

어휘정보를 이용한 문장작성에 관한 연구

A Study on the Sentence Generation using Lexical Information

황인정*, 민홍기*

Ein-Jeong Hwang*, Hong-Ki Min*

요약

본 연구는 언어 장애를 가진 사용자의 언어생활을 돕기 위한 문장작성 방식을 제안한 것이다. 제안한 문장작성 방법은 통신보조기기에 적용할 수 있도록 시스템으로 구현해 보았다. 통신보조기기는 개인 휴대장치로서 필요한 문장을 작성하여 출력하는 기기이다. 언어장애인들의 대표적인 의사표현 수단인 수화는 일반인들과의 의사소통에는 불편하기 때문에 다른 표현 방법이 필요하다. 자모를 모두 입력하여 문장을 작성하는 방식은 키의 수가 많기 때문에 대화할 때 시간이 많이 소용되는 불편한 방식이다. 그러므로 언어장애인을 위한 문장 작성의 가장 중요한 목적은 적절한 장소와 상황에 맞는 단어가 배열된 통신보조기기를 이용하여 적은 키의 수로 문장을 작성하는 것이다. 본 연구의 문장 작성을 구현하기 위한 어휘정보는 사용자 영역을 정하고, 그 영역에 맞는 단어와 문장을 수집하고, 수집된 어휘의 특징을 추출하여 구축되었다. 그리고 제안한 문장 작성 방식의 효용을 측정하기 위해 시스템을 이용하여 문장 작성을 위한 키의 수와 키보드 입력에 의한 자모수를 비교하였다.

ABSTRACT

This study suggests a sentence generating method to help those who have language impediment with their communication. The method suggested in this study was constructed into a system in order to be applied to AAC system. AAC system is a personal portable device that generates sentences. Those who have language impediment need another communication method, causes inconvenience when used in a conversation with those who don't have the same trouble. The method of inputting both consonants and vowels can be inconvenient and time consuming for a conversational communication because of the number of the key strokes. The lexical information for the sentence generating of this study defines the user's domain, collects the adequate words and sentences, and extracts and classifies the characteristics of the collected words. The comparison between the number of key strokes for sentence generating using the system and that of inputting consonants and vowels using a keyboard was made in order to evaluate the usefulness the sentence generating method.

Key words : sentence generation, lexical information, augmentative and alternative communication

I. 서론

본 연구는 통신보조기기의 목적에 맞도록 어휘정보를 구축하여 문장을 작성하는 것이다. 통신보조기기는 보완

-대체통신 (Augmentative and Alternative Communication : AAC) 중 고급공학체계이며, 일상적 언어생활이 불편한 사람들을 위한 보조기기[1]로서 필요한 문장을 작성하여 음성으로 출력하는 기기이다. 그러므로 통신보조기기는 사용자의 특성과 환경에 따라 입력방법과 언어능력에 따른 어휘의 수와 내용, 크기와 부착형태가 다르나, 통신보조기기의 일반적인 형태는 어휘부분, 표시부분, 제어부분으로 나뉘며, 일반인과의 대화 시 빠르게 문장을 작성할 수

*인천대학교 정보통신공학과

접수 일자 : 2004. 4. 14 수정 완료 : 2004. 7. 23

논문 번호 : 2004-1-15

※본 연구는 2004년도 인천대학교 멀티미디어 연구센터의 일부 지원에 의하여 수행되었음.

있는 것이 중요한 목적이 된다.

본 연구는 적은 키의 수로 문장을 작성하기 위한 방법에 초점을 맞추었고, 그 목적에 맞도록 어휘정보를 구축하였다. 어휘정보는 어휘발체와 분류, 영역의 계층구조, 말뭉치를 이용하여 구축되었으며, 구축된 어휘정보는 문장작성과 저장, 재사용에 이용되었다.

II. 어휘정보의 구축

어휘정보는 적은 키의 수로 문장이 작성될 수 있도록 구축하였다. 통신보조기에 적용하기 위한 어휘정보는 어휘의 미시적 특성이 아니라, 문장 내 명사와 동사의 위치, 연결정보를 이용하였다.

2.1. 어휘의 발체와 분류

효과적인 문장작성을 위한 어휘발체는 사용자의 언어 환경 즉, 연령, 이용하는 장소, 어휘능력을 고려한 어휘와 연령에 따른 사회적으로 통용되는 어휘를 포함하여 어휘발체가 이루어져야 한다. 사용자의 어휘능력과 이용 장소만을 사용하여 어휘발체를 한다면 한정된 어휘사용으로 어휘능력의 향상을 기대할 수 없으므로, 사회적으로 통용되는 어휘자료를 추가하여야 한다.

어휘발체는 사용자의 언어 환경을 고려하여 진행되어야 하나, 아직 사회적으로 통용되는 연령별 어휘가 정립되어있지 않은 관계로 본 연구는 초급 어휘를 발체하였다. 초급어휘의 발체는 초등교과서와 회화 책, 인터넷을 통한 자료모음과 수화 사이트 그리고 다른 연구에서 수집된 어휘와 논문[2][3]을 참고하였다.

어휘는 특정 영역에서 빈번히 사용되는 영역별 어휘와 특정 상황에 구애받지 않고 널리 사용되는 중심어휘 그리고 숫자, 연월일, 색, 단위, 시제 등의 단어를 포함한 영역으로 구분하여 발체하였다. 본 연구에서 특정 영역은 교통, 식당, 쇼핑, 집으로 분류하였다. 특정 영역으로 어휘를 발체하는 것은 특정 상황에서 빈번히 사용하는 어휘를 통신보조기에 배열하여 사용자가 쉽게 문장을 작성하기 위해서이다. 중심어휘는 인사말, 도움말을 포함한다. 표 1은 중심어휘의 예이다.

표 1. 중심어휘의 예

Table 1. The examples of the core sentences

안녕하세요. 도와주세요. 고맙습니다. 미안합니다, 괜찮습니다. 반갑습니다. 예. 아니오.

교통, 식당, 쇼핑, 집 영역에 따라 발체된 단문은 명사부와 동사부로 분류하였다. 문장 분류는 동사와 명사의 연결 정보를 제공해준다. 동사에 따라 연결되는 명사의 종류[4]를 알 수 있기 때문에 이러한 연결정보를 이용하여 문장을 작성할 수 있다. 문장 분류 시 분류하는 사람

의 의도에 따라 명사종류와 동사종류의 이름이 다를 수 있다. 명사종류, 조사, 동사종류의 연결정보는 동사에 따라 명사의 종류가 한정되므로, 새로운 어휘가 입력되었을 때도 기존의 명사와 동사로 결합된 연결정보[5]를 이용하여 문장을 작성할 수 있는 장점이 있다. 각 영역에 따라 발체된 문장을 이용한 문장 분류는 어휘정보를 구성하는 기초 자료를 제공한다. 표 2는 교통영역의 문장 분류의 예이다.

발체된 문장은 명사부분에 적당한 이름과 동사부분에 적당한 이름을 붙였다. 예를 들면 '학교는 어디에 있습니까?'에서 '학교' 명사는 '장소' 라고 하는 대표적인 명사종류를 붙였고, '어디에 있습니까?'는 '위치' 라는 동사종류를 붙였다. 문장 분류는 발체된 문장을 명사와 동사에 대표적인 명칭을 붙인 것이다. 동사종류와 명사종류는 각 영역에서 파생된 하위도메인의 개념이며, 각 영역에 따라 다시 세분화 될 수 있다.

표 2. 교통 영역의 문장 분류의 예

Table 2. The examples of the sentences are classified by nounal and verbal type in transportation domain

(장소)에서 내리려면 얼마나 걸려요? (시간)
 몇분이나 걸려요? 걸립니다. 같가요. (시간)
 (장소)는 어디에 있습니까? 있습니다. (거리)
 (장소)는 어떻게 가야 합니까? 가야 합니다.(거리)
 얼마나 더 가야하나요? 가야합니다.(거리)
 (교통수단)를 타고 가야 합니다. (거리)
 (장소)역까지는 얼마입니까? (요금)
 요금(요금)은 얼마입니까? (요금)
 교통 카드값(요금)은 얼마입니까? (요금)
 (장소)역 (몇장) 주세요. (기타)
 (요금)이 없습니다.(기타)

2.2. 어휘정보의 구축

어휘정보는 문장을 작성하기 위해 필요한 어휘 자료의 모음으로 정의된다. 어휘정보는 각 영역의 하위도메인 [6][7]과 어휘의 분류로부터 얻어진 동사종류에 연결되는 명사종류, 조사, 동사활용을 포함하고 있다. 각 영역의 하위도메인은 각 영역에서 어휘의 범위를 규정하여 수집할 수 있도록 한다. 동사활용은 질문과 답변, 청유, 요청으로 문장을 바꾸어 사용할 수 있도록 한다. 조사는 명사와 동사의 연결을 위한 요소이다. 각 영역의 하위 도메인의 구조는 문장의 분류로 이루어지며, 교통 영역의 경우 장소, 교통수단, 시간, 거리, 요금, 기타로 나누어져있고, 장소는 다시 지명, 전철역, 기차역, 위치로 나누어진다. 각 영역의 계층 구조는 세분화할 수 있다. 표 3은 교통영역의 하위도메인을 나타낸 것이다.

표 3. 교통영역의 하위도메인

표 3. 교통영역의 하위도메인

Table 3. The transportation domain's sub domains

교통 =>장소 -> 지명 ->
->전철역 ->
->기차역 ->
-> 위치 ->
수단 ->
요금(명사) ->
시간 ->
거리 ->
요금(동사) ->
기타 ->

작성된 각 영역의 하위도메인을 참고하여 어휘를 다시 수집하였다. 각 영역의 하위도메인의 범위를 벗어나지 않는 말뭉치[8]는 초기 어휘발췌에서 얻지 못했던 어휘들을 체계적으로 얻을 수 있는 장점이 있다. 표 4는 교통 영역의 말뭉치의 예이다.

표 4. 교통 영역의 말뭉치 예

Table 4. The examples of the corpus in transportation domain

번호 1. 장소
1.1 지명 : 인천, 서울,
1.2 전철역 : 제물포, 주안, 서울,..
1.3 기차역 : 서울, 부산, 대구,..
1.4 위치 : 이곳, 저곳, 그곳, 여기, 저기,
번호 2. 수단
2.1 교통수단 : 지하철, 전철, 버스, 택시. 기차..
번호 3, 요금
3.1 요금 : 가격, 값, 요금 전철비, 교통카드..
번호 4 기타
4.1 교통표시판 : 안내표시판, 좌회전, 유턴,
번호 5 시간 : 얼마나 가야 합니까? -가야합니다.
번호 6 거리 : 얼마나 걸립니까? -걸립니다.
번호 7 요금 : 얼마입니까? -입니다.

발췌된 어휘 중 동사는 기본형, 평서:답변, 의문, 요청, 허용, 부정으로 나누었다. 동사를 동사활용[9]으로 정리한 이유는 사용자가 질문과 답변, 청유, 요청, 부정으로 문장을 바꾸어 사용하기 위해서이다. 문장유형은 평서문이라 하더라도 의미기능에 따라 서술, 의문, 명령, 청유의 의미로 나타낼 수 있고, 의문문도 질문의문문, 요청의문문, 수사의문문으로 분류할 수 있고 더 상세한 문장유형분류도 할 수 있다. 문장 유형은 연구자에 따라 여러 형식으로 분류[11]한다. 그러므로 문장의 유형을 명확히 구분하는 것은 어려운 문제이다. 본 연구의 문장유형 분류는 평서문은 답변형식으로 의문문은 질문형식으로 정하였고, 요청, 허용, 허가는 동사의 기본형을 이용하였다. 부정의 경

우 평서, 요청, 의문, 허가로 동사 활용된 동사를 다시 부정형식으로 만들었다. 모든 동사가 동사활용의 다양한 문장유형을 갖지 않는다. 어휘 발췌와 분류로부터 얻어진 자료를 이용하여 문장 작성을 위한 어휘정보 사전에 구축한다. 어휘정보 사전에는 하위도메인, 동사의 기본형과 동사활용, 동사종류, 연결 가능한 명사의 종류, 조사를 포함한다. 조사는 각 명사의 종류에 따라 자세히 분석했다. 동사와 조사는 동사에 연결될 수 있는 명사의 종류를 한정[4][10]한다. 표 5는 문장작성을 위한 동사 어휘정보 사건의 예이다.

어휘정보 사전에서 동사에 따라 명사종류가 여럿 나타나는 이유는 여러 영역에서 중복되어 사용되기 때문이다. 조사는 명사 종류에 따라 차례대로 배열하였다. 은/는, 을/를 등의 조사는 문장을 작성할 때 명사의 중성을 검색하여 알맞게 붙여지도록 하였다. 조사는 한국어 초급 교재의 기초 어휘 선정에 관한 논문[3]을 참고하여 “까지, 게, 께서, 나, 도, 만부터, 서, 에, 예게, 에서, 와/과, (으)로, 은/는, 을/를, 의, 이/가”를 발췌하였고, 그 외 “도, 하고, 해서, 해, 동안, 하러, 하기, 하고, 께, 께서,..” 등 어휘 발췌에서 발견된 조사를 포함하여 27개로 배열하였다. 분류된 동사는 103개, 명사는 284개이며 여러 영역에 중복되어 사용된다.

표 5. 동사 어휘 정보 사건의 예
Table 5. The examples of the verbal lexical information dictionary

[문형번호1]
[기본형] (시간이) 걸리다.
[동사 하위 도메인 :동사종류] 시간
[명사 하위 도메인 :명사종류] 위치/수단
[조사] 까지, 으/로
[평서:답변] 걸립니다.
[요청] 걸리게 해주세요.
[의문] 얼마나 걸립니까?
[요청:허가] 걸려도 됩니다.
[평서부정] 걸리지 않습니다.
[요청부정] 걸리지 않게 해 주세요.
[의문부정] 걸리지 않습니까?
[허가부정] 걸리면 안 됩니다.

명사 어휘정보는 명사번호, 명사이름, 명사종류(하위도메인), 상위 도메인으로 작성하였다. 명사 어휘정보는 동사가 포함된 어휘정보와 연결된다. 명사 어휘정보 사전은 표 6에 나타내었다.

표 6. 명사 어휘정보사전의 예
Table 6. The examples of the nounal lexical information dictionary

[명사번호 1]
 [명사] 밝은색
 [명사하위도메인 : 명사종류] 물건
 [상위도메인] 색

III. 문장 작성

통신보조기기에 적용 가능한 문장작성에 관한 연구는 아직 미비하다. 한글을 이용한 문장작성에 관한 연구로는 어휘를 의미심볼과 함께 붙여 단순하게 사용자가 문장을 작성한 것[12]과 신경망을 이용하여 명사에 따라 연결될 수 있는 동사를 제시해주는 것[13]이 있다. 명사에 따라 동사를 제시해주는 것은 명사에 연결될 수 있는 동사가 많은 경우 키의 수가 늘어날 수 있으며, 새로운 어휘가 시스템에 입력되었을 경우, 다시 어휘 연결에 관한 정보를 모두 새로 작성해야 하는 불편이 있다. 그 외 가상네트워크를 이용한 방법[14]이 제안되었으나 이 방법은 어휘사이의 관련성을 명확히 수치화 시키는 것이 어려운 문제이다.

본 연구에서 제안한 방법은 명사 종류와 동사종류와 연결로 문장을 작성하는 것이다. 이 방법은 사용자가 어떤 명사를 선택하면 문장 분류에 의해 구축된 어휘정보를 이용하는 것이다. 이 방법은 영역으로 한정된 문장의 작성이 용이할 뿐만 아니라, 새로운 어휘가 시스템 내에 입력되었을 경우 기존의 명사와 동사의 연결 구조를 그대로 사용할 수 있는 장점이 있다. 그림 1은 교통 영역의 연결 구조를 나타낸다.

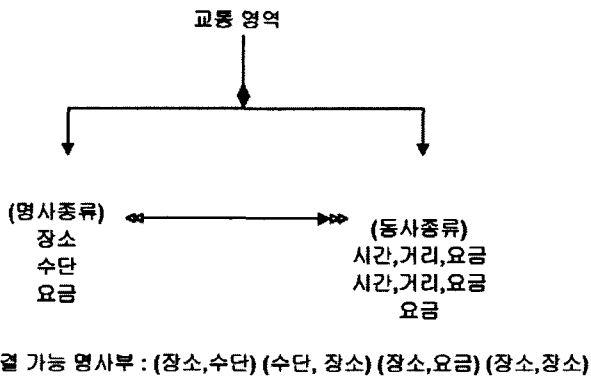


그림 1. 교통영역의 가변적인 연결구조

Fig. 1. The transportation domain's variable pattern net construction

문장 작성은 사용자가 명사종류 중 명사, 동사종류 중 동사를 선택하게 되면, 어휘정보에 구축된 동사종류와 명사종류의 비교에 의해 조사가 붙게 된다. 그리고 명사의 중성을 유니코드로 검색하여 알맞은 조사를 선택한다. 이러한 과정을 거쳐 문장이 작성된다. 조사연결에 관한 것

을 그림 2에 나타내었다.

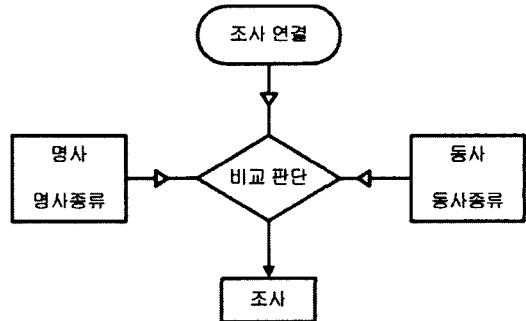


그림 2. 조사 연결

Fig. 2. A connection of postpositional word

문장 작성을 위한 어휘정보는 사용자가 어휘를 입력하여 문장을 작성할 때 동사종류, 명사종류에 따라 적절한 조사가 연결되는 형태를 가지고 있다. 그러므로 어휘정보는 문장의 작성 뿐만 아니라, 어색한 문장의 검출과 저장에도 이용할 수 있다.

어색한 문장의 검출은 사용자에게 의해 작성된 문장이 출력된 후 저장될 때 이용한다. 문장작성은 우선적으로 사용자 중심의 문장이다. 그러므로 어색한 문장이 작성되더라도 우선 출력한다. 그런 후 출력된 문장을 어휘정보를 이용하여 저장여부를 판단한다. 저장여부의 판단은 구축된 어휘정보의 명사와 동사종류의 연결 정보를 이용하여, 구축된 정보와 차이를 보일 때 저장하지 않는다. 저장된 문장은 재사용이 용이하도록 하기 위해, 동사의 기본형을 이용하여 빈도수를 표시하여 저장한 것이므로, 올바른 문장의 저장이 필요하기 때문이다.

올바른 문장으로 판단되었을 경우 출력된 문장의 저장 방법은 출력된 문장의 동사기본형과 저장 문장 내 동사 기본형을 검색하여 일치하는 동사기본형이 없는 경우는 빈도수에 의해 배열된 저장 공간 내 하위부분에 저장되고, 저장 문장 중 동사 기본형이 있는 경우에는 동사에 연결된 빈도수를 하나 늘려준다. 문장에 연결된 명사는 최근 사용된 명사가 저장된다.

문장의 재사용도 어휘정보를 이용하였다. 저장된 문장은 어휘정보와 연결되어 있고 동사에 빈도수가 포함된 형태를 가지고 있다. 저장된 문장은 어휘정보와 연결되어 있으므로, 명사를 바꾸어 출력이 가능하다. 명사를 바꾸고자 할 때, 그 명사종류에 해당하는 명사가 화면에 나타난다.

문장 작성을 위해 특정 영역과 어휘를 선택하여 문장을 작성하고, 저장과 재사용에 관한 내용을 전 과정을 그림 3에 나타내었다. 문장 작성, 저장여부, 재사용에 어휘정보가 이용되었다.

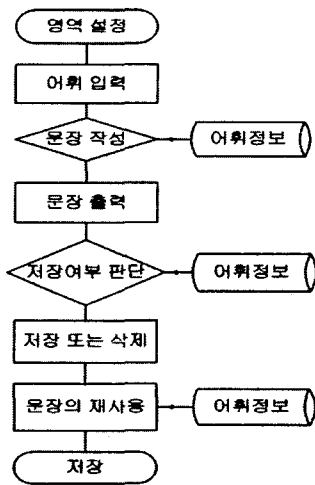


그림 3. 문장 작성 방법

Fig. 3. Method of the sentence generation

IV. 구현

구축된 어휘정보를 이용한 문장작성은 통신보조기기에 적용할 수 있도록 하였다. 통신보조기기의 일반적인 구성 요소는 어휘부분, 제어부분, 표시부분으로 나뉘어져 있다. 그러므로 본 연구의 문장작성은 통신보조기기의 일반적인 형태를 취하였다. 구성 요소는 선택된 어휘를 보여주는 표시부분과 문장의 삭제와 출력을 위한 제어부분 그리고 각 도메인의 어휘부분이다. 그림 5는 어휘정보를 이용한 문장작성의 구현 모델이다. 각 영역의 어휘가 하위 도메인에 배열되어 있고, 시제와 숫자는 앞부분에 배열하였으며, 색, 단위 등의 단어는 특정영역을 정하여 배열하였다. 어휘의 배치는 왼쪽에 명사종류 오른쪽에 동사종류를 배열하였다. 동사활용은 '선택'이라는 제어부분을 통해 표시된다. 동사활용은 제어 부분을 사용하면 다양한 문형의 동사가 출력된다. 선택된 어휘를 보여주는 표시부분은 상, 하 2부분으로 배열하였다. 사용자가 선택한 어휘는 표시부분 중 아래에 배치되고, 완성된 문장은 위쪽에 나타난다. 여러 영역 중 교통 영역을 선택하고 명사종류는 장소를 선택했으며 그 아래 하위도메인 중 전철역, 지명, 기차역, 위치가 표시되어 해당 어휘를 선택 할 수 있도록 보여준 것이 그림 4이다. 동사도 같은 방법으로 선택한다. 문장작성은 동사에 연결되는 명사를 한개 또는 두개를 선택할 수 있으며, 동사만 선택할 수도 있다.

명사와 동사를 선택하여 조사를 붙여 완성된 문장을 보여준 것이 그림 5이며 표시부분의 윗부분에 문장을 나타낸다. 출력된 문장은 어휘정보사전과 비교하여 문제가 없는 경우 저장된다. 저장될 때 출력된 문장의 동사기본형과 저장된 문장의 동사기본형을 비교하여 새로 저장되거나, 동사기본형과 연결된 빈도수를 늘려준다.

그림 6은 저장된 문장 중에서 명사를 바꾸어 출력하는 과정을 나타낸 것이다. 저장 문장 중 “인천역에서 얼마입

니까?” 문장을 명사 ‘서울역’으로 바꾸어 “서울역에서 얼마입니까” 문장을 작성한 것이다.

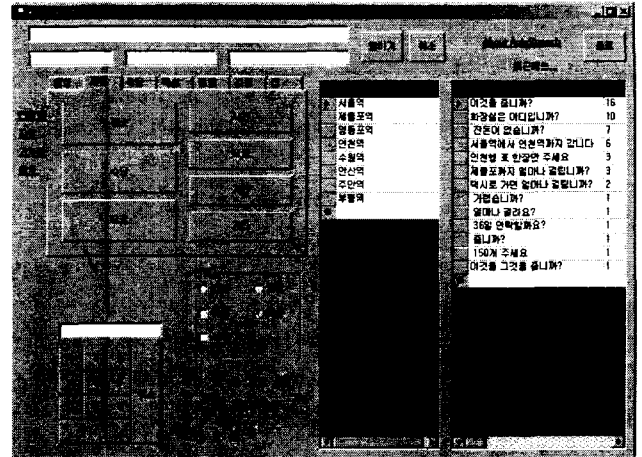


그림 4. 명사 선택

Fig. 4. A selection of noun

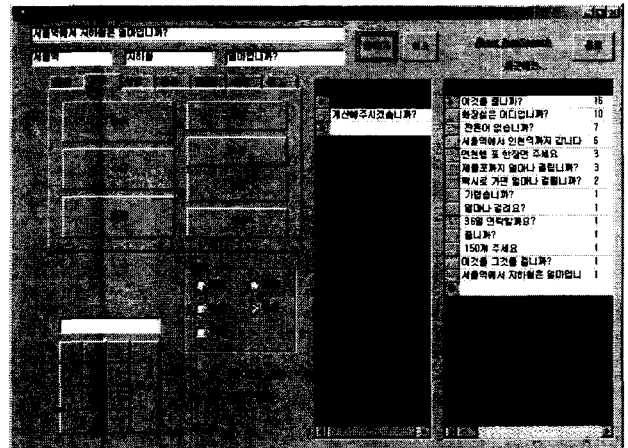


그림 5. 문장작성

Fig. 5. Sentence generation

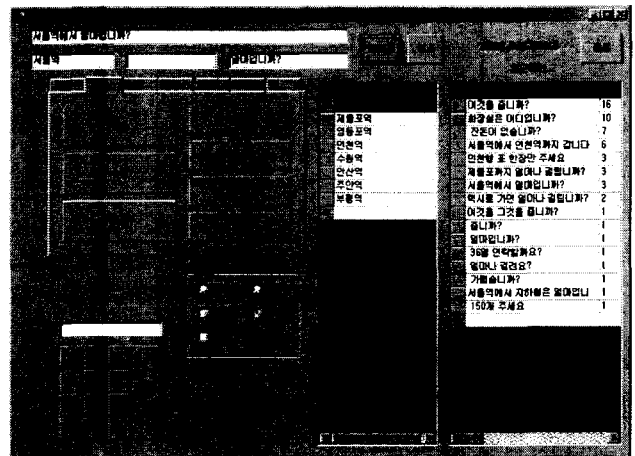


그림 6. 저장 정보의 예

Fig. 6. The examples of storage information

V. 검토

통신보조기기에 적용하기 위한 문장작성은 적은 키의 수를 이용해야 하기 때문에 구현된 시스템에서 문장작성을 위한 키의 수와 키보드를 이용한 자모수를 비교하여 시스템의 효용을 측정하였다. 시스템에서 문장작성을 위한 최대 키의 수는 영역선택, 명사종류, 명사하위도메인, 명사, 동사종류, 동사, 동사활용, 어휘정보 보완, 출력을 포함한 9개 키의 수이다. 키보드를 이용한 자모수는 마침표, 띄어쓰기, 쌍자음을 포함하여 측정한 것이다. 시스템에서의 키의 수와 키보드를 이용한 총 자모 수는 아래의 식에 대입하여 측정하였다.

$$\left\{ 1 - \frac{(keystrokes + menu\ selections)}{keystrokes\ needed\ without\ prediction} \right\} \times 100$$

표 7은 실험문장과 총자모수, 식에 의한 효용성을 측정 한 것이다. 실험문장은 시스템에서 명사와 동사로 이루어진 문장이다.

표. 7. 키의 수에 대한 효용성의 결과
Table 7. A result of keystroke savings

실험문장	총 자모 수	키의 수 9 일때 효용성 계산 (%)
서울역까지 얼마입니까?	29	68.9
이것은 얼마입니까?	24	62.5
몇 시간이나 걸려요?	23	60.9
체물포까지 가야합니다.	25	64.0
이것을 주세요.	17	47.1
메뉴판을 보여주세요.	23	60.9
계산서를 알려주세요.	24	62.5
서울역은 어디입니까?	25	64.0
인천까지 전철로 얼마입니까?	34	73.5
서울까지 가 주세요.	21	57.1
서울에서 가 주세요.	21	57.1
평균	24.2	61.7

효율의 측정은 영어의 경우 보완-대체 통신에 적용할 수 있는 영역을 정하여 단어예측을 이용하여 문장을 작성하였을 때 방법에 따라 차이를 보이지만 50% 이상의 효율을 보였다는 결과[15]가 있다.

본 연구의 시스템은 단어예측은 아니지만, 문장을 작성할 때 최대 키의 수 9개에 대한 효율은 평균 61.7%를 나타냈다. 실험문장은 시스템을 이용하여 만들 수 있는 문장들로 명사 1개 + 동사, 명사 2개+동사, 자모수가 적은 단어를 포함한 문장, 자모수가 큰 단어가 포함된 문장등을 실험문장으로 하였다. 일상대화 시 키의 수는 영역선택과 동사활용의 선택부분을 제외한 7개 키의 수를 주로 사용한다. 시스템에서 영역과 동사활용부분은 자주 바꾸

지 않기 때문이다. 그러므로 7개 키에 대한 효율은 표 7에 나타난 문장을 이용하였을 때 평균 70.2%로 측정되었다. 그리고 본 시스템에서 관용표현의 문장은 동사부분에 포함시켜 문장을 작성하였다. 관용표현의 문장은 명사와 동사로 분리하지 않고 한 문장으로 사용해야 올바른 의미를 전달하기 때문이며, 이러한 특성을 이용하여 동사부에 포함시킴으로써 키의 수를 줄였다. 그러므로 관용표현의 경우 4개 또는 3개 키의 수가 필요하므로, 높은 효율을 보인다. 그리고 저장문장을 다시 사용하는 경우 키의 수가 2개이므로 더욱 높은 효율을 얻을 수 있다.

VI. 결론

본 연구는 통신보조기기에 적용하기 위한 문장 작성에 관한 것이다. 일상대화에서 빠르게 사용할 수 있도록 키의 수를 줄여 문장을 작성하는데 초점을 맞추었다. 문장작성과 저장, 재사용은 어휘정보를 구축하여 사용하였으며, 어휘정보는 문장작성과 검토, 저장에 이용하였다. 어휘정보는 각 영역의 어휘발췌와 분류를 통해 이루어진 정보로서 각 어휘에 따라 동사에 연결될 수 있는 명사종류, 어휘의 적절한 연결을 위한 조사, 동사활용을 포함하여 구축되었다.

구축된 시스템의 효율을 측정하기 위해 시스템에서의 키의 수와 키보드를 이용한 자모수를 비교 측정하였다. 최대 키의 수를 이용한 값이 평균 61.7%의 결과를 얻었으며, 방법에 따라 좀더 높은 효율이 나타난다.

본 연구는 통신보조기기에 적용할 수 있는 문장작성의 초기연구이다. 향후과제는 실용적인 시스템이 될 수 있도록 저장목록의 검색부분 추가와 의미심볼과의 결합, 적절한 어휘발췌, 그리고 개인휴대장치가 되기 위한 사용자 어휘입력부분과 하드웨어의 연구가 필요하다.

참고문헌

- [1] S. L. Glennen and Decoste, The Handbook of Augmentative and Alternative Communication, Singular Publishing Group, Chapter 3, 1996
- [2] 수화사랑“http://myhome.naver.com/minirose/”
- [3] 조성문, “한국어 초급 교재의 기초 어휘 선정에 관하여”, 한국언어문화학회 한양어문, Vol.15, No.0, pp.317-348, 1997
- [4] 홍재성 외, ‘현대 한국어 술어 구문 사전’, 두산동아, 1997
- [5] 이정민, “한국어 술어 중심의 의미구조”, 한국인지과학회 춘계학술발표논문집, pp.32-40, 1997
- [6] 추교남, ‘개념 기반 정보 검색을 위한 한국어 어휘의 의미분석’, 인천대학교 석사학위청구논문, 1998.12,
- [7] M. O'Donnell, “Input specification in the WAG sentence generation system”, In INLG 8, pp. 41-50.

1996

[8] 김영택 외 10인, '자연 언어 처리', 교학사, 제7장, 1994

[9] Ann Copestake. "Augmented and alternative NLP techniques for augmentative and alternative communication", In Proceedings of the ACL workshop on Natural Language Processing for Communication Aids, pp. 37-42, Madrid. 1997

[10] Daniel Jurafsky & James HL.Martin, 'Speech and Language Processing', Prentice-hall, chapter 8, 2000

[11] 박영순, '한국어문장의미론', 도서출판 박이정, 2001.11

[12] 황인정, '의미심볼에 의한 통신보조기기의 설계', 인천대학교 석사학위청구논문, 1998.12.

[13] 이은실, '제한된 공간내에서 동사예측기법을 이용한 언어장애인용 문장발생장치에의 구현', 인하대학교 박사학위 논문, 2001.2.

[14] 황인정, 김효진, 이은주, 민홍기, "통신보조기기용 어휘예측시스템의 구조", 대한신호처리학회 하계학술대회, 2001.6

[15] Gregory Lesh and Gerard J. Rinkus, "Domain-Specific Word Prediction For Augmentative Communication", Enkidu Research, Inc.



황 인 정 (Ein-Jeong Hwang)

1990년 인천대학교 물리학과 이학사

1999년 인천대학교 정보통신공학과 공학석사

1999년2월-현재 인천대학교 정보통신공학과 박사과정

주관심분야 : 보완 대체 통신, HCI



민 홍 기 (Hong-Ki Min)

1979년 인하대학교 전자공학과 공학사

1981년 인하대학교 전자공학과 공학석사

1990년 인하대학교 전자공학과 공학박사

1985년10월-1991년8월 한국과학기술원
선임연구원

1991년8월-현재 인천대학교 정보통신공학과 교수

주관심분야 : 보완 대체 통신, HCI
