

한글문장-수화 번역기를 위한 사전구성

권경혁*, 민홍기

인천대학교 정보통신공학과

(032-770-8284, 032-772-2289, nainsys@chollian.net* hkmin@lion.inchon.ac.kr)

A Construction of Dictionary for Korean Text to Sign Language Translation

*K. H. Kwon, H. K. Min

Dept. of Information and Telecommunication Engineering, Univ of Incheon,

(032-770-8284, 032-772-2289, nainsys@chollian.net* hkmin@lion.inchon.ac.kr)

Abstract

Korean Text to Sign Language Translator could be applied to learn letters for both the deaf and hard-of-hearing people, and to have a conversation with normal people. This paper describes some useful dictionaries for developing korean text to sign language translator: Base sign language dictionary, Compound sign language dictionary, and Resemble sign language dictionary. As korean sign language is composed entirely of about 6,000 words, the additional dictionaries are required for matching them to korean written language. We design base sign language dictionary which was composed of basic symbols and moving picture of korean sign language, and propose the definition of compound sign language dictionary which was composed of symbols of base sign language. In addition, resemble sign language dictionary offer sign symbols and letters which is used same meaning in conversation. By using these methods, we could search quickly sign language during korean text to sign language translating process, and save storage space. We could also solve the lack of sign language words by using them, which are appeared on translating process.

1. 서론

청각장애인의 의사소통수단은 구화, 수화, 필담 등이 있다. 구화의 경우 건청인과의 의사소통이 용이하기는 하나 교육이나 직업 등 수많은 다양한 정보를 얻기에는 한계가 있으며 상대방과 근거리에서 마주보며 대화해야 한다는 어려움이 있다. 수화의 경우 청각장애인이 자신의 생각을 표현하는 수단으로 많이 사용하였으나 모든 사람에게 적용되는 통신수단이 아닌 일부 그 사용법을 알고 있는 사람들에게만 적용되는 제한적인 언어이다.

필담은 가장 일반적인 의사소통의 수단으로 건청인과 청각장애인이 의사소통을 할 때 많이 사용되고 있으나, 청각장애인중에는 문장력이 부족한 사람이 많아 의사소통이 잘 되지 않을 수도 있다. 필담으로 자유롭게 의사소통을 할 수 있는 청각장애인은 아주 적으며 많은 청각장애인들이 문자언어의 표현에 어려움을 겪고 있다. 미국의 경우 청각장애 고등학교 졸업자들은 표준화 읽기 검사 결과, 영어를 대략 초등학교 3,4학년 정도 수준으로 읽는 것으로 나타났다.[7] 따라서 일반적인 통신수단인 문자를 청각장애인에게 제공해 주는 것은 그들에게 교육과 직업, 정보 획득 등에서 지금보다 많은 기회를 제공할 것이다.

통신보조기에 대하여 적용되는 기술 중 한가지인 AAC(Augmentative and Alternative Communication)는 통신(말)을 사용할 수 없는 장애인들에게 제공되는 대체통신방법을 설명할 때 사용된다.[1] AAC 인터페이스는 장애인들이 가지고 있는 능력(speech, vocalization, gesture, signs, aided communication)을 최대한 활용할 수 있도록 해주는 복합적 통신이라고 하였다.[1] 최근 컴퓨터 기술의 발달로 음성과 영상의 출력기능을 가진 컴퓨터의 보급이 확산되고 또한 저 가격으로 판매되고 있으므로 이를 이용하여 AAC 시스템을 구현한다면 언어장애를 가진 사람들에게 좀 더 보편적이고 경제적인 것이다.

본 연구에서는 이러한 AAC 개념을 바탕으로 문자학습 환경이 부족한 청각장애인에게 수화와 문자를 동시에 교육시킬 수 있고, 또한 수화와 문자를 매개체로 하여 일반인과 대화할 수 있는 보조통신장치로서 문자를 수화로 바꾸어주는 시스템인 한글문장-수화 번역기(Korean Text to Sign Language Translator)에 적용하기 위한 사전구성을 제안한다.

한글 문장을 수화영상으로 변환하는 과정에서 필요한 수화변환사전들을 수화의 특성에 따라 분류하여 기본 수화기호 사전, 수화 유사어사전, 수화 복합어사전으로 구성하고, 이러한 사전들을 이용하여 청각장애인이 이해할 수 있는 수화언어를 개인용 컴퓨터 화면상에 수화기호의 동영상과 지문자의 연속으로 표현하는 실험을 수행하였다.

2. 한글문장-수화 번역기의 구성

2.1 수화관련 연구

현재 한국에서 사용되고 있는 수화는 방법적 유형에 따라 한글 문법과 일치하게 표현되는 문법식 수화와 농자사회에서 자연발생적으로 생성된 자연수화로 크게 구분된다.[3]

자연수화는 청각장애인에 의해 창조되고 발전되어온 일종의 고립어이며, 문법식 수화는 자연수화의 기호, 지문자, 지숫자로 국어의 문법에 맞게 체계화된 수화이다. 현실적으로 청각장애인 사회에서 문법식 수화보다 자연수화가 더 넓게 사용되고 있고 또한 청각장애인은 음성을 들음으로써 습득되는 음성언어의 문법을 학습하기가 매우 어려우므로 본 연구는 대부분의 청각장애인이 많이 사용하는 자연수화를 대상으로 진행되었다.

자연수화의 기호는 지사(指事), 모방(模倣), 상형(象形), 형지(形指), 형동(形動), 회의(會意), 전주(轉注)로 구분된다.[4]

- (1) 단어가 의미하는 것을 가리키는 동작을 해 보이는 지사(예 : 머리)
 - (2) 단어가 의미하는 동작이나 그 동작의 특성을 본 뜬 동작을 해 보이는 모방(예 : 보다)
 - (3) 단어가 의미하는 대상의 형체나 그 윤곽을 그려 보이는 상형(예 : 산)
 - (4) 단어가 의미하는 대상의 형체를 만들어 보이고 그것을 가리키는 동작으로 구성되는 형지(예 : 비탈)
 - (5) 단어가 의미하는 대상의 형체나 그 일부를 만들어 그 대상이 하는 동작을 해 보이는 행동(예 : 꼬끼리)
 - (6) 단어가 의미하는 것을 둘 또는 그 이상의 기존 기호를 합쳐 나타내는 회의(예 : 일요일)
 - (7) 기존기호를 다른 뜻을 가진 단어의 기호로 사용 하는 전주(예: 눈물=비참, 슬픔, 애도)
- 등 일곱 가지 기호 구성방법에 의해 구성되며[4] 이러한 자연수화의 기호는 대략 6000 여 개 밖에 안 된다.

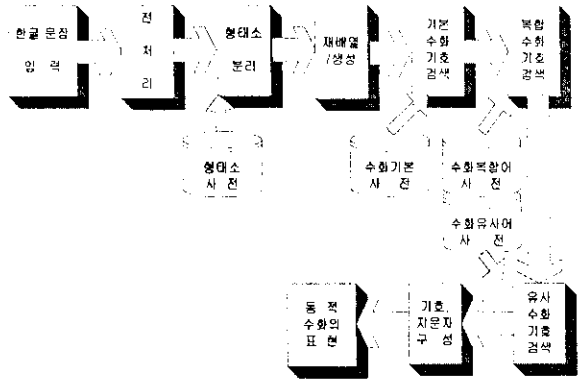


그림 1 한글문장-수화번역기의 구조

구성에 따른 분류로써 한국수화를 보면 한국 수화는 한글 지문자, 영문 지문자, 기본수화기호와 이러한 기호를 조합하여 생성한 복합수화 기호로 분류할 수 있다. 한국 수화의 단어 6000여개 중에서도 기본수화기호 동작들을 복합하여 표현되는 수화기호(회의)가 다수 존재한다. 복합수화 기호란 한자의 '田'자와 '力'자가 합쳐져 '男'자라는 다른 의미를 나타내는 글자가 되었듯이 수화에서도 둘 이상의 기존기호가 다른 의미를 나타내는 기호가 된 것을 말한다. 예를들어 '일요일'의 수화표현은 <빨갈다> + <문이 닫히다, 폐교하다> + <날>이 된다. 이와 같은 회의 수화기호의 예는 표 1 과 같다.

2.2 한글문장-수화 번역기

한글 문장을 문법식 수화로 번역한다면 이는 단순한 단어의 치환으로 변환이 가능하지만 자연수화로 번역 한다면 기계번역의 범주에서 고려해야 할 것이다.[5] 다음의 예문을 이용하면 문법식 수화와 자연수화가 상당히 다르다는 것을 알 수 있다.

예 1) 당신의 이름은 무엇입니까?
 문법식수화 : <당신>(의)<이름>(은)<무엇><이다><바니까>
 자연수화 : <당신><이름><무엇>

즉 한글문장을 수화로 표현 할 때 문법식 수화는 한글 형태소들을 한글문장에 맞도록 일대일 치환하고 일치하는 기호가 없으면 지문자로 표현하여 변환이 가능하다. 그러나 자연수화는 자체 언어적인 특성이 한국어와는 다르므로 한글문장을 자연수화로 변환하는 과정은 기계번역의 관점에서 처리해야 하며 적절한 방법은 형태소 분석과 사전검색과정을 거치는 직접방식을 사용해서 처리하였다.[6] 직접방식 기계번역의 과정은 번역할 문장이 입력되면 형태소 분석기를 통해서 문장을 어근과 조사로 분해하고 분해된 단어들을 사전 검색을 통해 번역하고자 하는 언어의 단어로 변환한다. 그후 재배열/생성단계에서 문장을 보완한 후 번역된 문장의 출력하는 과정을 거쳐 번역작업이 이루어진다. 직접 방식 기계번역은 한/일, 영/불 과 같은 동일 어족에 속한 언어의 번역방식으로는 적합하나 사전구성과 재배열/생성도들이 대단히 복잡해지게 된다.[6]

한글문장-수화 번역기의 구조는 그림 1에 나타냈다. 번역기는 입력된 한글문장의 형태소를 분석하는 형태소 분석기, 분석된 어근과 조사에 대해 적절한 수화로 바꾸어 주는 기본단어 추출기와 동적 수화를 저장하는 기본수화사전, 수화복합어 사전, 수화유사어 사전, 그리고 사용자에 게 표시되는 인터페이스로 구성된다.

3. 수화사전의 구성

직접방식 기계번역을 사용하여 한글문장을 수화로 번역하기위해 가장 먼저 고려해야할 사항은 사전의 구성이다. 한국 수화 사전구성을 위한 수화기호의 분류는 어원에 따른 분류보다는 구성과 형태에 따른 분류를 사용하여야 할 것이다.

표 1 회의 수화의 예

단 어	수 화 표 현
<언제>	<무엇> + <시간>
<입원>	<병원> + <눕다>
<계획>	<생각> + <자르재다>
<소화불량>	<위> + <어렵다>
<안개>	<흰색> + <희미하다>

복합수화기호는 기본수화기호로 형성된 것만 있는 것이 아니라 '카톨릭', '컴퓨터' 와 같이 수화기본기호와 지문자를 조합하여 만들어진 것도 있다. 회의를 사용하여 기존 기호의 조합으로 수화 기호를 만들어 내는 것은 지금까지도 많은 수가 존재하지만 계속해서 새롭게 만들어 질 수 있으며 수화만이 가지는 언어의 융통성을 제공해주고 수화단어의 부족성도 해결하는 방법이 될 것이다. 수화기호에 없는 단어의 경우 다음의 예문과 같이 회의를 사용하면 여러 가지 다른 의미를 나타내는 단어를 생성할 수도 있다.

예 2) 전입생
 자연수화의 회의 : <이사> + <학교> + <입학>

이러한 복합수화(회의)들은 기본 수화기호들을 조합하여 많은 의미적 확장이 가능하므로 한글문장-수화 번역시스템에서 복합수화사전을 추가, 삭제가 가능하도록 유연하게 시스템을 구성한다면 수화단어의 부족함을 해결할 수 있을 것이다.

또한 한국수화는 유사한 의미를 가진 한글단어에 대해 동일한 동작으로 그 의미를 표현하는 경우(전주)도 다수 존재한다.

이는 단어가 의미하는 것을 수화기호로 바꾸는데 한계가 있기 때문에 기존의 수화로도 의미전달이 가능한 경우

그대로 사용하는 경우가 많기 때문이다.

전주(轉注)는 수화에서만 볼수 있는 것이 아니라 한국어에서도 얼마든지 볼수 있으나 수화에서 일어나는 전주는 한국어나 한자등 언어문자에 비해 대단히 많이 존재한다. 수화에서 나타나는 전주의 예로는 <과학>, <뜻>, <이유>, <이치> 모두가 같은 수화 기호를 사용한다. 이러한 동일한 수화기호를 사용하는 유사어에 대해서 수화유사어 사전을 구성하여 번역과정에서 사용한다면 기본수화 기호 사전의 용량을 감소할 수 있고 빠른 사전검색이 가능한 것이다.

위와 같은 한국 수화만의 특수성을 고려하여 제안한 사전의 구조는 기본수화사전, 수화복합어사전, 수화유사어 사전으로 다음과 같은 구조로 구성되었다.

① 기본 수화기호 사전

기본 수화기호 사전은 수화기호 중에서 단일 기호로 그 의미를 표현할 수 있는 대표적인 단어이다. 따라서 레코드 구조는 키워드가 되는 단어와 동영상상을 포함해야 하며 그 구조는 다음과 같다.

K_Base 1 = (K_Part, K_Noun, K_Bigo, S_Animation)
 :
 K_Base n = (K_Part, K_Noun, K_Bigo, S_Animation)

단어 K_Base는 형태소 분석 결과와 재배열/생성 과정을 거쳐 수화로 표현 해야할 단어이다. 수화 특성상 품사를 구별하지 않으나 단어의 이해와 분류를 위해 K_Part에는 품사를 기록해 준다. K_Noun이 검색을 위한 기본키가 된다. S_Animation은 컴퓨터에서 수화동작을 표현하기 위해 사용되는 동영상 파일로 최근 인터넷상에서 많이 사용되는 Animated GIF 형태로 저장하여, 연속재생기법을 사용한다면 수화 동영상상을 저장하는 공간의 절약과 복합적인 수화표현에 적합할 것이다.

② 수화복합어 사전

수화복합어 사전은 기본 수화기호들의 조합으로 표현되는 복합수화기호들을 저장하는 사전이다. 따라서 복합해서 표현해야 하는 기호만큼의 기본수화기호 사전의 키워드를 저장하고 있어야 하며 그 구조는 다음과 같다.

(K_Part, K_Comp 1, K_Noun1, K_Noun2, K_Noun n)
 :
 (K_Part, K_Comp n, K_Noun1, K_Noun2, K_Noun n)

단어 K_Comp는 K_Noun들의 조합으로 구성된 복합수화(회의)이며 기본수화사전 검색에서 실패한 단어인 K_Base 와 동일하다. 수화복합어사전은 K_Comp를 키워드로 검색하여 발견하면 K_Noun들을 구한다, 그후 K_Noun를 키워드로 하는 기본 수화기호 사전 검색하여 S_Animation을 구하고 이들의 조합으로 최종적으로 표현하고자 하는 동영상상을 얻는다.

③ 수화유사어 사전

수화유사어 사전은 동일한 의미를 갖는 단어들을 한 개의 레코드로 저장하는 사전으로 그 구조는 다음과 같다.

(K_Part, K_Sign 1, K_Like1, K_Like2, K_Liken)
 :
 (K_Part, K_Sign n, K_Like1, K_Like2, K_Liken)

단어 K_Like는 복합수화사전 검색에서 실패한 단어이며

형태소 분석 결과로 나온 기본형이다.

수화 유사어사전은 K_Like 모두를 키워드로 구성된다. 따라서 사전 검색은 복수키 검색하여 발견하면 K_Sign을 기본 수화기호사전 검색과 복합 수화어 사전 검색과정으로 입력된다. 한 개의 기호로 여러 의미와 단어를 표현하는 수화의 특성을 반영한 사전이다.

4. 수화사전의 검색 실험

수화는 은, 는, 이, 가, 을, 를, 로 와 같은 조사가 대부분 생략되어 있으며 명사, 동사, 형용사, 관형사, 부사 등의 품사에 관여하지 않고 동일하게 사용한다. 예를 들어 <건강> = <건강하다> = <건강한> 등 품사에 상관없이 동일하게 사용된다. 또한 문장의 모든 요소를 표현하는 것이 아니라 의미진달에 필요한 부분만을 추측하여 사용해야 한다. 따라서 형태소 분석이나 사전구성에 있어서도 이러한 한국 수화만의 특성을 고려하여야 하며 재배열/생성 과정을 간단하게 처리할 수 있도록 설계되어야 한다.

그림 2 는 지금까지 제시한 사전들을 사용하여 한글문장을 수화로 번역하는 과정에서 기본적으로 적용 가능한 알고리즘을 나타낸다.

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

그림 2 한글문장-수화번역의 알고리즘

위의 과정을 거치면 한글 문장은 최종적으로 수화유사어사전과 수화복합어사전 검색을 통해 기본수화기호의 동영상 파일과 지문자, 지숫자로서 재구성되며 이를 컴퓨터 화면상에 지문자, 지숫자는 정적 그림화일의 연속으로, 기본 수화기호는 동적 이미지로 출력하게 된다.

한글문장을 입력하여 수화 데이터베이스를 검색하는 실험에서 사용된 한글 문장의 예는 그림 3 과 같다.

안녕하세요? 만나서 반갑습니다. 누구시더라? 어디서 된 것 같은데 생각이 잘 안 나는군요. 작년 겨울 근로자의 밤 때 우린 같은 팀 이었어요.
--

그림 3 입력한 한글문장

위의 예문을 한글공학연구소(HAM)의 한글 형태소분석기인 CORAN[2]에 입력하면 그림 4와 같이 명사, 용언, 부사, 어미, 조사 등 많은 품사를 가진 형태소로 분석된다.

```

[[안녕하세요?]]
안녕하세요
      (N "안녕")<N:25> + (t "하") + (e "세요")
?
(P "?")< :0>
[[만나서 반갑습니다.]]
만나서
      (V "만나")<IT:24> + (e "어서")<4>
반갑습니다
      (V "반갑")<Jb:20> + (e "습니다")
(P ".")< :0>
    
```

그림 4 형태소 분석 결과

형태소 분석된 원시 자료는 그림 5의 좌측 리스트와 같이 단어와 품사로 구성되며 재배열/생성 과정을 통해 그림 5의 우측과 같이 품사들 중에서 수화로 표현되지 않는 조사, 어미 등을 삭제하고, 그후 용언과 체언 중에서도 삭제, 도치, 변형 등의 작업이 필요한 부분에 대해 수화표현에 맞도록 재구성 처리를 한다.

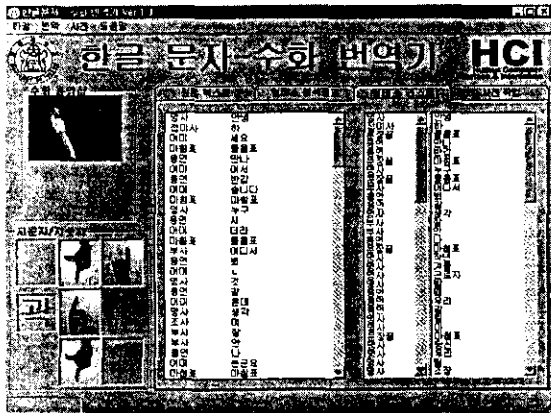


그림 5 재배열/생성 결과

재배열/생성 과정의 결과로 출력된 수화표현 단어들은 그림 6과 같이 수화 사전 검색과정을 거친다. 그림 6의 우측은 좌측의 재구성된 단어들에 대해서 수화사전 검색의 과정을 완료한 자료를 출력한 것이다. 각각의 단어들에 대해 기본수화사전, 유사수화사전, 복합수화사전을 차례로 검색한다. Base라고 표현된 단어는 기본수화사전을 검색하여 발견한 단어이며, Comp는 복합수화 사전을 검색하여 발견한 단어이고 Like는 유사 수화사전을 검색하여 발견한 단어이다. 수화사전 검색이 완료된 단어들은 데이터베이스에 저장된 GIF 형태의 동영상을 순차적으로 화면에 출력하여 입력된 문장의 의미를 전달하게 된다.

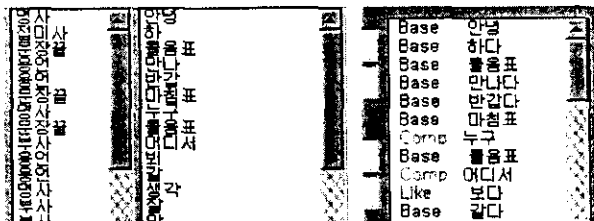


그림 6 수화사전 검색 결과

또한 그림 5와 그림 6은 지금까지 진행된 한글문장-수화 번역기의 사용자 인터페이스를 보여준다. 화면의 좌측은 수화 동영상과 지문자로 표시하는 공간이며 화면의 우측은 한글문장을 수화로 번역하는 과정을 차례로 보여주는 공간으로 할당했다. 지문자의 표현은 빠른 의미전달을 위해 한글의 자음, 모음을 순서대로 보여주는 손동작과는 다르게 한글의 초성, 중성, 종성을 한 화면에 출력하도록 설계하였다. 지문자의 출력위치도 사용자가 쉽게 인지하도록 화면에서와 같이 한글과 동일하게 6개의 위치를 갖도록 하였다. 한글 텍스트 문장의 입력은 저장한 파일에서 읽거나 화면에서 키보드로 직접 입력이 가능하도록 구성하였다.

5. 결론

본 논문에서는 청각장애인에게 문자를 교육시키고 건청인과 대화수단으로 사용할 수 있는 한글문장-수화 번역기를 개발하는 과정에서 필요한 세 사전들을 구성하였고 그 사전을 이용하여 한글문장이 수화로 번역되는 과정을 실험하였다. 수화사전을 크게 기본 수화기호 사전, 수화복합어 사전, 수화유사어 사전으로 분류하고 이러한 사전검색을 통해 직접방식 기계번역으로 한글문장을 수화로 변환하는 과정중 동영상 재생을 제외한 전 과정을 설명하였다. 이러한 방법은 수화사전의 저장공간을 감소시키고 한글문장-수화번역기 구성에 융통성을 주며 한국수화에서 나타나는 어휘의 부족문제도 해결할 수 있다.

예문을 이용한 수화 사전검색 실험과정에서 간단한 문어체 한글문장들은 청각장애인이 이해 가능한 수화로 구성되는 것을 확인할 수 있었다. 그러나 실제로 수화의 표현은 단순한 손동작만으로는 표현할 수 없고 얼굴의 표정이나 동작의 크기 등을 고려해야 하며, 화자의 입장에 따라 수화의 방향이 바뀌는 문제 등 단순한 데이터베이스 검색만으로는 해결하지 못하는 사항들이 많이 있다. 또한 재배열/생성과정에서 발생하는 수많은 예외들에 대한 처리도 많은 실험을 통해 보완해야 할 것이며 광범위한 언어영역에 대한 수화사전 구성작업도 많은 시간과 노력이 필요할 것이다.

본 논문에서 제시한 수화사전과 한글문장-수화 번역기는 음성-수화 번역기 개발의 한 부분으로 사용될 수 있으며 청각장애인에게 문자를 읽고 쓰는 능력을 가르치는데 매우 유효 할 것으로 사료된다.

참고문헌

- [1] Sharon L. Glennen & Denise C. DeCoste, The Handbook of Augmentative and Alternative Communication, Singular Publishing Group, 1996
- [2] 강승식, 한글공학연구소, <http://ham.hansung.ac.kr/>, 1998
- [3] 김기범 외, 청각장애편람, 한국 청각장애자 복지회, 1991
- [4] 김승국, 한국 수화 연구, 오성출판사, 1996
- [5] 박수현 외, 한국정보처리학회 논문지 제5권 제3호, 1998.3
- [6] 김재훈 외, 마이크로소프트웨어, 정보시대, 1997.1
- [7] Holcomb 외, http://www.cal.org/NCLE/digests/ESL_LITERACY.HTML