

# 영한 기계번역의 자연어 생성 연구

홍성룡\*

## 요약

기계번역에서 자연어 생성의 목적은 입력언어의 어구 분석을 이용하여 그 문장의 의미를 변환해주는 목적 언어를 생성하는 것이다. 그것은 언어적 구조, 낱말 전사, 대화체 언어, 어휘적 정보 등을 포함해야 한다.

본 연구에서는 대화체 자동 기계번역 시스템 구현계획의 일부인 음성, 음운 분야에서 담당하게 될 음성인식과 음성합성 알고리즘을 확립하기 위한 한국어 특질에 대한 기초조사를 하고자 한다. 또한 기계번역의 단계를 분석하여 형태소 분석 단계와 구문 분석 단계, 의미 분석 단계로 구분한다. 형태소 분석은 입력 문장을 받아 분리된 형태소를 사전 내에서 검색하여 품사 정보를 얻고 이웃하는 단어와의 접속 관계가 문법적으로 올바르게 되었는지를 점검한다.

본 연구의 결과가 대화체 기계번역 시스템 구현계획의 종합적 입장에서는 단순한 기초조사일 수 있지만, 한국어의 교육 및 기계번역 이해의 측면에서는 그 자체로 가치를 지닌다고 할 수 있겠다. 따라서 교육적 측면에서의 직접적 활용을 여러 측면에서 고려할 수 있을 것이다.

## A Study on the Natural Language Generation by Machine Translation

Sung-Ryong Hong\*

### Abstract

In machine translation, the goal of natural language generation is to produce an target sentence transmitting the meaning of source sentence by using an parsing tree of source sentence and target expressions. It provides generator with linguistic structures, word mapping, part-of-speech, lexical information.

The purpose of this study is to research the Korean Characteristics which could be used for the establishment of an algorithm in speech recognition and composite sound. This is a part of realization for the plan of automatic machine translation. The stage of MT is divided into the level of morphemic, semantic analysis and syntactic construction.

Key words : 자연어 생성, 음성 인식, 번역 단계, 기계번역

### 1. 서론

국내에서 개발 중이거나 개발된 기계 번역 시스템들의 대부분은 직접 번역 방식이나 변환 방식만을 채택하고 있기 때문에 그 발전은 이미 한계에 도달하였다. 이 두 방식의 문제점을 해결하기 위해서 다양한 번역 모델을 복합한 엔진 기계번역 모델의 개발과 대규모 이중 언어, 개념 사전, 다중 언어 어휘의 구축이 시급하다.

현재 우리나라에서는 여러 연구기관들에 의해 가장 효율 가치가 높은 영한 번역 연구가 시행되고 있다. 외국의 경우, 이전의 기계 번역의 한계를 인식하고 새로운 방식의 기계 번역 시스템을 만들기 위해 연구가 활발히 이루어지고 있다.

이처럼 전 세계적으로 기계 번역 시스템의 구현에

힘쓰는 것은 그 효율 가치가 높기 때문이다. 기계 번역이 제대로 이루어진다면, 무역이나 웹 문서의 번역으로 정보를 한층 많이 얻을 수 있게 된다. 본 연구에서는 대화체 자동 기계번역 시스템 구현계획의 일부인 음성, 음운 분야에서 담당하게 될 음성인식과 음성합성 알고리즘을 확립하기 위한 한국어 특질<sup>1)</sup>에 대한 기초조사를 연구하고자 한다. 본 기초조사는 한국어의 특질에 대한 기존의 국내연구를 조사 분석하고, 아울러 이미 음성인식 및 음성합성 분야에서 연구된 자료들을 수집, 검토함으로써 문제점을 분석하고 향후 음성인식 모델의 개발 연구방향을 기술하고자 한다.

1) 예를 들어 한국어에는 문법적 관계를 설명하는 기능어인 조사가 있어 자유 어순을 허용하고 있다.

\* 제일저자(First Author) : 홍성룡  
접수일 : 2005년 2월 22일, 완료일 : 2004년 2월 27일  
\* 남서울대학교 부교수

\* 이 논문은 2005년도 남서울대학교 교내연구비 지원에 의하여 연구되었음.

한국어 음성인식을 위하여 각 실험의 장단점 분석을 통한 기초자료 수집과 함께 음성인식 시스템들을 비교 분석하여 한국어 음성시스템 개발에 가장 적합한 모델을 찾는 것은 매우 중요한 의미를 가진다.

## 2. 기계번역의 개요

### 2.1 기계번역의 이해

기계번역이란 컴퓨터를 통해서 구문을 파악하고 단어를 대입하여 번역하는 것을 말한다. 기계 번역은 하나의 자연 언어를 동등한 의미를 지닌 다른 자연 언어로 번역하는 것으로 크게 원시 언어(source language)의 이해와 목표 언어(target language)의 생성으로 이루어진다<sup>2)</sup>.

자연 언어 처리(natural language processing)에 대한 필요성은 컴퓨터의 눈부신 발전과 고도화된 정보화 사회를 이룩하려는 사회적 욕구에 따라 점차 증대되고 있다. 그 중에서도 기계 번역은 급증하는 외국과의 정보 교류와 정보의 세계화에 따라 개발이 절실히 요구되는 분야이다[Allen95, Hutchins86].

기계번역은 컴퓨터가 소개된 후 곧 바로 연구의 주요 주제가 되었으나 쉬워 보였던 작업이 여러 가지 한계에 부딪치면서 한동안 관심 밖으로 떨어졌었다. 그러나 수년전부터 컴퓨터의 발달 및 계산 언어학 등의 발달로 다시 연구의 주제로 각광을 받아 오늘에 이르고 있다[Slocum 85, Tucker 87].

기계번역은 하나의 언어로 표현된 입력 문장으로부터 동일한 의미를 갖는 다른 언어로 된 문장을 출력하는 작업을 기계 즉 컴퓨터에 의해 일어나게 하는 것을 말한다. 한 언어로 쓰인 문장에서 뜻하는바 의미를 다른 나라에서 쓰이는 말로 같은 의미를 갖게끔 번역한다는 것은 단순히 문장을 분석하고 그 결과에 대해서 적절한 연어를 선택해서 적절히 나열한다는 것 이상의 의미가 있다. 다시 말하면 한국어 문장들을 영어로 번역하는 경우 우리나라의 말의 쓰임에 대한 배경(즉 문화적 배경)은 잘 안다고 하더라도 미국의 문화적 배경에 대한 이해가 있어야 하고 또 그 문장이 발화된 상황, 화자 및 청자에 대한 사전 지식 등등의 지식이 있어야 정확한 역어 문장을 얻을 수 있을 것이다.

그럼에도 불구하고, 선진국에서는 이미 많은 연구가 이 주제에 대해서 이루어져서, 영역을 제한한다면(비교적 어휘사용에 모호성이 적은 이공계 분야 등) 이미 부분적인 실용화 단계에 접어들고 있다[Kaji 90, Nagao 86]. 이것은 문장 자체의 의미 해석에 모호성이 적은 탓도 있지만 선진 기술을 짧은 시간에 얻고자 하는 필요성도 아울러 고려되기 때문이다.

### 2.2 기계번역의 역사

1950년대의 기계번역의 연구 개발은 자연언어처리 뿐만 아니라 언어 이론들이 미숙한 단계이던 1950년대 말부터 시작되었다. 제한된 컴퓨터 메모리 크기와 컴퓨터 처리 능력 때문에 기계번역 분야는 연구 개발 사항들이 많은 어려움을 겪었다.

1960년대는 컴퓨터의 메모리 자원 및 처리 성능이 빈약할 뿐만 아니라 높은 가격으로 인해 어휘나 문법 정보가 적은 양으로 구축되었고, 개발자들은 여러 시스템들을 토대로 기계번역시스템의 가능성을 부여하는데 초점을 맞추었다.

1970년대는 인공지능 연구자들이 관심 분야 중 한 분야로 자연언어 이해에 많은 관심을 가졌다. 이때 의미정보 분석에 있어서 심도 높은 연구가 이루어졌으며, 우위 의미론(Preferential semantics)이라고 불리는 방법을 제안하였다. 그러나 이들은 그때 당시 시스템에 적용을 하지 못한 이론적 방법으로도 남았다.

1980년대는 유럽과 일본에서 기계번역에 대한 많은 연구가 이루어진 시대이다. 여러 종류의 많은 프로젝트가 시작되었으며, 반도체 기술의 빠른 성장으로 기계번역 개발자를 위한 충분한 컴퓨터 시스템을 제공할 수 있었다. 80년대 중반에는 많은 상업적 기계번역시스템이 출현하였다. 그러나 너무 높은 가격과 번역 품질에 많은 불만을 가지게 되었다.

80년대 말부터 90년대까지 많은 기계번역 개발자들은 공유할 수 있는 언어 자원(LRs : Language Resources)에 대해 관심을 가져왔고 현재 많은 연구기관들에 의해 구축이 진행되고 있다. 앞으로도 다국어 기계번역을 위한 다양한 언어에 대한 자원이 구축되리라 본다.

이렇게 1950년 이후 1970년대를 거쳐 21세기 현재에 이르기 까지 컴퓨터의 발달과 더불어 공학적 측면에서 자연 언어 이해(natural language understanding)를 연구하기 시작했으며 제한된 영역에서 전문가 시스템(expert system) 및 기계번역 시스템(machine translation system)등 많은 연구 발표가 있었다. 최근 들어, 지능형 컴퓨터와 지식 정보화 사회에 대한 비전이 제시되면서, 그 핵심적인 요소 가운데 하나가 바로 자연언어처리라는 인식이 확산되고 있고, 컴퓨터의 비약적인 발달과 더불어 자연언어처리 기술이 크게 향상되어서 실용적 수준의 기계번역 시스템에 대한 시장 전망을 밝게 해주는 요인이 되고 있다.

그러나 현재로는 인간의 개입이 전혀 없는 완벽한 자동 번역은 불가능하며, 인간과의 대화식 방법으로 번역 작업을 고속, 대량으로 처리하는 데 사용된다. 앞으로 지식 공학의 도입에 의한 완전한 자동 번역의 실현을 위해 연구가 요구된다.

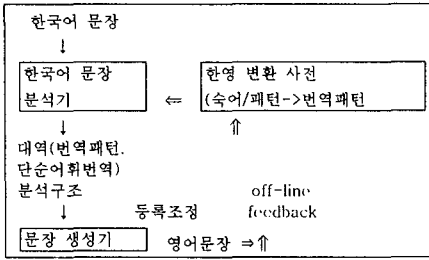
## 3. 기계번역 유형

번역에 영향을 주는 요인은 여러 가지가 있으나 의

2) [Hutchins86]

부의 요인을 크게 생각하지 않을 때 번역의 기본단위는 문장이다. 따라서 가장 간단한 형태의 기계번역방법은 원시언어(source language) 문장과 목적 언어(target language) 문장을 대응시켜 원시언어 문장이 입력됐을 때 목적 언어 문장을 출력하면 된다.

표 1. 한국어 기계번역 패턴



3.1 직접 방식

형태소 분석정도의 이해 과정을 통해 번역이 이뤄지고 번역은 형태소를 번역한 후에 번역된 형태소를 목적 언어의 구성에 적합하게 재배열하는 방식이다. 한-일 또는 영-불과 같은 동일어족에 속한 언어 번역에 적용한다.

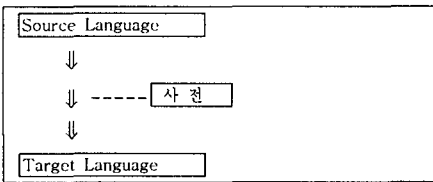


그림 1. 직접입력 방식

3.2 변환 방식

구문분석 정도의 이해과정으로 번역이 이뤄지며 약간의 의미 분석도 이뤄지는 방식이다. 원시언어 문장이 분석되면 일반적으로 트리(tree) 형태로 표현되고 트리는 목적 언어의 구조에 적합한 트리로 변환된다. 이렇게 변환된 수형도로부터 목적 언어의 문장을 생성한다.

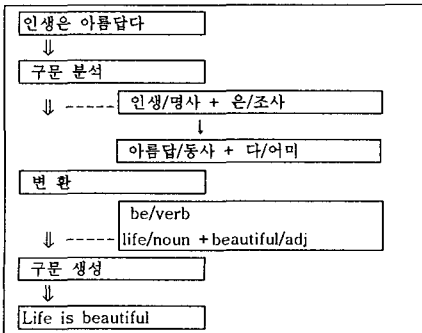


그림 2. 변환 방식

3.3 피봇(Pivot) 방식

중간언어 방식은 의미 분석 정도의 이해과정으로 번역이 이뤄진다. 즉 원시언어 문장이 분석되면 중간언어로 표현되는 것이다. 중간언어는 어떤 언어에도 의존하지 않는 의미구조 혹은 개념구조이며 주로 트리 형태로 표현한다. 목적 언어의 문장생성은 중간언어로부터 원시언어와 무관하게 이뤄지며 원시언어의 분석이나 목적 언어의 생성은 완전 독립적이다.

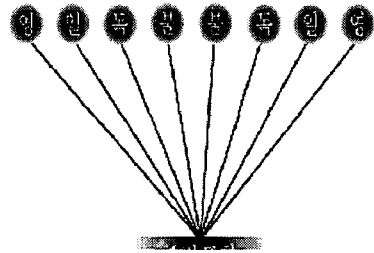


그림 3. 다중 언어간 번역

중간 언어(interlingual) 방식이라고도 하는데, 원시언어와 목표언어에 대해 공통적인 하나의 중립적인 이른바 피봇 표현을 상정하고, 크게 분석과 생성의 두 단계만을 거쳐 번역이 이루어진다. 변화 단계가 없기 때문에 완전한 분석이 이루어지며, 목표언어의 특성에만 의존하여 생성이 이루어진다.

4. 기계번역의 단계

자연어 문장을 처리하기 위해서 형태소를 해석하고 이를 토대로 문장의 구조를 해석하여 전체 문장의 의미를 파악하는 세 가지의 단계를 거쳐야 한다.

4.1 형태소 분석

자연어의 입력 문자열로부터 문장의 해석에 필요한 단위 문자열로 잘라서 입력 문장에 대한 각종 지식 정보, 즉 품사의 범주, 활용형 등 각종 정보를 추출해 내는 과정을 형태소 분석이라고 한다.

형태소 분석기는 사전을 검색하여 어절을 구성하는 단어들의 성분 등을 파악하는 것이 제일 먼저 이루어져야 하며 이를 파악한 후 다른 처리를 행한다. 그러므로 사전이 가져야 하는 정보에 있어서 형태소 분석에 필요한 품사, 활용 규칙 등이 필수적이다.

형태소 분석기는 한 문장에서 가능한 모든 분석 결과를 출력할 수 있어야 한다. 다음은 모호한 결과를 가지는 문이나 단어에 대한 모든 결과를 출력할 수 있도록 처리하는 예와 사전의 구성을 보인다.

한국어는 영어와 달리 단어 단위의 분리가 일정하지 않고, 명사에 결합되는 조사의 종류가 다양하며 용언의 활용이 복잡해서 형태소의 분석이 매우 중요하다. 영어의 경우 단어 구분이 분명하고 어형 변화를 나타내는

표지<sup>3)</sup>는 비교적 단조롭게 정형화되어 있지만, 한국어의 경우 '나는'이라는 어절은 ['나'(대명사)+'는'(조사)], ['날(fly)+'는'(어미)], ['나다'(be born)+'는'(어미)] 등 여러 가지로 해석될 수 있다.

형태소 분석을 위해선 다음과 같은 네 가지의 단위 모듈(module)이 필요하다.

- ① 언어적 최소의 단위인 형태소 단위로 사전을 구성
- ② 어절을 이루는 규칙성을 형태소 분석 문법으로 작성
- ③ 사전을 참조하거나 문법 검사를 통해 대상 어절을 형태소 단위로 분석
- ④ 형태소 단위의 사전 정보로부터 어절에 대한 언어적 정보 합성

형태소 분석 단계는 입력 문장을 받아 최소의 의미 단위인 형태소로 분리하는 과정으로서, 분리된 형태소를 사전 내에서 검색하여 품사 정보를 얻고 이웃하는 단어와의 접속 관계가 문법적으로 올바르게 되었는지를 점검한다.

## 4.2 구문 분석

구문 분석은 형태소 분석의 결과로 얻어진 결과를 이용하여 문장 내의 성분들이 문법적으로 어떻게 결합되는지를 규명한다.<sup>4)</sup>

구문 분석은 입력된 문장에 대해서 형태소를 분석한 결과로부터 문법 규칙 및 어휘 정보를 조사하여 문장의 구조를 해석하는 과정이다. 예를 들어 '철수는 영화를 좋아한다.' 라는 문장의 구문 해석의 결과는 다음과 같다.

S[NP[N(철수),JO(는)], VP[NP[N(영화), JO(를)],V(좋아한다)]]

구문 분석에 필요한 문법 규칙은 적용하는 문법 이론에 따라 다를 수 있으나, 단일화 기반 문법 이론을 기반으로 설명한다면, 문법 규칙은 문의 중심어 성분이 취하는 하위범주화 자질로서 설명될 수 있다[8],[9].

이제까지의 구문 해석은 주로 의존 문법을 기반으로 했었지만, 최근에는 중의성<sup>5)</sup> 해결의 문제가 초점이 되고 있다. 또한 한국어는 어절 자체가 구문 정보를 가지고 있는 형태이기 때문에 어절 간의 관계를 파악하는 것은 구문 구조간의 관계를 밝히는 것이라고 볼 수 있다.

## 4.2 의미 해석

단어 또는 문장이 나타내고자 하는 의미를 의미자질(semantic feature), 격(case), 의미표현(semantic representation), 논리(logic) 등의 수단으로 표현하는 것이다. 기계 번역에서 의미 해석은 구문 분석은 되나 의

3) 동사의 사제: -ed, -ing, 명사의 복수형: -s, -es

4) [서영훈 91, 김호영 90]

5) 기계번역이나 자연언어 처리에서 문장을 처리할 때 제일 문제가 되는 것은 해석의 결과가 유일하지 않아서 생기는 모호함이다.

미 해석이 안 되는 문장, 구문 분석의 결과가 모호한 문장일 경우에 중요하다. 예를 들어 '그가 배를 샀다'라는 문장에서 '배'의 의미가 '먹는 배'인지 '타는 배'인지 의미가 모호할 경우 앞 뒤 문장을 살펴보아야 한다.

## 5. 음성 인식과 음성 합성

자동음성인식과 음성합성의 목적은 자연스럽게 발화하는 음성신호를 컴퓨터가 인식하여 주어진 음성입력을 수용하고 나아가 컴퓨터로 하여금 주어진 발화를 적절한 음성신호로 내보냄으로써 인간과 기계와의 자연스러운 의사소통을 가능하게 하고자 하는데 있다.

### 5.1 음성 인식 연구의 문제점

음성인식과 음성합성 등의 자연언어처리의 연구는 단기적으로는 은행잔고조회, 영화시간안내, 전화통화 전환 또는 정보검색기능에 응용할 수 있을 뿐만 아니라, 음성을 이용한 정보입력처리에도 응용될 수 있다. 예를 들어, 의료진단이나 암실에서의 응용, 또는 휠체어를 이용하는 장애인이나 로봇의 음성보조기구로 이용 개발될 수 있다. 그리고 인공지능건물이나 인공지능 자동차와 같은 최첨단 장비에도 이용될 수가 있다.

1940년대 중반 스펙트로그램의 개발과 더불어 시작된 음성인식의 연구는 특정화자의 화자의존적인 소량의 어휘들에 대한 고품단어 음성인식에 있어서는 어느 정도 성공을 보이고 있으나, 다 용량 어휘의 불특정 화자, 그리고 연속음성시의 자연스런 음성인식의 수준에는 아직 크게 못 미치고 있는 형편이다. 이와 같이 구어체 연속음성인식에서 직면하고 있는 문제들을 연구의 차원에 따라 분류하면 다음 6가지로 나눌 수 있다.

첫째, 고품, 연결 및 연속음성인식의 문제이다. 고품 단어와는 달리 연속어의 경우에는 동시 조음(coarticulation) 등 음운규칙의 적용으로 단어경계를 검출하기가 어려울 뿐 아니라, 혼동하기 쉬운 단어나 구로 오인될 소지가 있다.

둘째, 어휘 량의 문제이다. 소량의 어휘와는 달리 어휘가 증가함에 따라 어휘간의 혼동이 증가할 뿐만 아니라 정확도가 감소하게 된다. 아울러 충분한 training data의 확보도 어렵게 된다.

셋째, Task와 언어제약의 문제이다. 주어진 용도에 따라 통사론과 의미론의 언어 모델을 제한할 필요가 있게 되는데, 이때 주의할 점은 입력형태 역시 크게 제한받게 될 수 있으므로 언어 모델은 최대한 제한하면서 입력형태는 최대한 자유를 부여하여야 한다.

넷째, 화자독립과 화자종속의 문제이다. 특정 화자를 위한 화자종속시스템은 한 화자에 대하여 모델계수를 결정하므로 비교적 높은 인식률을 보여주지만, 새로운 화자의 불편을 겪게 되는 단점이 있다. 이와는 반대로 불특정화자를 위한 화자독립시스템은 다양한 화자의 음성을 모델로 하기 때문에 낮은 정확도를 보이고 있다.

다섯째, 음향학적 모호성과 혼동의 문제이다. 어휘간의 모호성과 혼동가능의 정도에 따라 인식의 정도가 좌우되는데, 이를 위해 고성능의 음향패턴해석방식을 사용할 필요가 있다.

마지막 여섯 번째는 주변 소음의 문제이다. 예를 들어, 공장 내의 소음, 에어컨 소리, 비행기 조정석에서의 엔진소리, 여러 사람이 동시에 말할 때의 혼신, 그리고 기침, 재채기와 같이 화자에서 발생하는 잡음을 들 수 있다.

## 5.2 음성 인식 연구 모델 비교

한국어 구어체 음성인식을 위한 연속음성인식시스템 개발의 문제는 지금까지의 화자의존(Speaker Dependent), 제한된 단어 수, 고립단어의 음성인식과는 달리 화자독립, 무한단어, 연속음성인식을 위한 것으로써 기존의 음성인식방식과는 달리 보다 장기적인 계획이 설정되어야 한다.

음성 인식의 가능한 방법을 조사하면 다음과 같다.

패턴 비교방식(Template-based Approach)은 주어진 음성입력을 미리 입력된 표준패턴과 비교하여 가장 유사한 것을 채택하는 음성인식 방식으로써 제한된 단어에는 비교적 좋은 결과를 보여주고 있기 때문에 간단한 응용 분야에서 성공적으로 이용되고 있다. 거리계산에 대한 척도(distance metric)가 필요하며 또한 시간적인 변화를 보완해주는 DTW(Dynamic Time Warping) 등의 알고리즘이 요구된다.

전문가 인식시스템은 음양음성학 지식과 음운지식을 바탕으로 한 전문가 인식시스템은 화자 독립적, 연속음성인식에 적합한 방식으로 간주되고 있다. 대표적인 것으로 일본의 SPREX(A Speech Recognition Expert)와 V. Zue를 중심으로 한 MIT방식을 들 수 있다.

통계를 이용한 음성인식방식은 HMM(Hidden Markov Model)과 같은 통계적 모델을 음성인식을 하는 방식으로 IBM에서 주로 도입하고 있는 방식이다. 현재까지 알려진 대용량시스템으로서는 가장 성공적인 것으로 간주되고 있다. 음성분할(segmentation)과 음성인식의 문제를 동시에 해결해 주고 있기 때문에 연속인식 특히 적합하다고 할 수 있다.

신경회로망 방식(Connectionst Approach)은 최근에 대두되고 있는 음성인식방법으로 인공신경망을 이용하여 많은 간단한 연결마디들에 음성특징들을 분산 분포시켜 음성인식을 실현하자 하는 음성인식 방법이다.

## 5.3 음성 인식 모델 연구 방향

한국어 구어체를 위한 연속 음성인식 시스템에서는 이상과 같은 몇 가지 음성인식 모델의 장단점들을 비교 검토하여 한국어의 특질에 맞는 음성인식모델을 개발하여야 한다. 한편 연속음성인식을 위한 가능한 모델로 전문가 인식시스템을 도입할 경우, 한국어에 대한 음성 데이터베이스를 구축해야 할 필요가 있다.

음성인식의 첫 번째 기본 단계는 일반적으로 음소인

식을 위한 특징자질 추출에 있다. 따라서 음향 음성적으로 또는 컴퓨터에서 인식 가능한 자질들의 특징적 변수들을 산출해 내는 실험과 함께, 이렇게 하여 산출된 자질들을 음성합성실험을 통한 음향심리실험을 거침으로 한국어의 특징적인 불변의 자질들을 추출하는 일을 제 우선적으로 수행하여야 한다. 아울러 문맥에 따라 변이되는 변이의 폭과 기타 가능한 모든 변수들을 추출하는 실험 역시 이와 병행되어 수행되어야 한다.

이러한 음성인식의 실험에는 기존의 주파수 분석, 선형예측분석, 상관관계분석 뿐만 아니라, pitch, formant, voicing등 음운자질의 유무개념을 이용한 자질인식 실험을 통하여 한국어 음소의 모든 자질들을 실험할 필요가 있다. 이때 음운적으로 주요한 이론들인 자질수형도 이론이나 잠재표기이론 그리고 유표성 이론과 같은 음운 이론을 도입해야 한다. 즉 음운 자질들 중에서도 핵심적인 인식의 자질들을 그룹별로 주요 자질과 이차적 자질 등으로 나누어 추출해야 할 것이며 이와 더불어 어떤 자질들이 잉여적인가라는 문제도 함께 검토되어야 한다. 한편 잠재표기이론과 유표성 이론을 이용한 자질수형도이론은 여러발생시 원위치로 찾아가는 여러 검색 통로를 자리위에 따라 제한시켜줌으로서 여러 검색시 소요되는 시간을 절약시켜줄 수 있다.

## 6. 결론 및 제언

본 연구가 음성인식 및 합성의 기본 알고리즘을 설정하거나 이미 설정된 것을 개량하는데 충분한 이론적 바탕이 될 수 있을 것이다. 또 실제 대화상황, 화자의 심리 등에 따른 음성, 음운적 변이요소들을 고려하여 실험적 결과를 얻은 후, 아직까지 확립되지 않은 한국어의 발음기호를 확립하는 기본모체로도 사용될 수 있을 것으로 기대된다.

한국어 자모음에 대한 불변의 음성자질들과 문맥에 따라 변하는 변이음적 자질들의 확인은 음성인식 실험과 더불어 추출된 자질들을 음성합성을 이용한 음향심리실험을 거침으로써 온전히 검증될 수 있다고 보기 때문에 음성인식의 실험과 음성합성의 실험은 함께 병행하여 진행되어야 한다. 따라서 한국어에 대한 기존의 음성지식의 검증에는 성별, 연령별, 지역별 특성을 고려하여 이에 대한 불변의 고유자질과 변이자질들을 음향 음성실험을 통해 검증해 나가야 하며, 아울러 한국어 자모음에 대해 추출된 모든 자질들을 음성합성의 음향 음성 심리실험을 통하여 검증해나가는 일을 병행해야 한다.

6) 조음방식 자질(sonorant, fricative, stop, nasal 등), 조음 위치자질, 유성 자질 등

참 고 문 헌

[1] [서병락96] 서병락, 김영택, “한영기계 번역을 위한 번역 패턴에 기반한 영어 문장 생성기”, 정보과학회논문지, 제 23권, 5호, pp. 520-529, 1996

[2] [옥철영93] 옥철영, 한영 기계번역을 위한 구 단위 변환 사진, 서울대학교 컴퓨터공학과 박사학위 논문, 1993

[3] [이익섭83] 이익섭, 임홍빈, 국어 문법론, 학연사, 1983

[4] [이정민87] 이정민, 배영남, 언어학 사진, 박영사, 1987

[5] 박성익·강명희·김동식(1998). 교육공학 연구의 최근동향. 서울: 교육과학사.

[6] [이하규94] 이하규, 숙어 분산 특성을 이용한 한영 번역 숙어의 인식, 서울대 학교 컴퓨터공학과 공학박사 학위논문, 1994

[7] [이호석93] 이호석, 영한 기계번역을 위한 말뭉치에 기반한 변환사전의 자동구축, 서울대학교 컴퓨터공학과 박사학위 논문, 1993

[8] [장석진93] 장석진, 정보기반 한국어 문법, 언어와 정보, 1993

[9] [Barnett91] J. Barnett, I. Mani, E. Rich, C. Aone, K. Knight, J. C. Martinez, "Resolving Translation Mismatches and Divergences in an Interlingua-Based MT System," International Workshop on Fundamental Research for the Future Generation of Natural Language Processing, pp. 127-146, 1991.

[10] [Bateman90] J. A. Bateman, "Finding Translation Equivalents : An Application of Grammatical Metaphor," Proceedings of COLING-90, vol. 2, pp. 13-18, 1990.

[11] [Dorr90] B. Dorr, "Solving thematic Divergences in Machine Translation," Proceedings of the 28th Annual Meeting of the ACL, pp. 127-134, 1990.

[12] [Harbusch94] K. Harbusch, G. Kikui, A. Kilger, "Default Handling in Incremental Generation," Proceedings of COLING-94, pp. 356-362, 1994.

[13] [Huettner87] A. K. Huettner, M. M. Vaughan, D. D. McDonald, "Constraints on the Generation of Adjunct Clauses," Proceedings of the 25th Annual Meeting of the ACL, pp.207-214, 1987.

[14] [Kikui92] G. Kikui, "Feature Structure Based Semantic Head Driven Generation," Proceedings of COLING-92, vol. 1, pp. 32-38, 1992.

[15] [Kim95] N. Kim, Y. T. Kim, "Pattern from large Corpora for Korean-English Machine Translation," NL PRS'95, pp. 383-387, 1995

홍성룡



1983년 충남대학교 영어영문학과 (학사)  
 1985년 경희대학교 영어영문학과 (석사)  
 1993년 경희대학교 영어영문학과 (박사)  
 1995년 경기도청 전문위원(국제협력)  
 1996년 신성대학 전임강사  
 1997년 ~ 현재 남서울대학교 부교수  
 관심분야 : 기계번역, 원격교육, 매체제작