

국어정보화에 관한 남북 교류 전망

진용옥(경희대학교수:정보통신공학전공)

1. 서

남북한 간에는 통행, 통신, 통상등 여러가지 장벽이 있지만 그 중에서도 통신교류는 가장 시급하고 절실한 과제중의 하나이다. 통신교류에는 우편과 전신전화가 있고 넓게는 방송과 전산분야도 포함된다고 보겠다. 전자가 직접적 교류라면 후자는 간접적 교류에 해당한다.

우편과 전신전화의 경우 분단초기에서 48년까지 지속하다가 중단된 이래 오늘에 이르고 있으며 방송은 상호비방의 수단으로 전락하고 말았다. 이에 비해 전산분야는 최근에 대두된 문제이다. 그러나 통일이 정보시대에 이루어질 것으로 가정했을때 체제를 달리 할지라도 통일이전 이나 그 이후에도 동일한 체계에서 이루어지는 것이 소망스러운 일이다. 특히 전산이용은 한글과 한말(한국어)과도 밀접한 관계가 있으므로 언어문제와도 직결되어 동질성 회복에 있어 중요한 수단이므로 쉽사리 합의가 어려운 점은 있으나 체제변화에 대한 효과가 직접적이지 않은 관계로 의외로 양방간의 합의가 비교적 용이한 분야이기도 하다.

2. 국어정보화에 관한 남북교류 진행

최근 국어정보학회에서는 남북한 및 국외 동포들이 모인 전산 관계 학술회의를 공동으로 주최한바 있다. 1994년 8월 5~9일까지 중국연변에서 열렸으며 북한전문가는 25명이 참석하고 11편의 논문이 발표되었다. 김일성 사후, 그것도 대규모 북한학자들이 나온 것은 극히 이례적인 현상이며 어떤 면에서는 분단이후 대규모 학술교류의 하나로 자평할 수 있을 것이다.

이같은 이 대회에서 논의된 내용을 중심으로 남북한의 차이와 학술교류전망에 대해서 북한측 발표논문을 중심으로 기술한다.

3. 국어 정보분야의 남북문제

국어정보화란 한글과 한말(한국어)을 정보통신 방법으로 실현하는 것을 말한다. 이에는 여러가지 분야가 있으나 입력장치로써 글씨판(자판)문제, 정보처리와 교환용 부호화 체계 그리고 출력방식으로 글꼴(폰트)문제가 있다. 그 이외에 한글자료기지(DB)와 글틀(워드프로세서) 그리고 문자인식 등 인공지능기술을 활용하는 분야가 있다. 여기에서는 주로 위 3가지 분야 즉 글씨자판, 부호 및 글꼴 문제에 대해 집중적으로 검토한다. 이들 3분야는 국어정보화의 핵심과제이며 국내에서조차 의견의 일치를 보지 못하는 분야도 많다.

예컨대 글씨자판에서 3벌식과 2벌식, 부호체계에서 통합바이트문제나, 완성형과 조합형이 그것이다. 북한 역시 마찬가지이며 우리의 초기적 고민(5~7년전)과 비슷한 양상인것 같으며 이수락¹⁾의 논문에서 개괄적으로 잘 지적하고 있다. 이수락은 이들 문제를 해결하는 원칙에 대해서도 언급하고 있는데 눈길을 끄는 대목은 과학적 공정성과 주체적 입장을 강조하고 있다는 점이다. 주체적 입장이란 민족적 이해와 독창성 그리고 자력해결의 원칙을 강조한 것으로 원론적 입장에 지나지 않는다. 또한 국제협력을 강조하고 있으면서 동시에 자력해결을 주장하고 있다. 그러나 자력해결이 부족하거나 미진할 때 정체가 불가피하며 나아가 대외의존을 촉진시키는 점을 잊지 말아야 한다.

4. 분야별 내용과 전망

(1) 소리마디글자와 자모순 문제에 대하여

문영호²⁾의 논문에서는 소리마디글자(음절)와 자모의 배열 순서에 대하여 논급하고 있다. 그러나 자모순 배열에서 우리와는 순서가 다르다. 이 논문에서는 소리마디글자(음절문자)는 언어학적, 경험적, 통계적으로 취급할 수 있는 총수 4160자를 제시하였다. 남북의 언어와 문자가 이질화 되고 있는 것은 사실이지만 전체적 시각으로 보아 본질적 변화는 아니라고 본다. 그러나 언어학적 입장과는 달리 전산처리에서 하나라도 다르면 공통사용이 불가능 함으로 완전 일치에 이르려면 상당한 노력과 시일이 들것으로 전망된다.

1) 북한의 전자 자동화 공업위원회 컴퓨터 센터 처장

2) 사회과학원 언어연구소 실장

(2) 정보교환 부호에 대하여

김명규³⁾의 논문은 한글 부호화의 표준안에 대하여 논술하고 있다. 문영호의 논문에서 제시된 것처럼 7비트 체계는 본질적으로 남측과 대동소이하다. 다만 자모 배열이 다르므로 배열위치가 다르고 쌍자음과 겹모음의 배열순이 다르다. 2바이트 부호체계에서는 자모수에서 남한 94자인데 북한은 80자로 제한하였다. 이는 옛글자를 취사선택 여부에서 오는 차이이다. 또한 한자는 제외된 점이 두드러진다. 한국에서 완성형 글자수는 4280으로 확장되었는데 북한에서는 2350자에서 4136자인 것 같다. 한자, 라틴어, 일본어 문자세트가 제외되어 있는 것이 눈에 띄는 차이이다. 따라서 한자영역이 제외되어 남은 구역이 많아졌지만 어디에 사용할 것인가에 대한 구체적 대안에는 언급이 없다.

(3) 자판배치

권중성⁴⁾의 논문은 입력코드수와 자판배치기술에 관한 것으로 이번 학술발표에 가장 주목되는 부분이다. 자판의 통일은 모든 전산처리 분야의 상징적 분야인데 결론적으로 말한다면 남북 양측의 성실한 노력 여하에 따라 통일자판에 대한 구체안이 마련될 것 같은 전망이다. 지금까지 북한에서는 26, 41, 51 타건 방식이 있었던 것으로 알려지고 있다. 26 타건은 1991년 3월 ISO에 제출하여 국제 규격화를 시도했으나 무산된 바 있는데 받침 쌍자음 ㄱ, ㅋ의 처리방식이 규정되지 않아 모아쓰기 효율이 떨어지는 자판이었다. 아직 실용화여부는 불투명하지만 이번 논문에서는 이에 기초하여 쌍자음 5개를 상단에 추가하여 위 타건으로 제안하였다. 이런 구조는 현재의 남한의 자판과 매우 유사하다. 배열도 자음 왼쪽 모음 오른쪽의 구조를 유지하고 있다. ㅋ, ㆁ의 처리와 일부 모음과 자음의 위치를 조정하는 문제는 있으나 기본구조상의 차이는 없다. 상세한 구조는 아래 그림과 같다.

3) 북한의 전자자동화위원회 컴퓨터 센터 실장

4) 북한 사회과학 통보 연구소 부소장

ㄱ	ㅋ	ㆁ	ㄷ	ㅌ	ㅍ	ㅊ	ㅅ	ㅆ	ㅇ	ㅈ	ㅊ				
ㄴ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ			
ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ				

그림 1. 남한의 자판배열(KSC 5715)

ㄱ	ㅋ	ㆁ	ㄷ	ㅌ	ㅍ	ㅊ	ㅅ	ㅆ	ㅇ	ㅈ	ㅊ				
ㄴ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ			
ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ				

그림 2. 북한의 31타건(시안으로 제안된 것임)

(4) 북한에서의 글꼴(폰트) 문제

1985년 부터 인자용 24점 청봉체와 터미널 현시용 16점 천리마 글꼴을 사용해 왔으며 조립형 반조립형과 완전글자를 배합 사용하여 왔다.

조립형의 요소는 모두 579개로 10734자를 형성시키고 반조립으로 324자, 완전 글자로 14자를 생성하는 방법을 취한 것이다. 그러나 점폰트 방식은 인쇄방식에서는 작용이 쉽지 않아 윤곽선 글꼴 생성방법을 고찰하여 타상 출판체계용 레이저인자기, 윈도우 3.1 체계에 적용하고 있다. 그러나 보다 전문적인 출판체계에서는 서체(글꼴이 아님)의 다양성이 중요한데 북한에서는 고딕, 청봉, 명조, 붓글씨체 등이 있다고 한다. 이들 서체에서 적용되는 위치 형태 변화를 고려하여 분석한 다음 조립식 반조립식 완전글자로 나누어 처리하는 윤곽폰트 작성을 고려하였다.

그러나 보다 구체적 내용에는 언급이 없으며 이들 모두에 적용할 수 있는지는 분명치 않다. 그러나 조립식, 반조립식, 완전글자의 아이디어는 살만하지만 역시 다양성에는 못미칠 것으로 생각된다.

참고로 문체부에서 개발한 서체변형을 위한 요소는 902개가 되어야 하는 것으로 조사된바 있는데, 이로 볼때 579개로는 섬세한 부분까지 난점이 많을것이며 다양성을 살리지 못할 것이다.

(5) 정보처리용 조합형 부호화문제

정보처리용부호체계는 ISO-2022에 따라 완성형을 쓰고 있지만 이 부호를 이용하면서 개인컴과 매킨토시에서 사용가능한 조립식 2바이트 부호체계로는 MS/DOS체계를 사용한다. 처음소리 5, 가운데소리 5, 나중소리 5비트로 구성된 2바이트 조합형부호를 MS/DOS에 배치한 것이다. 부호이용율이 저하되는 단점이 있으나 검색과 정정이 용이한 장점이 있다. 이는 조합형글꼴과도 연결이 쉬워진다. 통합 2바이트코드에도 적용가능한 이점이 있다.

그러나 조합형(이북에서는 조립식)부호의 최대난점은 국제적 호환성 여부인데 이점에 대해서는 구체적 언급이 없는 것을 볼 때 북한은 이 문제가 심각히 대두되는 상황이 아닌 것으로 짐작된다.

결어

이상에서 자판, 부호 및 글꼴에 대한 문제를 고찰하였다. 한마디로 말하자면 북한은 이용수준과 보급율의 저하로 앞으로 더 큰 문제의 심각성보다는 초기적, 그리고 이론적(명분적)인 면에 치우친 감이 든다. 당연한 귀결일 것이다.

자판의 경우를 보면 91년 ISO제출안보다 우리의 표준안에 훨씬 근접한 안을 제시하고 있는 것은 사용하면서 결점이 나타나는 전산의 특징을 잘 반영하는 예일 것이다. 부호체계 또한 유사한 경향인 바 앞서간 우리의 경우를 외면할 수 없다. 중국 또한 마찬가지다. 우리는 시행착오의 경험을 가지고 있으므로 이의 노하우와 자료를 더 많이 보내 그들의 시행착오를 줄여줄 의무가 있다.

그런면에서 이번의 3자모임은 그야말로 황금의 3분할이다. 내년 5월에 다시 만나 자판, 글꼴, 부호에 대한 분과회의를 가지자는 협의도 있는 만큼 희망적인 낙관을 가져도 좋을 것이다. 그러나 실제 시행되기까지는 멀고 험난한 길이다. 왜냐하면 99%가 합의되도 부호 하나가 다르면 시행될 수 없는 것이 전산분야의 맹점이기 때문이다. 설사 합의가 이루어진다 해도 실제로 이용되기까지는 얼마나 많은 세월이 필요할지 모른다. 희망사항이지만 가장 바람직한 것은 보급률이 낮은 북한에서 수정해오는 것이 전 민족적인 경지에서 경제적인 것이다.

이런 문제외에도 다가올 문제 특히 남북에서 아직 확정되지 못한 문제들에 대해 지속적인 관심이 요구된다.