

## □ 기획연재 □

## 컴퓨터 과학 산책(22)

## 우리말과 정보처리

시스템공학연구소 박동인\*

## 우리말과 우리글

훔볼트(Humboldt)는 모든 민족의 모국어에는 그 민족 고유의 얼이 담겨 있다고 말하였다. 왜냐하면 언어는 그 언어를 사용하는 공동체 구성원들의 사유와 지각과 삶의 세계에 대한 이해를 일정하게 제약하고 이를 통해서 그들만의 고유한 의식 구조와 가치관과 문화 창조를 가능케 하기 때문이다. 그러므로 언어는 그 언어를 사용하는 공동체의 정체성을 의미하며 그들과 운명을 같이하는 상호 불가분의 긴밀한 관계를 가진다.

우리에게는 우리말과 한글이라 불려온 우리글이 있다. 우리 민족의 얼을 음성으로 담은 것은 우리말이요, 그 말을 담은 것이 우리글로서 우리말과 우리글은 우리 민족의 흥망의 비전인 것이다. 그러기에 일제 식민 통치는 우리 민족의 얼을 빼앗기 위해 우리말과 우리글의 사용을 금하였었고 우리의 선열들은 우리 겨레의 생명인 우리말과 우리글 수호에 온 몸을 던졌던 것이다. 우리말과 우리글을 지키고 다듬는 일은 우리 민족의 자존과 발전을 위한 길과 이어지는 이명동질(異名同質)의 것인 까닭이었다.

### 보다 용이한 정보 기기 사용 환경으로서의 우리말과 우리글

종래의 농업화 사회와 공업화 사회에서는 가치를 생산하는 데 있어서 물질과 에너지가 가장 중요한 자원이었으나 오늘날에는 정보가 가장 중요한 자원이 된다. 그 어느 때보다 냉혹

하고 치열해진 국제 경쟁의 관계에서는 오직 정확한 정보를 빨리 얻는 자만이 무한 경쟁에서 승리할 수 있기 때문이다. 정보의 중요성이 강조되는 오늘날의 사회를 흔히들 정보화 사회라고들 부른다. 정보화 사회의 가장 두드러진 특징으로는 일반화된 정보 기기의 사용과 다량의 정보 유통 그리고 그에 따른 신속 정확한 정보 획득의 필요성의 대두 등이다. 사정은 우리의 경우에도 마찬가지이다. 급증하고 있는 정보 기기의 보급과 일반화된 활용은 우리에게도 정보화의 물결이 닥치고 있음을 피부로 느끼게 한다. 우리에게도 정보 기기는 더 이상 전문가들만의 전유물이 아니라 누구나 자신의 정보를 저장하고 필요한 정보를 검색하기 위해 사용할 수 있는 일상적인 생활의 도구로서 자리 잡게 된 것이다.

이와 같은 정보 기기 사용의 일반화는 새로운 정보기술의 등장을 요구하게 되었다. 사용이 보다 용이한 정보 기기 사용 환경(User Interface)에 대한 요구가 그것이다. 물론 초창기의 명령어 입력 방식(Command Line User Interface)에 비해 요즘의 GUI(Graphic User Interface) 방식은 정보 기기 사용을 훨씬 편리하게 하였다. 그러나 GUI 방식 또한 오늘날의 사용자들의 요구를 만족시킬 수 있는 궁극적인 해결책이라고 보기 어렵다. 여전히 사용자들은 사용자 특성이 충분히 고려되지 않은 “획일적인 정보 기기 사용 환경”을 배워야 하며, 그에 따른 심리적 부담감은 사용자들의 정보 기기 이용의 효용성 극대화를 가로막는 걸림돌로 작용하고 있기 때문이다. 그렇다면 사용자를 위한 진정한 정보 기기 사용 환경이란

\*중신회원

무엇인가? 이 질문에 대답하기 위해, 우리는 인간의 가장 자연스러운 의사 소통 수단인 언어, 즉 자연어(Natural Language)를 사용해서 작업을 할 수 있는 환경을 사용자를 위한 진정한 작업 환경들 중 하나로 제시할 수 있을 것이다. 그와 같은 작업 환경 하에서는 사용자는 자신의 모국어로 자연스럽게 작업을 할 수 있으며 사용법을 배우는 데에 투자되는 노력의 낭비를 줄임으로써 작업 효율성의 극대화를 기대할 수 있을 것이다. 자연어를 이용한 정보처리 기술은 예의 사용자 환경에서뿐만 아니라 정보검색이나 자동번역 등과 같은 분야들에서도 그 개발이 절실히 요구되고 있다. 보편화된 정보 기기의 사용의 결과 유통되고 있는 자연어로 된 다량의 유익한 정보들을 자연어 정보처리 기술을 이용하여 검색하고 분석하며 때로는 번역을 하여 필요한 정보를 신속하게 획득하고, 궁극적으로는 개인 및 국가의 경쟁력 향상을 도모할 수 있기 때문이다.

자연어를 이용한 정보처리 기술에 대한 이와 같은 비전은 새로운 것이 아니다. 이미 서구 선진국들은 우리보다 앞서 자신들의 언어를 사용한 정보처리 기술 개발을 시작하였으며 우리의 경우에도 우리말을 이용한 도구 개발이 절실한 상태이다.

### 관련 학문 분야들간의 공조체제

우리말과 우리글을 이용한 정보처리 기술 개발은 경제, 산업 등 다양한 분야에 효율성 증대, 비용 절감 그리고 경쟁력 강화라는 경제적 측면에서의 긍정적인 파급 효과를 가져다 줄 수 있을 것이다. 그러나 무엇보다도 중요한 것은 문화적 측면에서의 의미이다. 우선, 자연어 정보처리 기술 개발은 정보처리학뿐만 아니라 언어학을 포함한 언어 연구에 관련된 다양한 학문 분야들 간의 협력을 필요로 한다. 따라서 우리의 말과 글에 대한 관련 분야들에서의 연구의 확대 및 심화를 통해 우리말 정보처리는 우리의 말과 글을 더욱 더 다듬고 발전시키는 데에 큰 기여를 할 수가 있을 것이다. 그러나 우리말 정보처리 기술이 만약 우리가 아닌 남의 손에 의해 개발되고 그 기술이 우리의 정보

사회를 지배하게 된다면, 우리는 막대한 경제적 손실 이외에도 엄청난 문화적인 혼란을, 나아가서는 우리 스스로의 모습마저 잃어버리게 되는 민족 정체성의 위기까지도 맞이할 수 있을 것이다. 우리 민족의 사유와 지각과 삶의 세계에 대한 이해를 이끌어 가는 일이 절여된 왜곡된 말과 글이 우리의 진정한 글과 말을 잠식할 것이기 분명하기 때문이다.

### 언어에 대한 인식의 차이

우리말 정보처리 연구는 분야의 특수성으로 인해 정보처리학자들과 특히 언어학자들의 공동 작업을 필요로 한다. 물론 정보처리학자들만의 단독 작업도 불가능한 것만은 아니다. 그러나 그 경우 별도로 언어 자료 분석을 병행해야 함에 따른 작업량과 필요 노력의 증가뿐 아니라, 언어 자료 분석 결과가 경험과 기존의 자료가 풍부한 언어학자들의 분석 결과만큼 만족스럽지 못 할 수도 있다는 문제점이 있다. 따라서 언어학자들은 정보처리를 전제로 하여 언어 자료들을 분석을 하고 정보처리학자는 그 분석된 결과들에 근거하여 정보처리 도구를 구축하는 것이 두 학문 분야간의 가장 이상적인 공조체제일 것이다. 그러나 실제로는 이 두 학문 분야간의 공동 작업의 예는 매우 드물다. 관점은 다를 수 있겠지만, 동일한 우리말과 우리글에 대한 연구를 하면서도 정보처리학 분야와 언어학 분야는 서로를 의식하지 않은 채 그 고유한 연구 영역 내에만 머물러 있는 것이 우리의 현실인 것이다.

아마도 이러한 우리의 현실은 다음의 몇 가지 점들로 분석이 될 수가 있을 것이다. 우선 두 학문 분야의 연구 대상인 언어에 대한 인식의 차이를 들 수 있다. 정보처리학자들은 종종 처리 대상인 인간의 언어를 비교적 그 내부 구조가 간결한 프로그래밍 언어와 동일시하는 경향이 있다. 이와 같은 언어에 대한 견해는 언어 자료 분석 시 언어의 복잡한 측면을 간과케 하거나 혹은 단순화, 일반화를 시키려는 태도로 직접적으로 드러난다. 물론 이러한 견해는 지난 수년간의 연구 과정 중에 많이 개선된 바가 없지는 않으나, 여전히 그 문제점들이 지적

될 수 있을 정도로, 많은 부분에서 아직도 발전이 된다. 정보처리학자들의 이와 같은 태도는 모국어 화자로서 모국어에 대한 지식이 있으므로 그를 바탕으로 그 모국어 정보처리를 할 수 있다는 인식과도 관련이 있다. 물론 언어 자료의 기술 및 분석에 그 언어에 대한 지식이 근본이 됨에는 이론의 여지가 있을 수가 없을 것이다. 그러나 그 언어에 대한 지식이 있다는 것이 그 언어를 객관적이고도 정확하게 분석을 할 수 있는 능력과 자질을 가지고 있다는 것을 의미하지 않는다는 것 역시 자명하다. 반면, 언어학자들의 경우, 언어는 다양하고 이질적인 요소들이 뒤엉켜 있는 복잡한 구조로 인식이 된다. 이와 같은 언어에 대한 언어학자들의 이해는 언어에 대한 모델링이나 이론적 접근 시도에 대한 그들의 회의적 태도에 의해 단적으로 드러난다. 그들은 항상 예외 현상에 대한 많은 배려를 하며, 언어 현상 전반에 대한 폭 넓은 분석보다는 작은 현상에 대한 깊은 분석을 하며 또한 문제의 언어 현상에 개입하는 다양한 이질적인 요소들을 열거하고 그것들에 대한 자세한 주석을 한다. 그러나 정보처리 학자들에게도 언어학자들의 연구 결과를 그들의 연구의 출발점으로 삼기가 어려운 나름대로의 타당한 이유가 있다. 우선 정보처리학자들에게는 처리해야 될 언어의 전체적인 기술이나 규칙들이 필요하다. 그러나 대개의 언어학적 연구들은 전반적인 연구보다는 작은 현상에 대한 부분적인 연구가 주를 이루고 있을 뿐이다. 또한 개별 연구들이 근거한 이론적 배경들 간의 비호환성과 연구 결과물들 간의 불일치 등 역시 정보처리학자들이 언어학적 연구 결과물

들에 근거하여 작업을 하는 것을 어렵게 만드는 큰 걸림돌들인 것이다. 우리말 정보처리를 위한 정보처리학과 언어학의 공동 연구에 대한 또 다른 장애 요인으로 언어학자들의 정보처리학에 대한 이해의 결여를 지적할 수 있다. 물론 언어학은 정보처리학과는 구분된 연구 분야로서 그 연구의 독립성을 갖는다. 그러나 기술 방법이나 이론적 배경의 특성상 정보처리 도구로 구현될 수 없는 연구물들은 그 연구 결과의 높은 가치와 투지한 노력을 고려할 때 정보처리학자들에게는 많은 아쉬움을 남긴다.

### 우리말 정보화의 시대적 당위성

우리는 지금까지 우리말 정보화를 위해서는 관련된 제반 분야들, 특히 정보처리학과 언어학간의 긴밀한 협조 체계가 필요함을 지적하였다. 그러나 이러한 바람과는 달리 이 두 학문 분야간의 공조 사례는 매우 드물다는 사실을 보았고 그 공조체제를 가로막는 원인들에 대한 간단한 분석을 하였다. 물론, 공동의 목표를 위해서라고 할지라도, 인문과학과 공학의 경계선을 넘나드는 것이 우리에게 아직 익숙하지 않을지도 모른다. 그러나 역사는 우리에게 우리말을 우리 방식대로 쓸 수 없는 것이 얼마나 큰 아픔인지를 이미 보여주었다. 그러므로 이제는 우리의 말과 글, 더 나아가서는 우리의 자존과 발전 도모라는 민족의 당위적 과제를 위해 서로를 이해하고 서로에게 귀 기울이며 인문학과 공학이라는 기존의 구분 틀을 넘어선 공동 연구 수행을 활성화하기 위해 함께 노력을 해야 할 때가 아닌가 생각한다.