

한국어-수화 번역 시스템 설계

권 경 혁[†] · 우 요 섭^{††} · 민 흥 기^{†††}

요 약

본 논문에서는 청각장애인의 문자 습득과 진검인과의 대화 수단으로 사용할 수 있는 한국어-수화 번역 시스템을 설계하고 구현하였다. 번역방식은 직접분석 기계번역을 사용하였으며 이 과정에서 필요한 수화 사진들을 정의하였고, 이를 이용하여 한국어 문장을 수화 동영상으로 변환하였다. 수화 사진은 기본 수화 기호 사진, 수화 유사어 사진, 수화 복합어 사진의 세 개로 구성하였다. 기본 수화 기호 사진은 한국 수화의 기본적인 기호와 수화 동영상을 제공하며 수화 복합어 사진은 기본 수화 기호로 구성되는 복합 수화에 대한 경의만을 제공하였다. 또한 수화 유사어 사진은 동일한 의미로 사용되는 수화들에 대한 관련성을 제공하였다. 수화 사진에서 검색된 수화 기호에 대한 동영상의 표시는 GIF 형식의 연속 수화 동작 재생과 한국어 코드 분석에 의한 지문지 표시로서 화면에 출력하였다. 설계한 한국어-수화 번역 시스템은 한국 수화의 특성을 반영한 사전 구조를 사용하여 번역 과정에서 빠른 수화 검색이 가능하며, 수화 단어의 부족함도 해결될 수 있다. 또한, 공간적인 수화의 표현에 동영상어 아닌 그림 파일의 연속 재생 기법을 사용하여 수화 사진의 저장 공간을 절약할 수 있었다.

Design and Implementation of a Korean Text to Sign Language Translation System

Kyung-Hyuk Kwon[†] · Yo-Seop Woo^{††} · Hong-Ki Min^{†††}

ABSTRACT

In this paper, a korean text to sign language translation system is designed and implemented for the hearing impaired people to learn letters and to have a conversation with normal people. We adopt the direct method for machine translation which uses morphological analysis and the dictionary search. And we define the necessary sign language dictionaries. Based on this process, the system translate korean sentences to sign language moving picture. The proposed dictionaries are composed of the basic sign language dictionary, the compound sign language dictionary, and the resemble sign language dictionary. The basic sign language dictionary includes basic symbols and moving pictures of korean sign language. The compound sign language dictionary is composed of key-words of basic sign language. In addition, we offered the similar letters at the resemble sign language dictionary. The moving pictures of searched sign symbols are displayed on a screen in GIF formats by continuous motion of sign symbols or represented by the finger spelling based on the korean code analysis. The proposed system can provide quick sign language search and complement the lack of sign languages in the translation process by using the various sign language dictionaries which are characterized as korean sign language. In addition, to represent the sign language using GIF makes it possible to save the storage space of the sign language dictionary.

※ 본 연구는 인천대학교 교내연구비의 멀티미디어 연구센터의 인부 지원에 의하여 수행되었음.
† 준 회원 인천대학교 대학원 정보통신공학과
†† 정 회원 인천대학교 정보통신공학과 교수
††† 종신회원 인천대학교 정보통신공학과 교수
논문집수 1999년 12월 1일, 심사완료 · 2000년 2월 23일

1. 서 론

우리에게 언어란 생활 그 자체이다. ASHA(American Speech-Language Hearing Association)에 따르면 통신(communication)은 인간 생활의 핵심이라고 했다[1]. 언어는 대표적인 의사소통 수단으로서 인간의 사상과 감정을 표현하는 매개체이다. 따라서 말과 글이 없는 인간 사회를 생각하기란 불가능한 일이다. 즉 말이란 풍요로운 삶을 이루는 기본 요건이라 할 수 있다.

청각장애란 청각 능력이 없거나 손상되었음을 지칭하는 용어로 청각을 통하여 언어적 정보를 처리하는데 장애가 되는 상태를 의미한다[3, 4]. 청각장애의 발생이 음성언어를 습득한 후 발생한 후천성 청각장애인은 단지 듣기만 못하므로 구화 등의 방법을 사용하면 건청인과의 의사소통이 가능하다. 그러나 청각장애의 발생이 유아기 이전에 나타나는 청각장애인의 대부분은 음성언어를 습득할 기회를 상실하여 언어장애까지 수반하는 복합장애인이다. 따라서 정보의 습득과 전달이 어려워 사회 참여에 많은 제약은 받게 된다.

청각장애인들이 사용하는 의사소통 수단으로는 음성언어와 비음성 언어가 있다[4]. 음성언어로는 잔존 청력과 독화를 사용하는 구화가 있다. 구화는 건청인과의 의사소통이 용이하기는 하나 교육이나 직업 등 수많은 다양한 정보를 얻기에는 한계가 있다. 비음성 언어로는 수화, 필담 등이 있다. 수화는 청각장애인이 자신의 생각을 표현하는 수단으로 많이 사용하고 있으나 일부 사람들에게만 적용되는 제한적인 언어이다. 필담은 가장 일반적인 의사소통의 수단으로 건청인과 청각장애인의 의사소통을 할 때 많이 사용되고 있으나, 청각장애인중에는 문장력이 부족한 사람이 많아 의사소통이 잘 되지 않을 수도 있다.

읽기는 글을 이해하는 수단이며, 쓰기는 글을 표현하는 수단이다. 글을 이해하고 표현하기 위해서는 내적 언어 능력과 사고가 발달해야 하는데 청각장애인은 읽기를 위한 준비성과 감각이 결여되며, 문자를 사용하기도 어렵다. 많은 청각장애인들은 표준화 읽기 검사 결과, 미국의 경우 영어 읽기는 대략 초등학교 3.4학년 정도인 것으로 나타났으며 글쓰기에서는 단어와 구조상의 오류와 보어나 관계사절이 포함된 복잡한 구조의 문장을 구사하는데 어려움을 보였다[8].

오랜 기간 동안 청각장애는 병적인 상태라고 생각되어 왔으며 이러한 관점에서 청각장애인의 문자 능력 부

족을 설명하였다. 그러나 최근에는 언어식 문제와 읽고 쓰는 문제가 청각장애인들이 문자 능력 부족을 가지오는 원인이리고 설명된다[9]. 이러한 원인은 건청인과 청각장애인 사이에서 발생하는 언어학상의 차이와 교육적 차이로 해석된다. 언어학상의 차이란 청각장애인에게 음성언어는 들어 본 적이 없는 언어이거나 제한된 방법으로만 들었던 언어라는 것이다. 한글을 음성언어로 접할 수 없기 때문에 건청인들 보다 읽고 쓰는 능력을 획득하는데 어려움을 겪는다. 청각장애인들이 문자를 습득하기 어려운 또 다른 원인은 교육적인 차이이다. 최근에 발표되고 있는 청각장애인들을 위한 문자 교육 방법들은 문자언어와 수화를 통합적으로 교육할 수 있는 두 개의 언어 혹은 문화의 동시 접근방법(bilingual/bicultural access), 수화 비디오와 문자언어를 컴퓨터화하여 동시에 교육하는 상호작용 비디오 디스크 등이 있다[8]. 문자가 나오는 폐쇄 문자 TV 프로그램(closed captioned TV programs)은 미국의 경우 TV 디코더법(television decoder circuitry act)을 제정하여 13인치 이상의 모든 TV에 폐쇄자막을 표시하는 디코더 내장을 의무화하였다[10]. 이런 방법들의 공통점은 언어와 문자를 동시에 교육시킨다는 것이다. 즉 문자언어를 습득할 때 시각과 청각 모두를 활용하는 교육 방법이 높은 학습 효과를 나타내며 말과 문자가 연결되어야 쉽게 이해 할 수 있다는 것이다.

수화는 얼굴의 표정과 손을 이용하여 어떠한 모습 및 특정하게 정해 놓은 동작을 표현 언어의 대체 수단이다. 수화는 인류가 이 지구상에 태어났을 때부터 시작되었다고 생각되며 그때의 수화는 자연발생적인 손짓, 몸짓이었다. 수화가 우리 나라 정규 학교에 도입된 것은 Rosetta Sherwood Hall이 1909년 평양에 농아학교를 설립한 후 이다. 한국 수화의 정립은 1947년 국립 맹아학교 초대 교장인 윤백원에 의하여 한국 수화 지문 자기 만들어지면서 그 형태를 갖추기 시작하였고, 1982년 서울 농아학교 교직원들이 만든 약 6,000여 단어가 수록된 표준 수화 사진이 발간되어 수화 보급에 획기적인 기틀이 마련되었다[3, 4]. 수화는 청각장애인들에게 의사 표현 또는 수용의 중요한 수단으로 사용된다. 즉 수화는 청각장애인들의 의사소통에 가장 적합할 뿐만 아니라 청각장애인들이 활용하는 가장 적당한 의사 표현 방법이다[6]. 청각장애인의 정확한 의사소통을 위해 수화의 필요성이 강조됨에 따라 지금은 통합적인 의사소통 방법이 강조되고 있다.

현재 한국에서 사용되고 있는 수화는 방법적 유형에 따라 한국어 문법과 일치하게 표현되는 문법적 수화와 청각장애인 사회에서 자연발생적으로 생성된 자연 수화로 크게 구분된다[3,5]. 자연 수화는 청각장애인에 의해 창조되고 발전된 일종의 고립어이며, 문법적 수화는 자연 수화의 기호, 지문자(指文字), 지숫자(指數字)를 사용하여 한국어의 문법에 맞게 체계화된 수화이다. 청각장애인 사회에서 문법적 수화보다 자연 수화가 더 넓게 사용되고 있고 또한 청각장애인은 음성을 들음으로써 습득되는 음성언어의 문법을 학습하기가 매우 어려우므로 본 연구에서는 자연 수화를 대상으로 진행되었다[7]

자연 수화의 기호는 어원에 따라 지사(指事), 모방(模倣), 상형(象形), 형지(形指), 형동(形動), 회의(會意), 전주(障注)로 구분된다[5]. 이러한 자연 수화 기호의 분류에서 지사, 모방, 상형, 형지, 형동 등 다섯 가지는 하나의 형태소로만 형성된 단순형태소사(單純形態素詞)이며 회의는 둘 이상의 형태소로 형성된 복합형태소사(複合形態素詞), 그리고 수화운용방법에 따라 전주로 구성되어 있음을 알 수 있다[5]. 이러한 한국 수화의 기호는 대략 6000여개 정도로 한국어 자연 언어에 비하면 그 어휘 수가 현저하게 적다[7] 따라서 한국어 문장을 수화로 번역할 때 모든 한국어 어휘와 수화 기호가 일대일로 대응되는 표현이 불가능하다.

이러한 수화를 한국어로 번역하는 연구가 최근 국내에서도 일부 보고되고 있지만, 주로 화상인식의 관점에서의 접근이 주류이고, 수화가 언어라는 측면에서의 접근은 아직 미진한 실정이다[7, 19]. 따라서 본 논문에서는 경험적 방법에 의한 수화사전을 구성하여 수화를 한국어로 번역하는 시스템을 구현하였다. 즉 문자 학습 환경이 열악한 청각장애인에게 수화와 문자를 동시에 교육할 수 있고, 또한 수화와 문자를 매개체로 하여 일반인과 대화가 가능한 보조통신장치로서 문자를 수화로 바꾸는 한국어-수화 번역 시스템의 기본구조를 설계하고 그 타당성을 입증하기 위해 시스템을 구현하였다. 즉 한국어를 수화영상으로 변환하는 과정을 기계 번역의 관점에서 진행하였으며, 이 과정에서 필요한 세 수화 변환 사전들을 구성하고 이러한 사전들을 이용하여 한국어-수화 번역시에 발생하는 부족한 수화 어휘의 문제를 해결하였다. 수화 변환 사전은 수화의 특성에 따라 세 가지로 나누어 구성하였다. 즉 기본적인 수화기호의 영상 정보를 저장하는 기본 수화 기호 사전,

기본 수화 기호가 절할되어 새로운 의미가 되는 복합 수화를 저장하는 수화 복합어 사전, 그리고 유사한 의미를 가진 한국어 단어에 대해 동일한 동작으로 표현되는 수화들을 저장하는 수화 유사어 사전으로 나누어 구성하였다 또한 이 사전들을 이용하여 한국어 문장을 수화 동영상으로 변환하기 위한 알고리즘을 제안하고 이를 설계 및 구현하였다.

2. 한국어-수화 번역 시스템의 설계

번역시스템을 형태소 분석과 사전검색의 과정을 거치는 직접방식의 기계번역으로 설계하고, 번역의 과정에서 필요한 수화 사전의 종류를 구분하여 구성하고 형식을 지정한다 또한 구성된 수화사전들을 효율적으로 검색하기 위한 알고리즘을 개발한다.

2.1 고려 사항

컴퓨터를 사용하여 청각장애인에게 수화와 문자를 동시에 교육시킬 수 있고 또한 수화를 모르는 건청인이 문자를 입력함으로써 청각장애인에게 수화로 표시되는 통신 보조 기기의 개발하기 위해서는 다음과 같은 사항을 고려해야 한다

한국어 문장을 수화로 번역하기 위해서는 우선 한국어 문장에 대한 자연 언어 처리가 필요하다. 번역에 관련된 자연 언어 처리는 형태소 분석, 구문 분석, 사전 구성 등 원시언어의 해석과정과 목적언어로의 변환 및 생성 과정 등 많은 작업을 수행해야 한다[7, 12]. 형태소 분석은 한국어 문장의 분석에 필수적인 자연 언어 처리의 해석 과정이다. 본 연구에서는 형태소 분석에 관련된 내용은 이미 발표된 연구결과를 사용하였다[17] 즉 본 연구에서는 전체 번역시스템의 설계와 형태소 분석 라이브러리를 사용한 사용자 인터페이스 부분 그리고 수화 사전의 구성에 제한하였다

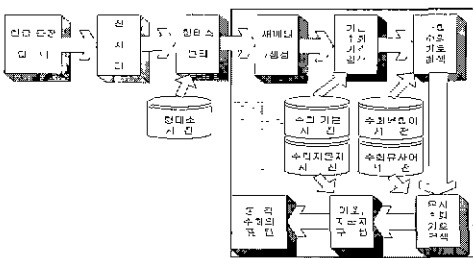
또한 언어의 다양성 문제도 주요한 고려사항이다. 어떠한 종류의 한국어 문장이 입력되어도 이를 수화로 변환시키는 것이 이상적인 한국어-수화 번역 시스템이다. 이를 해결하기 위한 방법은 우리가 말하는 모든 단어를 수화로 바꾸어 저장하고 또한 효과적인 탐색 방법을 개발하는 것이다 그러나 국어사전에 있는 모든 단어들을 수화로 바꾸어 컴퓨터 저장 장치에 저장한다는 것은 경제적으로나 시간적으로 비효율적이다 따라서 본 연구에서는 일반인이 일기 쉬운 방법으로

수화 사전을 구성하고 필요할 때마다 단어의 입력, 수정이 가능하도록 사전의 사용자 인터페이스를 융통성 있게 구성하였다

수화는 손짓과 표정으로 표현하는 3차원 언어이다 따라서 수화는 시간적이고 공간적인 표현이 가능하여야 의미가 정확하게 전달되는 경우가 대부분이다. 컴퓨터에서 이 문제를 해결하는 방법은 수화의 표현을 AVI, MOV, MPEG 형태의 동영상으로 출력하는 것이다. 그러나 이러한 형식의 동영상들은 많은 프레임 수를 사용하므로 저장 공간이 매우 크며, 수화 영상의 제작에도 많은 노력과 비용을 필요로 한다. 따라서 제작하고자 하는 시스템에서는 사용자가 수화를 이해 가능한 정도로 동영상의 프레임 수를 줄여야 하고 수화의 영상과 영상의 연결이 쉽게 구현되며, 또한 쉽게 제작 가능한 형태를 사용해야 할 것이다. 이 요구 사항에 만족하는 형태가 최근 인터넷 등에서 많이 사용되는 GIF (Graphics Interchange Format) 형태의 그림 파일이다[13].

2.2 한국어-수화 번역 시스템의 구조

원시언어인 한국어의 번역된 수화는 자연수화이다. 자연 수화는 자체 언어적인 특성이 한국어와는 다르므로 한국어 문장을 자연 수화로 변환하는 과정은 형태소 분석과 사전 검색 과정을 거치는 직접 방식을 선택하였다[7] 직접 방식의 기계 번역 구조를 갖는 한국어-수화 번역 시스템을 (그림 1)에 나타냈다



(그림 1) 한국어-수화 번역 시스템의 구조

(그림 1)에서 제시한 한국어-수화 번역 시스템의 구조는 입력된 한국어 문장의 형태소를 분석하는 형태소 분석기, 분석된 형태소를 적절한 수화로 바꾸어 주는 재배열/생성 모듈과 기본 수화 기호 사전, 수화 복합어 사전, 수화 유사어 사전, 수화 지문자를 저장하는 수화

데이터베이스 및 사용자에게 표시되는 인터페이스로 구성된다.

번역의 과정은 한국어 문장이 입력되면 먼저 형태소 분석기를 통해서 문장을 이근과 조사 등 기본단위로 분석한다. 한국어와 한국 수화는 기본적인 이순은 같지만 조사 및 접속사의 사용, 어순의 도치, 시제의 표현 등 표현의 차이가 있으므로 재배열/생성 단계에서 한국어와 수화 표현의 차이에서 발생하는 구성의 변경 작업을 처리한다. 재배열/생성 단계의 출력은 목적 언어인 수화로 표현해야 할 단어들로서 사전 검색의 키워드이다. 최종적으로 재배열된 단어들은 사전 검색을 통해 번역하고자 하는 목적 언어인 수화의 동작으로 표현된다.

2.3 수화 사전의 구성

한국어 문장을 수화로 번역하려면 수화 사전이 필요하다. 한국 수화 사전의 구성시 수화 기호는 어원에 따른 분류보다는 구조와 형태에 따라 분류하여야 한다. 한국 수화의 특징중 한가지는 많은 복합 수화의 사용과 유사 수화의 사용이다. 복합 수화의 사용이란 한국 수화의 단어 6000여 개 중에서도 기본 수화 기호 동작들을 복합하여 표현되는 수화 기호가 다수 존재한다는 것이다. 즉, 둘 이상의 기본 기호가 합하여 다른 의미를 나타내는 기호가 된다. 예를 들어 '일요일'의 수화 표현은 <빨강다> + <문어 닫히다, 폐교하다> + <날> 이 된다. 이러한 기본 수화 기호의 조합으로 생성한 수화 기호는 이미 다수 존재하지만 계속해서 새롭게 만들어 질 수 있으며 수화만이 가지는 언어의 융통성을 제공해 주고 수화 단어의 부족성도 해결하는 방법이다.

수화사전 구축시 복합 수화 기호를 기본 수화 기호와 동일한 형태로 구성한다면 계속해서 생성되는 복합 수화 기호들을 수용할 수 없다 즉 수화 복합어, 사전은 기본 수화 기호의 조합으로만 표현되도록 별도로 구성하여야 한다 그리고 수화 복합어의 추가, 삭제가 가능하도록 시스템을 구성한다면 번역 시스템에 확장성을 제공하며 수화 단어의 부족함도 해결할 수 있다.

또한 한국 수화의 또 다른 특징은 유사 수화의 사용이다. 즉 유사한 의미를 가진 한국어 단어에 대해 동일한 동작으로 그 의미를 표현하는 경우가 다수 존재한다. 이것은 단어가 의미를 수화 기호로 바꾸는데 한계가 있기 때문에 기존의 수화로도 의미 전달이 가능한 경우 그대로 사용한다. 이러한 예로는 과학, 뜻. 이유, 이치 모두가 같은 수화 기호를 사용한다. 이러한 유

사이에 대해서 수화 유사어 사전을 구성하여 번역 과정에서 사용한다면 수화 기호를 중복 저장하지 않아도 되므로 기본 수화 기호 사전이나 수화 복합어 사전의 용량을 감소할 수 있다. 또한 수화 단어의 부족함을 해결하는 방법이 될 수도 있다.

위와 같은 한국 수화의 특수성을 고려하여 제한한 수화 변환 사전은 기본 수화 기호의 동작을 저장하는 기본 수화 기호 사전, 기본 수화 기호의 조합으로 표현되는 단어를 저장하는 수화 복합어 사전, 동일한 수화 기호를 사용하는 단어들을 저장하는 수화 유사어 사전으로 구성된다.

2.3.1 기본 수화 기호 사전

기본 수화 기호 사전은 수화 기호 중에서 단일 기호로 그 의미를 표현할 수 있는 단어를 저장하는 사전이다. 따라서 레코드 구조는 키워드가 되는 단어와 동영상상의 정보로 구성되며 그 구조는 (그림 2)와 같다. K_Base는 형태소 분석과 재배열/생성 과정을 거쳐 수화로 표현해야 할 단어이다. 수화 특성상 품사를 구별하지는 않으나 단어의 이해와 분류를 위해 K_Part에는 품사를 기록한다. K_Noun이 검색의 기본 키이며, S_Animation은 수화 동작을 표현하기 위해 사용되는 동영상 파일의 정보이다.

K_BASE 1 = {K_Part, K_Noun, K_Bigo, S_Animation }
 :
 K_BASE n = {K_Part, K_Noun, K_Bigo, S_Animation }

K_Part	K_Noun	K_Bigo	S_Anim
명사	안녕		D:\Program\Sign\1\안
명사	안녕		D:\Program\Sign\2\안
명사	안녕		D:\Program\Sign\3\안
명동부	안녕	안녕	D:\Program\Sign\4\안
명동부	안녕		D:\Program\Sign\5\안
명사	안녕		D:\Program\Sign\6\안
명사	안녕		D:\Program\Sign\7\안
명사	안녕		D:\Program\Sign\8\안
명사	안녕		D:\Program\Sign\9\안
명사	안녕		D:\Program\Sign\10\안

(그림 2) 기본 수화 기호 사전의 구조

2.3.2 수화 복합어 사전

수화 복합어 사전은 기본 수화 기호 사전의 검색과정에서 실패한 경우 사용하는 사전으로 기본 수화 기호들의 조합으로 표현되는 복합 수화 기호들을 저장하는 사전이다 따라서 표현하는 기호만큼의 기본 수화 기호 사전의 키워드를 저장하고 있어야 하며 그 구조는 (그

림 3)과 같다

단어 K_Comp는 n개의 K_Base의 조합으로 구성된 복합 수화어나 수화 복합어 사전은 K_Comp를 키워드로 검색하여 n개의 K_Base를 구한다. 이후 수화 복합어 사전 검색에서 구한 K_Base들을 키워드로 기본 수화 기호 사전을 검색하여 S_Animation을 구하고 이들의 조합으로 표현하고자 하는 동영상상을 얻는다.

{K_Part, K_Comp 1, K_Base 1, K_Base 2, K_Base n}
 :
 {K_Part, K_Comp n, K_Base 1, K_Base 2, K_Base n}

K_Part	K_Comp	K_Base1	K_Base2	K_Base3	K_Base4	K_Base5	K_Base6	K_Base7
명사	안녕	안	안					
명사	안녕	안	안					
명동부	안녕	안	안					
명사	안녕	안	안	안				
명사	안녕	안	안					

(그림 3) 수화 복합어 사전의 구조

2.3.3 수화 유사어 사전

수화 유사어 사전은 기본 수화 기호 사전과 수화 복합어 사전에서 검색에 실패할 경우 사용하는 사전이다. 한 개의 기호로 여러 의미와 단어를 표현하는 수화의 특성을 반영한 사전으로서 (그림 4)와 같이 동일한 의미를 갖는 단어들을 한 개의 레코드로 저장하는 구조로 되어 있다.

단어 K_Like n은 수화 복합어 사전 검색에서 실패한 단어이다. 수화 유사어 사전은 K_Like n 모두들 키워드로 구성된다. 따라서 사전 검색은 복수키 검색하여 K_Sign을 구하고 이후 기본 수화 기호 사전 검색과 수화 복합어 사전 검색 과정으로 입력된다.

{K_Part, K_Sign 1, K_Like 1, K_Like 2, K_Like n}
 :
 {K_Part, K_Sign n, K_Like 1, K_Like 2, K_Like n}

K_Part	K_Sign	K_Like1	K_Like2	K_Like3	K_Like4	K_Like5	K_Like6
명사	안녕	안	안				
명사	안녕	안	안				
명사	안녕	안	안				
명사	안녕	안	안				
명사	안녕	안	안				
명사	안녕	안	안				
명사	안녕	안	안				
명사	안녕	안	안				

(그림 4) 수화 유사어 사전의 구조

2.4 수화 사전의 검색

수화는 '은, 는, 이, 가, 을, 로'와 같은 조사가 대부분 생략되며 명사, 동사, 형용사, 관형사, 부사 등의 품사에 상관없이 동일하게 사용한다. 예를 들어 <건강>, <건강하다>, <건강한> 등은 동일하게 사용된다. 또한 수화는 동사를 많이 사용하고 형용사가 많지 않기 때문에 대부분의 표현들이 단순하다. 그리고 문장의 모든 요소를 표현하는 것이 아니라 의미 전달에 필요한 부분만을 추출하여 사용한다. 따라서 형태소 분석이나 사전 구성시 이러한 한국 수화의 특징을 고려하여야 한다.

step1	Classification of basic element of Korean text
step2	Extraction of element and rearrangement/generation
step3	Search for sign language dictionary do every extracted word(Sword)
	1) Search for base sign dictionary if found then goto v)
	n) Search for compound sign dictionary if found then do each base sign repeat 1) procedure
	m) Search for resemble sign dictionary if found then change Sword into search result repeat 1) procedure
	iv) analyze Hangul code make finger spelling link
	v) Adding sign moving picture array
step4	Display the moving picture array on screen

(그림 5) 한국어-수화 번역 과정

(그림 5)는 제시한 사전들을 사용하여 한국어 문장을 수화로 번역하는 과정에서 적용한 알고리즘이다. 이 알고리즘은 원시언어인 입력된 한국어 텍스트 문장에 대해 형태소 분석, 제배열/생성, 수화사전의 검색, 그리고 목적인어인 수화동작의 표현 순서로 진행되는 기본적인 번역 과정을 기술하고 있다. 수화 사전의 검색 과정은 기본 수화 기호 사전, 수화 복합어 사전, 수화 유사어 사전 순서로 진행된다. 수화 사전 검색 과정을 기치면 한국어 문장은 최종적으로 수화 사전 검색을 통해 발견한 기본 수화 기호의 동영상 파일과 수화 사전 검색에 실패한 단어를 분석하여 표시한 지문자, 지숫자로서 재구성되며 이를 컴퓨터 화면상에 출력하게 된다.

3. 시스템의 구현

3.1 사용자 인터페이스

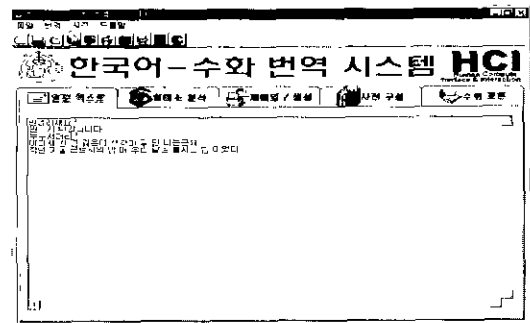
한국어-수화 번역 시스템을 이용하여 청각장애자에

게 한국어 문자를 교육하고, 건청인과 대화 수단으로 사용하려면 효율적인 사용자 인터페이스가 필요하다. 사용자 인터페이스는 한국어 문장이 수화로 번역되는 과정을 직관적으로 확인 할 수 있도록 원문 작성, 형태소 분석, 제배열 생성, 수화의 표현, 그리고 사전 작업의 다섯 부분으로 구성되었다.

첫 번째는 원문작성 과정이다. 입력된 한국어 문장은 저장된 텍스트 파일에서 읽거나 화면에서 키보드로 직접 입력이 가능하도록 구성하였다. 실험에서 사용된 한국어 문장의 예는 <표 1>과 같으며 한국어 텍스트 입력이 표시된 사용자 인터페이스는 (그림 6)과 같다.

<표 1> 실험에 사용한 한국어의 예

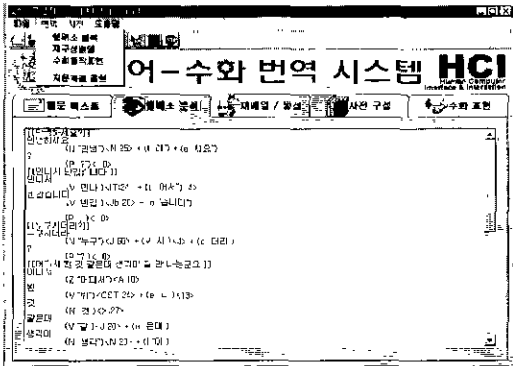
예제 한국어 문장	인녕하세요? 만나서 반갑습니다. 누구시더라? 어디서 뵈 것 같은데 생각이 잘 안 나는군요 작년 겨울 근로자의 날 때 우린 같은 팀 이었어요
실제 수화 표현	<ul style="list-style-type: none"> • <안녕> <만나다> <반갑다> • <누구> <의문> • <어디서=어편+곳> <보다> <같다> <생각> • <같> <나디> <아니다> • <작년> <겨울> <근로자=일-사람> <밤> • <때=시간> <우리> <같다> <팀=모임>



(그림 6) 입력된 한국어 텍스트

두 번째는 형태소 분석과정이다. 원문 작성 과정에서 입력된 한국어 문장을 형태소 분석기인 HAM 라이브러리를 이용하여 분석한 결과 값을 화면에 보여준다. 형태소 분석의 출력 화면은 (그림 7)과 같다.

세 번째는 제배열/생성 과정이다. 제배열/생성 과정의 처리는 (그림 8)의 좌측과 같이 형태소 분석에서 출력된 자료에서 품사와 단어들을 제외한 나머지 정보는 삭제한 후 제배열한다. 이 리스트에서 수화에서 사용되지



(그림 7) 형태소 분석 결과

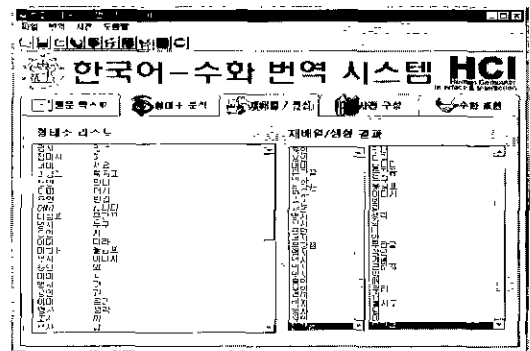
않는 조사, 어미와 같은 품사들을 삭제하며, 수화의 표현에 맞추도록 도치, 변형 등의 작업을 처리한다. 자연수화의 경우는 대부분의 경우에는 한글과 그 이순이 거의 같다. 그러나 수화 자체는 실제 대화에서 사용하는 언어이므로 한글의 문어체보다는 대화체에 가까운 특성이 있고, 또한 조사등이 생략되는 관계로 문장의 의미는 대화자의 경험적인 상식에 의존하는 부분도 매우 크다. 이러한 측면에서 어순이 다소 자유롭게 되는 경우도 존재한다. 물론 부사/어미가 '가능', '부정', '시제' 등의 의미 때문에 용언에 후치하는 등의 특성도 있지만, 이러한 부분은 별도의 예외처리차원에서 처리하였다. 따라서 일반적인 품사를 기반으로 처리하게 되며, 이에 관한 기준은 <표 2>와 같다 또한 이 과정을 처리하면 (그림 8)의 우측과 같이 수화에서 표현되어야 할 단어들의 리스트가 출력된다.

마지막은 수화 표현의 과정이다. 재배열/생성된 단어를 한 개씩 추출하여 기본 수화 기호 사전 검색 과정을 진행한다. 기본 수화 기호 사전 검색 과정에서 검색키와 일치하는 수화를 발견할 경우 수화 동작을 화면에 출력하고 다음 단어의 검색 과정으로 진행한다 발견하지 못할 경우 수화 복합어 사전 검색 과정을 진행한다. 수화 복합어 사전 검색에서 발견할 경우 수화 복합어를 구성하는 각각의 기본 수화 단어에 대해 기본 수화 기호 사전 검색을 반복하여 진행한다 수화 복합어 사전 검색에서 발견하지 못한 단어에 대해서는 수화 유사어 사전 검색을 한다. 수화 유사어 사전을 검색하여 발견할 경우 해당 레코드의 기본 수화 단어에 대해 다시 기본 수화 사전 검색 과정을 거쳐 수화를 화면에 출력한다 수화 유사어 사전 검색 과정에서도 검색에 실

패한 경우에는 한국어 코드를 분석하여 지문지로 화면에 표시한다

<표 2> 품사의 변환규정

명사	'것'을 제외한 모든 명사는 수화로 표현.
용언 선어말어미	용언은 기본적으로 수화로 표현 선어말어미는 해당수화표현이 있는 경우에만 수화로 표현 존칭, 심양에 관한 어미는 생략 부정Flag가 Set된 경우는 용언과 부정을 나타내는 수화를 도치
마침표	물음표일 경우 의문을 나타내는 수화기호를 표현 나침표일 경우는 그냥 문장을 종료
형용사 보조용언 부사	부정을 나타내는 '안'이면 부정Flag를 Set 이를 제외한 형용사, 부사는 모두 수화로 표현 보조용언은 해당수화표현이 있는 경우에만 수화로 표현
계사 이밀어미	~ 니다 일 경우 수화로 표현 이밀어미는 의문형에 대해서만 수화로 표현 나머지는 표현하지 않음
관형사, 감탄사, 조사, 명사형어미, 접두사, 접미사	표현하지 않음



(그림 8) 재배열/생성과정의 결과

기본 수화 기호 사전 검색이 완료된 단어들은 데이터 베이스에 저장된 CIF 형태의 동영상상을 순차적으로 화면에 출력하여 입력된 문장의 의미를 전달하게 된다.

(그림 9)에 한국어-수화 번역 시스템의 결과를 표시하였다. 화면의 좌측은 수화 동영상상을 출력하는 공간이며 중간은 지문자를 표시하는 공간이다. 화면의 우측은 재배열/생성 과정을 거쳐 수화 사전을 검색한 결과를 보여주는 공간이다. 수화 사전 검색에서 기본 수화

기호 사전에서 검색 성공한 단어는 Base라고 표시하며 수화 유사어 사전 검색에서 성공한 단어는 Like, 그리고 수화 복합어 사전에서 검색 성공한 단어는 Comp라고 표시하였다. 이러한 인터페이스는 하나의 화면에 수화동영상, 지문자, 한국어문자 모두를 표현하여 각각의 한국어 문자가 어떠한 수화기호와 지문자로 대응되는지를 보여준다. 이것은 수화 사전의 검색 결과를 쉽게 알 수 있고 수화 유사어일 경우 동일한 의미로 사용되는 수화 기호가 무엇인지를 인지하게 하며, 수화 복합어일 경우 그 각각을 구성하는 기본 수화 기호가 무엇인지를 알게 해 준다

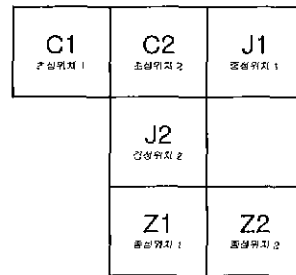


(그림 9) 수화표현

3.2 수화 지문자와 동작의 표현 기법

본 연구에서 지문자의 표현은 실제 사용되는 순차적인 방법이 아닌 동시 표현이라는 새로운 방법을 사용했다. 예를 들어 “꼭”이라는 문자를 청각장애인이 사용하는 지문자로 표현할 경우 “ㄱ → ㄱ → ㅍ → ㅍ → ㄱ” 순서로 수화 지문자를 일정한 시간 간격을 두고 표현해야 한다. 이러한 방법은 한 개의 문자를 표현하려고 문자의 표현에 사용된 자음, 모음 수만큼의 시간이 요구되어 변환 속도에 영향을 준다. 빠른 의미 전달을 위하여 한국어의 초성, 중성, 종성을 동시에 출력하도록 설계하였다. 하나의 문자를 수화 지문자로 표현할 때 2~6개의 지문자를 순차적으로 표현해야 하지만 제한한 지문자의 표현 방법은 하나의 화면에 지문자를 표시하므로 소요 시간을 절약할 수 있다. 지문자의 출력 위치도 사용자가 쉽게 인지하도록 (그림 10)과 같이 한국어와 동일하게 6개의 위치를 갖도록 설계하였다. 수화 지문자의 표현은 한국어 완성형 코드 값을 분석하여 검색하며 완성형 코드 2350개 모두에 대해 (그

림 10)에서 제시한 위치 값을 포함하도록 데이터 베이스를 구성하여 저장하였다. 본 논문에서는 한글 텍스트가 주로 완성형이기 때문에 수화 지문자 표현 시 완성형코드를 사용하였지만 조합형코드 혹은 유니코드로의 적용도 가능하다



(그림 10) 지문자의 위치구조

수화 기호의 표현을 위해 사용한 GIF는 수화의 동작 유형에 따라 보통 1~8개의 프레임으로 구성된다. 안녕'이라는 수화를 표현하기 위해서는 (그림 11)과 같이 4개의 프레임을 사용한다. (그림 11)의 각 프레임을 구하는 방법은 수화 애니메이션을 사용하는 방법과 디지털 카메라를 사용하는 방법이 있다. 수화 애니메이션 방법은 작업이 복잡하여 제작시간이 많이 걸리지만 수화동작의 특징을 잘 표현할 수 있으며 16색 이하의 색상 수와 낮은 해상도로 작업이 가능하므로 제작된 GIF 파일의 크기를 줄일 수 있다.



(그림 11) 수화 생성

디지털 카메라로 촬영하는 방법은 고정된 디지털 카메라로 사진을 촬영한 후 구분된수화 동작의 사진을 GIF 파일로 변환하는 방법이다. 이 방법은 디지털 사진의 특성 때문에 해상도가 높아지고 제작된 GIF 파일의 크기가 커지게 되지만 작업시간이 적으며 간단하게 수화 GIF를 제작할 수 있는 장점이 있어 본 연구에서는 디지털 카메라를 사용하여 수화 GIF를 제작하였다. 실제로 '만나다'라는 단어를 수화로 표현할 경우 GIF로 표현된 크기는 27K바이트이지만 AVI로 표현한 경우는 172K바이트이다.

4. 결 론

본 연구에서는 청각장애인에게 문자를 교육하고 건청인과 대화 수단으로 사용할 수 있는 한국어-수화 번역 시스템을 개발하는 과정에서 요구되는 제 사전들을 구성하였고, 구성된 사진을 이용하여 한국어 문장이 수화로 번역되는 과정을 구현하였다. 수화 사전은 기본 수화 기호 사전, 수화 복합어 사전, 수화 유사어 사전으로 분류하였다. 사전 검색을 수행하여 직접 방식의 기계 번역 방법을 이용하여 한국어 문장을 수화로 번역하였다. 구성된 수화 사전은 사전의 저장 공간을 줄이고 한국어-수화 번역 시스템의 구성에 융통성을 부여하며 한국 수화에서 나타나는 어휘의 부족 문제도 해결할 수 있었다. 또한 연속 GIF를 사용하여 수화 동영상의 시간, 공간적인 수화로 적절하게 출력하였다.

예문을 이용한 수화 사전 검색 실험 과정에서 간단한 문이체 한국어 문장들은 청각장애인이 이해 가능한 수화로 번역되며, 또한 번역에 사용한 사전의 유용성도 확인하였다. 그러나 실제 수화의 표현은 단순한 손동작만으로는 표현되지 않고 얼굴의 표정과 동작의 크기 등을 고려해야 하며, 화자에 따라 수화의 방향이 바뀌는 문제 등 단순한 데이터베이스 검색만으로는 해결할 수 없는 문제들이 많이 있다. 또한 채택/생성 과정에서 발생하는 많은 예외의 경우들에 대한 연구와 다양한 한국어에 대한 수화 사전 구성에 많은 시간과 노력이 필요하며 수화 GIF 생성을 위해 사용한 디지털 사진을 색상 수와 해상도가 적은 애니메이션으로 변환하여 출력 속도를 향상시키고 저장공간을 줄여야 한다.

본 논문에서 제시한 수화 사전과 한국어-수화 번역 시스템은 음성-수화 번역기 개발의 한 부분으로 사용될 수 있으며 청각장애인에게 문자를 읽고 쓰는 능력향상

에도 유효할 것으로 사료된다.

참 고 문 헌

- [1] Sharon L. Glennen, Denise C. DeCoste, *The Handbook of Augmentative and Alternative Communication*, Singular Publishing Group, 1996, pp.3-6.
- [2] 서경수, 국어문법. 뿌리깊은 나무, 1994. pp.5-10
- [3] 김기범 등, 청각장애편람, 한국 청각장애자 복지회, 1991. pp.233-240, 254-256, 456-474. 567-578
- [4] 한국 농아어린이 이웃사랑회, 열린수화. 도서출판 푸른귀. 1998.1, pp.231-232
- [5] 김승국, 한국 수화 연구, 오성출판사, 1996, pp.35-58.
- [6] Sherman Wilcox, "American Sign Language as a Foreign Language," http://www.unm.edu/~wilcox/ASLFL/asl_f.html.
- [7] 박수현, 강석훈, 권혁칠. "한국어-수화 번역시스템을 위한 형태소 변환", 한국정보처리학회 논문지 제 5권 제3호, 1998.3, pp.688-702.
- [8] Tom Holcomb, Joy Kreeft Peyton, "ESL Literacy for a Linguistic Minority The Deaf Experience," http://www.cal.org/NCLE/digests/ESL_LITERACY.HTML.
- [9] Charrow, V. R., *The written English of deaf adolescents In M.F. Whiteman (Ed.), Writing. The nature, development, and teaching of written communication*, Vol.1, Hillsdale, NJ: Lawrence Erlbaum Associates, 1981, pp.179-187.
- [10] 손연기. "청각장애인을 위한 복지정보통신 정책방향", '97 청각장애인 복지 통신 제2호를 위한 심포지엄, 청음회관, 1997, pp.109-140.
- [11] David R. Beukelman, Pat Mircnda, *Augmentative and Alternative Communication*, Paul H. Brookes Publishing Co, May 1995
- [12] 엄미숙, "한국수화의 통사론적 특징 분석", 석사학위 논문, 대구대학교 대학원, 1996
- [13] CompuServe Incorporated, "Graphics Interchange Format Programming Reference," CompuServe Incorporated Document, July 1990.
- [14] 김재훈 등, '기계번역', 마이크로소프트웨어, 1997.1, pp.386-396.
- [15] 강석훈, "대화체 일한 번역 시스템의 설계 및 구현", 박사학위 논문. 한양대학교, 1995.

- [16] 김영덕, 자연언어처리, 교학사, 1994.5, pp.56-93.
- [17] 강승식, “한국어 형태소 분석기와 한국어 분석 모듈 (HAM: Hanguk Analysis Module)”, <http://ham.hansung.ac.kr/ham/ham.html>.
- [18] 강승식, “한국어의 형태론적 특성과 형태소 분석기법”, 정보과학회지 제12권 제8호, 1994. 9, pp.47-59
- [19] 국제오토시스템, 수화통역사, <http://www.vivashop.com>



권 경 혁

email : nansys@chollian.net
 1991년 광운대학교 전자계산기 공학과 졸업(공학사)
 1999년 인천대학교 대학원 정보통신공학과 졸업(공학석사)
 1999년 인하대학교 대학원 정보통신공학과 박사과정

1990년~1993년 삼성전관 컴퓨터사업부 사원
 1993년~1994년 인우시스템 개발과장
 현재 인천대학교 공과대학 정보통신공학과 박사과정.
 나인정보기술 대표
 관심분야 : 대체통신공학, CTI(Computer Telephony Integration)



우 요 섭

e-mail : yswwoo@hon.inchon.ac.kr
 1986년 한양대학교 전자통신공학과
 1988년 한양대학교 대학원 전자통신공학과(공학석사)
 1992년 한양대학교 대학원 전자통신공학과(공학박사)

1992년~1994년 인천대학교 정보통신공학과 조교수
 1994년~현재 인천대학교 정보통신공학과 부교수
 관심분야 : 한국어 정보처리, 멀티미디어 정보검색, VOD



민 흥 기

email : hkmin@hon.inchon.ac.kr
 1979년 인하대학교 전자공학과 졸업(공학사)
 1981년 인하대학교 대학원 전자공학과 정보공학전공(공학석사)

1990년 인하대학교 대학원 전자공학과 정보공학전공(공학박사)
 1985년~1991년 한국과학기술연구원 선임연구원
 1991년~현재 인천대학교 정보통신공학과 부교수
 관심분야 : 디지털신호처리, AAC, CTI