 되고 있다.



Korean 정보처리 국제학술대회에는 언어학 분야에서 국내외적으로 명성을 떨치고 있는 김일성종합대학의 김영환 박사(사진 가운데)가 참석했다.

언어 정보 처리를 인민경제 초석으로 인식

이러한 문제를 해소하기 위해 1994년부터 5차례에 걸쳐 중국 연변에서는 'Korean 정보처리 국제학술대회'가 열렸다. 이 학술대회에는 남북한의 학자와 중국을 비롯한 해외동포 학자들이 참석하여 용어, 자판, 자모순, 코드의 통일을 위한 연구를 진행했다. 그 결과 자판과 자모순에 공동안을 제시했으며 1999년에는 <국제표준정보기술용어사전>을 2002년에는 이의 증보판인 <ISO2382 기준 한·영·조·중·일 정보기술표준용어사전>을 남북한과 중국이 공동으로 발간하였다.

또한 2002년 8월에는 좀더 나아가 우리언어 정보처리를 산업화하여 경제활성화에 기여하고자 '정보화 시대에 따르는 민족어의 통일적 발전과 언어정보산업 표준에 관한 학술모임'을 중국 베이징에서 개최하였다. 이 학술모임에서는 민족어의 정보처리

를 위한 표준화 문제, 전문용어 처리문제, 자료교환을 위한 문자코드 처리 문제, 효율적인 자판 사용을 위한 문제, 문장 분석기의 성능을 높이기 위한 코퍼스 및 고급언어자원 구축문제에 대한 발표와 토론이 활발하게 전개되었다.

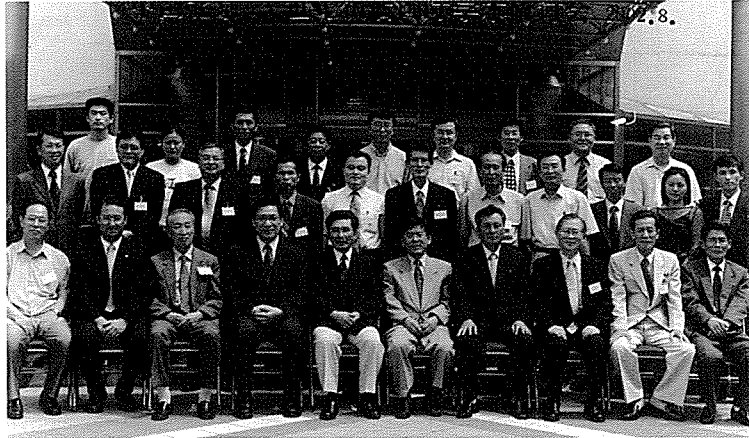
북한에서는 우리글 정보처리에 대한 관심이 오래 전부터 지대하였으며 많은 연구와 함께 저서도 나오

고 있다. 최근에는 김일성종합 대학 언어정보화 연구집단에서 <정보시대와 언어학>이라는 잡지도 발간하고 있다. 북한은 언어 정보처리를 인민경제 모든 부문의 생산기술 공정과 생산방법, 경영활동을 새로운 과학적 토대 위에 올려놓기 위한 중요한 연구사업으로 보고 있다.

그 이유는 과거에는 생산공정의 로봇화, 전자계산기화가 실현됨에 따라 컴퓨터가 수학적 계산이나 판단의 도구로만 쓰였으나 앞으로는 점차 인간의 지식을 대신하여 주는 지적기계로 발전해 나가기 때문이라 한다. 즉 컴퓨터가 인간이 진행하는 것과 같은 지적활동을 모방하려면 자연언어를 이해하고 합성할 수 있어야 한다는 것이다.

우리글 자유롭게 구사하는 능력 컴퓨터에 부여

조선어 정보처리의 목표는 정보화의 시대적 흐름을 반영하여 조선어를 자유로이 쓸 수 있는 능력을 컴퓨터에 부여하는데 있다. 여기서 정보처리는 구체적으로 정보의 자동처리를 말하며 정보 전달 수단으로서의 글자와 정보자료로서의 언어자료(본문)를



지난 2002년 8월 중국 베이징에서 열린 언어정보산업 표준에 관한 학술 모임 참석자들.

컴퓨터가 읽고 이해한 다음 여러 가지 과업 즉 번역, 초록, 색인, 검색, 해독을 하는데 맞게 가공하는 것을 말한다. 오늘날 정보의 자동처리는 이론적 측면에서는 정보공학의 중심을 이루며 실제적 측면에서는 정보산업의 중심을 이룬다고 보고 있다. 이러한 특성으로 인하여 응용언어학으로서의 정보처리는 글자와 언어 자료를 자동 처리하는 데에서 일어나는 언어학적 문제를 기본 연구대상으로 하면서도 수학, 전자공학, 기호학, 서지학, 번역이론 등과 밀접한 연관을 맺고 있다.

조선어 정보처리는 글자와 언어자료의 측면에서 조선어를 기본자료로 취급하고 있으며 컴퓨터와 관련하여 조선글자와 조선어 자료의 입·출력, 조선어 분석종합의 자동화 원리와 방법, 그리고 번역, 검색, 질의응답 등을 주된 연구대상으로 하고 있다. 이러한 연구는 우리글자, 단어, 문장을 읽고 이해하고 번역, 검색하는 언어학적 문제를 해결하는 것으로써 컴퓨터로 하여금 조선어를 자유로이 쓸 수 있는 능력을 가지게 하는 것과 관련되어 있어 매우 어렵고 복잡하다.

〈정보시대와 언어학〉에 실린 주목할만한 논문들

김일성종합대학의 〈정보시대와 언어학〉 제 1권은 2001년 8월에 발간된 것으로 '정보처리를 위한 탐구 : 언어학의 새로운 전개'라는 특집으로 엮여져 있다. 이 잡지의 집필자는 모두 14명으로 주로 김일성종합대학 문학대학 교수들이지만 외국어문학부, 사회과학센터 및 컴퓨터과학대학 교수도 포함되어 있다.

특히 이 책의 필자 가운데 김일성종합대학 문학대학 언어학 강좌 교수인 김영학 박사는 조선과학원 후보원사로서 언어학 분야에서는 국내외적으로 명성을 떨치고 있다. 김 박사와의 2001년 2월 연걸에서 열렸던 'Korean 정보처리 국제학술회의'에서

처음 만났는데, 우리말의 로마자 표기에 대한 학식이 풍부할 뿐 아니라 우리글의 컴퓨터처리에 대해서도 일가견이 있었다. 김 교수가 〈정보시대와 언어학〉에 게재한 글 가운데에는 '우리말의 로마자 표기법 문제'라는 논문 외에 '디지털 활자와 정음자 코드', '유니코드(국제문자 부호계)', '한자의 입력방식', '음성자동통역기술' 등 컴퓨터와 관련된 글이 눈길을 끈다.

〈정보시대와 언어학〉 1권에 게재된 다른 중요한 논문으로는 '정보시대와 계산언어학의 최근 발전동향(최정호)', '정보시대와 인지언어학의 역할(박영균)', '컴퓨터에 의한 어휘생성의 자동처리 문제(배광희)', '조선부 엔트로피에 의한 본문 정보측정의 한가지 방법(안성득)', '영·조 기계번역에서 분석과정과 분석결과의 기호적 표기문제(김호룡)' 등을 들 수 있다. 이밖에도 북한에는 우리글 정보처리 분야의 학자가 여러 명 있다.

조선컴퓨터센터의 기술처장으로 있다가 지금은 교육성 산하 프로그램교육센터 소장으로서 있는 리수락 박사는 컴퓨터에 의한 조선어 처리기술에서 제기되는 여러 가지 과학기술적 문제의 해결과 표준화 방면에서 많은 연구를 하였으며, 조선 사회과학원 언어학연구소장인 문영호 박사는 컴퓨터처리를 위한 조선음절자 구성과 자모순 문제에 대한 연구, 평양정보센터의 최경호 종합실장은 조선글 전자출판 체계 수립에서 제기되는 문제 연구, 그리고 사회과학원 언어연구소의 최병수 연구사는 조선문자와 기계번역기술에 대한 연구를 하였다.

'지리적 통일에 앞서 겨레의 언어동질성 회복을 통하여 겨레의 얼과 전통문화가 정보통신 기술을 활용하여 자자손손 발전 번영하게 한다'는 '남북언어정보산업표준위원회'의 모토와 같이 남북이 언어의 동질성을 하루 속히 회복하게 되기를 기대해 본다. ☞

글 박찬모 | 포항공대 대학원장