

세계어로서의 한국어 정보처리

이 동 한

충북대학교 자연과학대학 통계학과

Information Processing of the Korean Language as a World Language

Lee, Dong Han

Department of Statistics, College of Natural Sciences,

Chungbuk National University

요 약

보편적으로 일반 문법의 질서에 부합될 수 있는 자연언어들이 상호 번역이라는 수단을 통해 언어 변환이 가능하다면 이들 언어는 세계어이다. 한국어도 그 하나에 속한다. 본 논문에서는 한국어를 세계어라고 하는 보다 넓은 관점에서 보아 가면서, 이에 따라 고려해야 할 문제들을 검토함으로써 한국어 정보처리의 개념을 정립하고, 나아가 세계 속의 한국어 위치를 다듬어 보고자 하는 것이 본 논문의 목표이다.

1. 세계어의 개념

여기서 언급하는 세계어란 에스페란토와 같은 국제보조어나, 아니면 국제적으로 공용될 수 있는 범지구화된 언어를 뜻하는 것은 아니다. 상당한 수의 인구가 문화적 및 역사적 전통을 계승해 오면서 사용하는 언어로서 개별 문법 질서하에 체계화될 뿐만 아니라, 이것이 일반문법[2]의 보편성에 조화되며 그 이론의 틀에 기여될 수 있는 특정 국가의 국어를 지칭한 개념이다. 이런 관점에서 보면 특정 국가의 국어로 사용되는 모든 자연언어는 일단 모두 세계어의 범주에 속한다 할 수 있다.

한국어는 위에서 말한 개념에 부합될 뿐만 아니라, 한글 자모의 문자체계는 그 조직성과 과학성이 세계적으로 인정되기 때문에 한국어는 세계어로서 자격이 있는 언어이다. 각 국가들의 언어가 문법이라고하는 프로토콜을 통해 국제적 정보 교환을 수행할 수 있을 경우 국가 수준의 언어는 바로 세계 수준의 언어가 되는 것이다. 그러나, 만약 한국어에 국제적인 언어 개념을 완전히 수용할 수 없는 부분이 존재한다면 한국어는 세계어의 자격을 얻는데 하나의 흠이 된다할 것이다. 이런 의미에 있어서 한국어 문법체계는 어휘를 필두로 해서 여타 국가의 문법체계와의 교환을 위해 보다 조직화되고 체계화되어야 한다. 한·

중·일의 통합한자 공동 연구회의 활동과 같은 것은 이들 국가들의 국어가 세계어의 과정을 취하는 좋은 보기가 될 수 있다[4].

어떤 프로토콜로서도 상호 정보교환(문법의 교환)이 될 수 없는 고립어(예, 古漢文)라면 그런 언어는 세계어가 될 수 없다. 그러나, 대부분의 자연언어는 상호 번역이 보편적이기 때문에 현재 각 국가가 국어로 사용하고 있는 언어는 일단 세계어로 볼 수 있다. 한 국가의 국어를 독립된 개별 국가의 국어라고 하는 입장에서 한 걸음 더 나아가 세계어의 관점에서 고찰해야 할 필요성은 세계의 모든 국가들이 추구하는 정치적, 경제적 및 문화적 민주정신(분권사상)에서 부터 비롯되는 것이다.

동일한 의미의 세계가 상이한 표현 방식의 차이로 인하여 모호해 지거나 복수의 의미의 세계를 지적하게 된다면 정보교환의 효율성은 없어진다. 이런 불편을 없애기 위해 적어도 기술적인 분야에 있어서만은 통일된 의미의 세계를 견지하도록 하는 자연언어의 세계어 지향적인 노력은 국제적 관심사이기 때문에 우리도 한국어의 여러 체계들을 이런 시각에서 재검토해 보고 나아가 한국어의 정보처리에 관한 여러 문제점들을 분석해 보아야 할 필요가 있다[5].

2. 자본의 세계화와 정보의 증대

정보는 물질이나 에너지에서 검출되는 어떤 형태 또는 질서이다[3]. 에너지는 자연적인 것은 물론이고, 인위적인 것도 포함된다. 정보는 모든 사물에서 검출되며 인간의 두뇌와 사고 속에서 일어나는 사상, 관념, 의지 등과 같은 여하한 종류의 개념적 에너지에서도 검출될 수 있다. 검출된 것이 어떤 형태를 갖거나 질서를 갖으면 그것이 정보가 된다. 정보는 언어라는 도구를 통해 현실화 된다.

공간적으로 보면 정보의 생산지역은 한 국가의 국어가 지배하는 모든 공간적 영역이다. 농경시대에 있어서는 정보의 생산지역은 그 국어가 지배하는 제한된 지역이었다. 그러나, 범세계적으로 경제적 활동과 자본 시장이 보편화되고 한편, 무역 거래가 빈번히 일어나는 현대 사회에 있어서는 정보의 생산 공간은 자국의 국어가 지배하는 영역을 벗어나 전세계적인 규모로 확대되고 있다해도 과언이 아니다. 우리나라는 무역거래면에 있어서 세계 상위권에 속해 있고, 오대양 육대주에 거래선이 다변화되고 있으므로 한국어의 정보생산 공간도 자연적으로 한반도에서 전세계적인 공간으로 확대 되어가고 있다.

한국어의 언어공간은 한반도가 그 전부는 아니다. 해외 교포의 거주지(집단적인)나 우리나라에서 투자하는 해외 자본시장은 모두 한국어의 언어 공간이 될 수 있다. 활발한 무역과 해외 이민은 한국어의 언어 공간을 넓히는 중요한 계기이다. 따라서, 한국어가 생산하는 정보공간이 한반도를 넘어 전세계에 산재해 있다고 볼 수 있는 것이다. 그러므로,

자본의 세계화가 촉진되면 언어로 표현되는 정보의 생산량도 증대되는 것이다.

상품의 수입과 수출은 그것이 인간의 사고 활동과 지적 산물의 교환이란 면으로 볼 때, 상품의 수출입은 언어의 수입이고 언어의 수출이라 할 수 있다. 언어의 수입과 수출을 통해 정보 교환이 이루어진다. 무역거래에 있어서 언어의 교환은 목적으로하는 상품 그 자체보다도 오히려 더 우선적이다. 재화의 유통은 언어 유통과 이를 바탕으로하는 정보의 유통이 있으므로 비로소 가능한 것이기 때문이다. 그러므로, 상품의 국제적 표준화는 당연히 언어의 국제적 표준화의 필요성을 불러 일으키게 되는 것이다. 최근에 국제 용어은행[6]을 만들자고하는 움직임도 모두 이런 관점에서 출발된 것이라 볼 수 있다. 따라서, 각 국가는 자국의 국어가 세계어로 반전되기 위한 필요한 조치를 취하게 되는 것이다.

3. 언어 민주주의

지적 소유권이 국제 사회의 법률적인 요소로 등장한 것은 최근의 일이지만, 개인이나 국가적 단위의 지적 공간의 불가침을 선언하는 것은 바로 자기 언어의 불가침성을 주장하는 것과 같은 것이다. 순수한 학문적 성과에 대해 화폐단위로 그 소유권을 주장한 일은 과거에는 없었던 일로서 경제적 가치가 인류 사회를 풍미하고 있는 오늘날에는 매우 중요한 국제 관계의 요소가 되었다. 이것은 다시 말하면 개인이나 국가가 각자의 언어공간의 영유권을 주장하는 것이라 말할 수 있는 것이다.

언어공간의 영유권을 주장하는 것은 자기의 권익을 확보하려는 의도로서 이는 언어민주주의라고 표현할 수 있다. 즉, 언어민주주의란 언어의 다원주의를 뜻하는 말이다. 민주주의 하에서는 예를 들어 “사람”과 “man”이 동일한 의미의 세계를 지칭한다 할지라도 어느 한쪽의 존재성이 양보되거나 소멸될 수 없다. 말하자면 한쪽은 계속 “사람”이라는 표현을 영구적으로 사용한다면 다른 한쪽도 역시 “man”이라는 표현을 영구적으로 사용하는 것이 바로 언어민주주의다. 예를 들면 한국어의 언어세계를 영어가 침범할 수 없고 영어의 언어세계를 한국어가 침범할 수 없다는 것이 언어민주주의이다.

비록 두 언어사이에 접촉이 긴밀하게 됨으로서 상호의 언어 체계에 영향을 주는 요소가 없지는 않겠지만 마치 사람이 소고기를 먹는다고 하여 소가 되지 않는 것과 같이 한국어에서 영어단어를 외래어 수준으로 사용한다해서 한국어의 체계가 손상되거나 또 그런 체계가 무시될 수는 없는 것이다. 더구나, 지적 소유권과 같은 고급가치를 자산으로 확보하기 위해서는 자기의 언어 영유권을 상실하면 그런 소유권을 획득할 기반을 잃게 되는 것이다. 비근한 예로서 세계 각국이 자국의 고유한 상품이나 독특한 상표를 개발하고자하는 노력도 바로 언어 민주주의를 견지하고자하는 일환이라 볼 수 있다.

금후, 언어 민주주의를 견지 하고자하는 노력은 정치적 독립 또는 경제 사회의 불간섭주의를 고수하는 것 만큼이나 절실히될 것이고 또 그런 경쟁은 지구상의 모든 국가들 사이에 더욱 치열해질 것이다. 역사적인 면으로 관찰해 보면 자국의 영토를 확장하고자 한 다든지 정치적 패권주의와 경제적 실익을 얻기 위해 벌리는 인류의 욕망의 경쟁을 전쟁이라 볼 수 있으나, 다른 한편으로 보면 언어의 침략이 바로 전쟁이고, 약소국의 입장으로 보면 언어민주주의를 존속시키기 위한 방어수단이 곧 전쟁이라고 볼 수도 있는 것이다.

유럽 공동체가 정치적 및 경제적 통합은 이루었다 할지라도 언어민주주의만은 쉽게 없어지지 않을 것이며 오히려 그런 주의가 강화될 것으로 필자는 전망하는 것이다. 사실 언어민주주의가 소멸되는 수준에서 역내의 여러국가들이 통합을 이룬다면 복수의 민족은 단일 민족화되는 것이다. 이런 현상이 앞으로 지구 내에 존재하는 민족국가들 사이에 어떻게 전개될런지는 예측할 수는 없다. 그러나, 언어민주주의에 대한 인간의 관습은 쉽사리 없어지지 않을 것이다. 이것은 인간의 두뇌를 비롯한 생리적 구조와도 밀접한 관계가 있을 것이며 또 지구표면상의 각 지역적 환경이 인간에게 주는 에너지(氣)와도 깊은 관계가 있을 것이다.

4. 언어의 또 다른 측면

언어속에는 실로 많은 것이 농축되어 있다. 인간의 사고, 상상력, 의지, 소망, 권력, 생활양식 등 실로 다양한 국민적인 요소가 언어 속에 잠복되어 있다. 그런 언어를 사용한다는 것은 그런 사고방식이나 인간의 모든 정신적 활동을 그대로 반영한다고 볼 수 있다. 따라서, 언어는 그것을 사용하는 민족이나 국가와 동의어라 할 수 있다. 그러므로, 한국어를 세계어의 관점에서 다룬다고 하는 것은 한국인의 국제화를 한층 더 높히는 결과가 되는 것이다. 특히, 금후의 정보화 사회에 있어서는 정보가 사회의 기간 회로망이 될 것이기 때문에 언어가 질서화되고 창조적이 되면 그것을 기반으로 해서 구축되는 조직이나 국가의 정보력도 커지게 마련이다.

노동자가 자유로이 자기의 언어로서 의사표시를 하는 것은 그 노동자의 힘이다. 만약, 그들에게 그런 언어가 없다면 노동자는 힘을 발휘하지 못하게 될 것이다. 마찬가지로 국가에도 국어력이 미약하면 그만큼 국력도 미약하게 된다. 그러므로, 금후의 정보화 사회에 있어서의 국어는 국가를 상징하고 국력을 발휘할 수 있는 기본적인 에너지가 된다고 할 것이다.

언어가 정보를 만들고 정보가 에너지가 될 수 있다는 것은 언어의 의미론적 측면보다는 언어의 통신론적 측면에서 비롯된다[1]. 의미론은 그 언표의 일방적인 선언에 불과하다. 의미론은 언어가 제삼자에게 어떤 의미로 전달 되는가 하는 것 보다는 화자의 입장에

서 어떤 의미를 나타 내고자 하는 화자의 의도 그 이상이 못된다. 그러나, 통신론적 측면에서는 화자의 의도가 청자에게 그대로 전달될 때 비로서 그 언어는 정보가 되며 언어 본래의 목적 즉, 언어의 의미를 완성하게 되는 것이다. 요는, 언표가 어떻게 되어야만 청자에게 화자의 의도를 완성시킬 수 있는 기능이 발휘될 수 있는가를 다루는 것이 통신론이다. 전통적인 언어학에 있어서는 구문론, 의미론, 음운론의 3요소로 연구 하지만, 하나의 자연언어를 세계어라는 관점에서 다루고자 할 때는 통신론적 분석을 언어학의 요소로 추가해야 할 것으로 생각되는 바이다. 음운적 범주에서 탈락되는 문자 간의 공간 배치, 괄호, 구두점과 같은 각종 문자 이외의 부호, 자형의 크기에 따른 통신 속도(인지 속도)와 정보 통신망에서 이루어지는 검사 부호등은 모두 언어의 통신적 요소라할 수 있다.

언어가 통신적인 기능을 발휘할 수 없다면 세계어라는 관점에서 볼 때 언어 본래의 소기의 목적을 달성하고 볼 수 없다. 하나의 동일한 사실을 일의적으로 상호 통신하기 위해서는 언어에는 보다 체계적이고 조직적인 프로토콜의 설정이 필요하게 되며 이런 프로토콜은 전통적인 언어학에서 아직까지 취급하지 않았던 통신론적 분석이 그 주된 요소가 될 것이다.

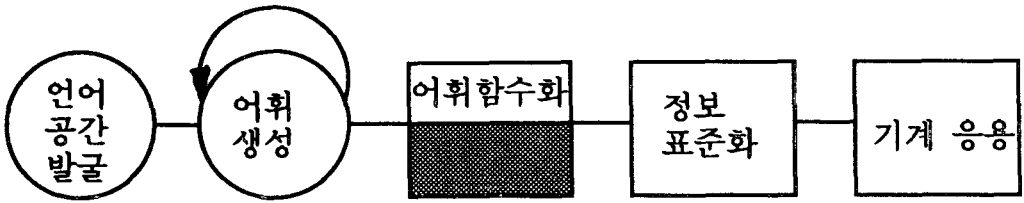
자연언어의 체계가 보다 인공적으로 조직화되면 통신적인 특성이 강화될 것이지만 비제약적인 예외사항이 많거나 문자 그대로 자연적인 면이 강하면 의미론적 특성은 나타날지는 몰라도 통신적인 특성은 매우 약화될 것이다. 어떤 문장이 의미론적 문장으로 손색이 없다하더라도 청자에게 일의적인 사실을 정보로서 전달하지 못한다면 그런 문장은 통신론적 문장이 될 수는 없는 것이다. 금후의 정보화 사회에 있어서는 의미론적 문장보다는 통신론적 문장이 보다 보편화될 것으로 전망되는 바이다.

예를 들어, 기계 설명서가 의미론적인 문장으로 쓰여진 것이라면 경제적 가치는 별로 없을 것이다. 기계설명서와 같은 제품은 모든 문장이 기계를 사용하고자 하는 사람에게 통신적 기능을 충분히 발휘하지 못한다면 그것은 설명서의 역할을 충분히 수행한다고 볼 수 없는 것이다. 엄밀히 말하면 정보란 통신적인 성격이 의미론적 성격을 포괄한다고 볼 수 있다. 통신론적 문장이란 화자보다는 청자에게 어떤 인지적 효과를 나타내는 가에 주목하여 기술되는 문장이라할 것이다. 따라서, 한국어 정보처리를 위해서는 인지론적 연구를 깊게해야 한다.

5. 한국어 정보처리의 체계

문자등을 기계에 입력하고 이를 저장 검색하며 논리적 조작을 하는 이른바 언어의 기계적 처리가 한국어 정보 처리의 전부는 아니다. 한국어 정보처리란 [그림]과 같이 언어의 시원에서 부터 기계처리에 이르는 총체적이고 시스템적인 전과정을 망나해야 한다. 왜

나하면 언어의 정보처리란 인간과 기계라고 하는 두 실체의 공동노력과 협조에 의해 이루어지기 때문이다.



[그림] 한국어 정보처리의 기본 체계

기계는 인간의 논리에 의거하여 그런 논리를 어김없이 실현하는 물체이고, 인간은 기계동작의 원인을 제공해 주는 것이기 때문에 어느 한쪽의 작업을 생략한다면 언어의 정보처리란 성공적으로 이루어질 수 없다. 따라서, 한국어 정보처리란 한국어라는 원어민과 정보처리라고 하는 기계가 협력하여 이루는 이른바 [인간 - 기계] 체제이다.

위 그림에서 첫째, 언어공간 발굴이란 한국어의 원어민이 갖는 언어 창조력이다. 이 창조력은 원어민의 의식에서 나오는 일종의 에너지이다. 한국어의 원어민은 육체적 경험과 정신적 직관을 통해 그 의식 세계에 언어공간을 생성 확대해 나가는 것이다. 언어공간의 발굴에서부터 원어민의 일체의 언어적 생활이 시작되는 것이다. 이런 언어 공간의 발굴은 개인적 및 국가적 노력이 함께 요구된다.

둘째, 어휘 생성이란 새로운 언어적 개념을 담은 어휘가 개인적인 또는, 집단적인 노력에 의해 만들어지는 원어민의 발표력을 말한다. 원어민이 학문을 연마하거나 기술을 개발할 경우 결론으로 얻어지는 사항들이 한국어라고 하는 특수한 언어적 도구로 발표되는 각종 어휘의 집합이 어휘생성이다. 이 단계는 동일한 학문적 또는 기술적인 결과들이 복수의 어휘로 개발되어 표준화 조치를 기다리는 시점이다. 이 단계에서는 어떤 개념을 단일한 어휘로 묶으려고 고집할 필요는 없으며, 가능한한 다양한 표현의 후보 어휘가 제안되는 것이 바람직하다.

셋째, 어휘 함수화한 위 둘째 단계에서 생성된 어휘 집합을 표준화해서 통일적인 국어로 제작하는 조치를 말한다. 표준화 작업에는 어휘 생산자와 일반의 관련자들이 참여해서 미주적인 토의와 절차를 거쳐 다양한 후보 어휘를 단일 안으로 묶어 그 개념을 확정하는 일련의 과정이 모두 포함된다. 이럴 경우 하나의 어휘는 전통적인 사전적 개념보다 더 조직화되고 강력한 의미전달을 내포하기 때문에 통신의 효율성을 증대시킨다. 언어공간 발굴, 어휘 생성 및 어휘 함수화까지의 단계는 한국어의 창조와 재생산 과정이겠으나, 세계언

어인 다른 외국어의 국내화 과정에도 준용되어야 한다. 또, 한국어를 국제 표준으로 제안하는 일도 여기에 포함된다.

넷째, 정보 표준화란 언어의 기계처리에 기본이 되는 자형의 제작과 언어의 부호화 및 이에 수반되는 작업을 말한다. 이 단계는 위에서 말한 언어의 통신논적 측면이 언어 처리를 위해서 보다 강화되는 단계라 말할 수 있다. 정보 표준화에는 언어의 구문논적, 의미론적 및 통신논적 개념을 기계화로 유도하는에 이미 성공했거나 머지 않은 장래에 성공할 것이 기약되지만, 언어의 중요 요소의 하나이며 오히려 근본적이라 할수 있는 음운논적 개념의 정보표준화에 대해서는 아직 이렇다할 성과가 보이지 않고 있다. 이 분야는 계속적인 연구 과제가 될 것이다.

다섯째, 기계 응용이란 언어 정보의 입력과 처리 및 결과의 3단계를 기계화 하는 전 과정을 뜻한다. 언어 처리에 있어 인간적인 것을 표준으로 해서 보면 기계는 질서의 극치인 반면, 기계적인 것을 표준으로 해서 보면 인간은 무질서의 극치이다. 따라서 양자의 거리는 상당한 것인데, 양자의 간격을 좁히기 위해서는 계면 기능을 강화 해야 한다. 그러나, 인간을 기계에 접근시켜서는 안되며 어디 까지나 기계를 인간적인 측면으로 접근시킬 수 있는 계면기능의 개발이 요구되는 것이다.

위와 같은 한국어 정보처리의 개념 정립은 시스템 설계적이다. 언어 정보가 인간의 문화 발전과 더불어 발전을 거듭하고 있기때문에 어휘 함수는 장기간 동안 고정 될수 없다. 즉 어휘함수의 생면주기는 과학 기술의 발전속도에 비례하여 설정될 것이다. 과학기술의 발전과 더불어 정보통신력이 없는 언어는 사라질 것이며 새롭고 강력한 정보통신력을 갖는 언어가 속속 생겨날 것이다. 이런 거시적인 안목에서 한국어의 정보처리를 발전시켜야 할것이다. 언어학의 분야를 구문론, 의미론, 음운론, 통신론의 4가지 범주로 생각해 볼 때 이 4가지 범주를 상당한 수준으로 처리할 수 있는 기계응용을 개발한다면 언어의 정보 처리는 획기적인 발전을 가져올 것이다. 이런 발전의 노력은 기계론적 분야와 인지론적 분야의 합작으로 이루어 질것으로 예상되기 때문에 한국어 정보처리는 한국민의 총체적 경험을 필요로 하는 것이다.

6. 앞으로의 과제

서로 다른 언어 공간들이 변환과정을 통해 상호 변환 될 수 없으면 세계어가 못된다. 세계어가 될 수 있는 가장 핵심적인 요소는 변환 과정을 통한 언어공간의 상호 일치이다. 이런 일치는 정보통신을 능률화하고 국제적 경제성을 증대시킨다. 또 이것은 한 언어가 세계어의 자격을 계속 유지함을 뜻한다. 한 국가의 국어가 세계어의 자격을 계속 누리기 위해서는 부단히 기술발전에 따른 국제적인 언어공간의 생성에 적극 참여하여 다른 언어

와의 상호변환이 가능토록 노력해야 할 것이다. 예를 들어 어떤 영어의 어휘를 번역할 한국어가 없다면 한국어는 적어도 그 부분에서만은 세계어의 구실을 못한다고 말할 수 있다. 현재 정보 처리 분야의 많은 어휘들이 아직도 한국어로 번역되지 못하고 있다는 사실은 한국어의 세계어화 과정에서 넘어 가야할 관문이며 커다란 과제이다. 명실 공히 한국어가 세계어로 되기 위해서는 국내활동과 국제활동의 두가지 방면에서 국가적 및 국민적 노력이 요망된다.

첫째. 국내적 활동에서는 한국어 고유의 언어공간을 넓히고 사유를 촉발하여 기술을 발전시키는 일이다. 이렇게 될 때 우리 고유의 기술 개발도 가능하여 국제 경쟁력도 높일 수 있다.

둘째. 국제적 활동에 있어서는 한국어를 세계어로 계속 유지 발전시키는 노력이다. 국제성 어휘들을 한국어로 번역 수용하고 우리가 창조하는 한국어를 국제화 함으로서 한국어는 세계어로서의 자격이 유지되는 것이다.

이와 같은 국내적 국제적 노력은 한국어의 정보처리를 활성화 시킴과 동시에 과학 기술의 여러 분야에 있어서 국제성을 발휘케되는 기본 요소가 될 것이다. 더욱 한국이 무역 의존형 국가라는 것을 감안할 때 세계어로서의 한국어 정보처리에는 모든 분야를 뒷받침할 수 있는 기반이 될 것이 분명하다.

[참고 문헌]

- [1] 이동한, 정보화사회에 있어서의 한국어 경제론, 우리말정보화잔치, 국어정보학회, 1992. 6, p. 2
- [2] 李乙煥, 韓國語文法論, 敎學社, 1981, p. 18
- [3] Jacques. attali, La parole dt l'outil(日역판), 日本評論社, 1983, p. 52
- [4] 정보산업표준원, 정보산업표준원 주간 소식 모음, 1992.1.27 - 3.30
- [5] 정보산업표준원, ISO 규격해설, 1992
- [6] 정보산업표준원, 국제용어은행 설립에 관한 자료, 1992. 8