

# 대화음성인식 시스템 구현을 위한 기본 플랫폼 개발

정민화\*, 서정연\*, 이용주\*\*, 한명수\*  
서강대학교 컴퓨터학과\*, 원광대 전기전자 및 정보공학부\*\*

## Development of a Baseline Platform for Spoken Dialog Recognition System

Minhwa Chung\*, Jungyun Seo\*, Yong-Ju Lee\*\*, Myungsoo Han\*  
Department of Computer Science, Sogang University\*  
Division of Electric and Electronic Engineering, Wonkwang University\*\*  
E-mail : {mchung, seojy}@sogang.ac.kr, yjlee@wonkwang.ac.kr,  
uuhan@nlpeng.sogang.ac.kr

### Abstract

This paper describes our recent work for developing a baseline platform for Korean spoken dialog recognition. In our work, We have collected about 65 hour speech corpus with auditory transcriptions. Linguistic information on various levels such as mophology, syntax, semantics, and discourse is attached to the speech database by using automatic or semi-automatic tools for tagging linguistic information.

### I. 서론

사람이 음성 대화를 통해서 컴퓨터에게 정보 및 명령을 전달하려면 컴퓨터가 사람의 대화 음성을 이해해야 한다. 그러기 위해서는 다양한 대화 음성의 수집을 통한 대화의 특성 연구가 필요하다. 따라서 다양한 대화 영역 및 대화 환경 특성이 반영된 대용량 대화음성 DB의 구축과 대화체 문장의 언어적 특성 정보 추출을 위한 대화 정보 태깅 표준안 수립이 중요하다. 또한 그러한 언어적 특성 정보를 쉽고 빠르게 부착할 수 있도록 대량의 대화 말뭉치 구축 및 말뭉치 구축을 위한 태깅 및 분석 도구 개발이 중요하다.

본 논문에서는 텔레뱅킹 영역의 대화음성 DB를 녹음하고 발음정보 및 언어정보 태깅을 수행하는 과정을 설명하였으며 이를 통해 대화음성 인식시스템 개발의 기반이 되는 데이터를 구축하는데 필요한 고려사항을

보여준다.

본 논문의 구성은 다음과 같다. 1장의 서론에 이어 2장에서는 대화음성 코퍼스 수집방법 및 특성을 설명하고, 3장에서는 대화체 음성 DB 전사 및 발음열 태깅 도구 개발, 4장에서 언어정보 분석 및 태깅방법을 설명한다. 5장에서는 구축된 DB의 규모와 태깅 예를 보이고 6장에 결론을 서술한다.

### II. 대화음성 DB 구축

#### 2.1 대화음성 DB 수집 방법

##### 가. 대화 영역 설정

대화 음성 DB의 수집 대상 영역으로서 텔레뱅킹을 설정하고 시나리오를 작성하였다. 또한 본 연구에서는 텔레뱅킹 도메인에서 사용자의 자연스러운 발화를 얻기 위해 데이터 및 시스템의 오류에 대한 화자의 반응을 살펴보기 위한 대화 음성 DB를 제작하였다.

##### 나. 시나리오 제작

대화 음성 DB 제작을 위한 시나리오의 제작 목적과 제작 방법은 다음과 같다.

- 텔레뱅킹 영역
  - 은행에서 서비스하는 실제 웹사이트와 전화상담 시스템을 분석하여 텔레뱅킹 시스템의 흐름에 맞는 시나리오를 제작하였다.
- 시나리오 작성
  - 가상의 음성인식 시스템 역할을 하는 시나리오를 구성[1]

- 시스템의 음성 인식 오류에 대한 화자의 반응
  - 인식 키워드에 대한 인식 오류의 유형을 설정하고 이에 맞는 시나리오를 제작하였다.
  - 시스템이 사용자의 음성 입력을 확인하는 경우에만 하여 인식 오류 시나리오를 제작하였다.

표 2. 텔레뱅킹 DB 시나리오 영역

서브 도메인	내용
환율 정보	환율 정보서비스, 환율조회, 환전수수료, 환율 할인
대출 정보	대출 상품안내, 대출금액한도, 대출금리
예금상품정보	예금상품안내, 예금상품선택
암호변경	텔레뱅킹 암호 변경, 이체암호변경
예금 조회	잔액조회, 입출금 내역조회, 계좌이체
계좌이체	수신 은행, 수신인이름, 수신인계좌번호
공과금 납부	전화 요금 납부, 전화 요금 납부 내역조회, 아파트 관리비 납부, 아파트 관리비 납부 내역조회
신용 카드	혜택 안내, 결제 대금 조회, 사용 내역, 사용 한도 조회, 현금 서비스 이체, 사용 명세서 주소지 변경
분실 신고	현금 카드 분실, 직불 카드 분실, 수표 분실, 통장 분실, 인감 분실
서비스	우편 서비스, 팩스 서비스

다. 녹음 환경과 방법

대화체 음성은 조용하고 방음 시설을 갖춘 환경에서 마이크와 유선전화 두 채널로 다음 [표 3]의 장비를 사용하여 녹음하였다. 녹음 진행자는 별도의 공간에서 프로그램을 조작하여 시스템 발화를 발성자에게 전달하고 발성자의 음성을 녹음하였다.

표 3. 녹음 장비

장비명	기능
TEAC TASCAM M-08	Mixer
Sennheiser HMD 25-1	Headset
Dialogic D/4 PCI	Telephony Board
SoundBlaster Live	Sound Card
유선전화기	전화기
Pentium III PC	Computer

### III. 대화체 음성 DB 전사 및 발음열 태깅 도구 개발

#### 3.1 데이터 전사

전사 규칙의 목적은 이를 이용하여 음향 모델과 언어모델을 훈련시키고자 함이다. 음향 모델을 훈련하기 위해서는 발성된 대로의 표기가 필요하며 언어모델을

위해서는 사투리나 잘못 발성된 단어들에 대해서는 표준 표기로 바꾸어 주는 것이 필요하다. 전사규칙은 발성된 대로 표기함을 기본 원칙으로 하고, 이에 표준 표기를 덧붙여 준다.

#### 3.2 발음열 분석 및 발음열 태깅 도구 개발

이 시스템은 형태소 분석을 선행한 후, 한국어에서 빈번하게 발생하는 음운변화 현상의 분석을 통해 정의된 음소 변동 규칙과 변이음 규칙을 선택적으로 적용하여 발음열을 생성한다. 규칙의 적용은 테이블 참조 방식의 오토마타를 사용하여 처리 시간을 줄였다. 또한 정교한 모델을 구축하기 위하여 실제 발화된 음성과 비교하여 결과를 비교 조정하고 이를 통해 발음열 생성기의 안정화 작업을 수행하였다.[2]

또한 대화체 발음열 생성시 고려되어야 하는 사항 중에 하나는 대화 현상으로 인한 비음운 규칙 발생이다. 대화체 음성의 경우, 낭독체 문장과는 달리 비표준 발음으로 발성하는 경우가 빈번하게 발생한다. 표준 전사 표기에서도 제기된 바와 같이 표기 형식은 '비표준발음(발성표기)/표준발음(표준표기)'으로 기재한다. 발음열 생성도 역시 실제 발음된 형태의 발음열 리스트와 표준 발음에 의한 발음열 리스트 두 가지에 대해서 작성하여 음향모델 학습시 정확한 모델을 얻을 수 있도록 작성하였다.

### IV. 언어정보 분석 및 태깅도구 개발

#### 4.1 형태소 태깅

본 연구에서는 규칙에 의한 방법으로 형태소 분석을 한 뒤, 그 결과를 사용하여 다시 통계적 처리에 의해 형태소 태깅을 수행하는 시스템을 구현함으로써 규칙의 확장에 따른 어려움과 견고성 및 일관성 유지의 문제를 해결하고자 하였다. 통계적 처리 모델은 기본적으로 은닉 마코프 모델(HMM)을 사용했으며, 형태소의 모호성을 해결하기 위하여 Viterbi 알고리즘을 이용했다. 원래, 통계적 방법을 구현한 시스템은 엄청난 양의 계산 시간을 필요로 하는 데 반해, 규칙과 통계의 혼합적인 방법을 통해 구현된 형태소 분석 및 태깅 시스템은 통계적 방법에서 요구되었던 많은 양의 계산 시간을 줄일 수 있을 뿐만 아니라 규칙에 의거한 방법에서 단점으로 지적되어진 규칙 확장의 문제와 견고성 및 일관성 유지에 대한 문제를 해결할 수 있다는 장점이 있다.

본 연구에서 사용한 형태소 분석 태그 집합은 총 51개의 tag set으로 구성되어 있으며 "국어 정보 처리 기반 구축을 위한 연구"에서 설정된 표준 태그 집합을 기본으로 사용하였으며, 추가적으로 구어체 문장을 분

석하기 위하여 구어체에서 자주 나오는 고유 명사를 처리하기 위한 태그들을 추가하여 정의하였다. [3]

## 4.2 구문 태깅

### 가. 의존 문법

한국어는 문법적인 성격을 표현하는 방식으로 기능이 발달한 언어이다. 이로 인해 문장을 이루는 문장 성분들 간의 위치적인 정보는 비교적 중요한 의미를 갖지 않으므로 부분적으로 자유로운 어순을 허용한다. 따라서, 문장 구조를 문장 성분들 사이의 상대적 위치에 의해 정의하는 구구조 문법은 영어와 같이 위치적 제약이 강한 언어의 분석에는 적당하지만 한국어와 같이 부분 어순의 자유성을 가지는 언어의 분석에는 적당하지 않다. 한국어의 이러한 특성을 반영하여 최근 문장 성분들 간의 상대적인 위치보다는 해당 성분들 간의 수식 관계를 중심으로 문장 구조를 표현하는 의존 문법이 한국어 구문 분석에 많이 사용되고 있다. 의존 문법은 문장 성분들 간의 의존 관계를 기술하는 문법 이론으로 문장 내의 서로 다른 등급의 성분들 중에 지배 성분에는 어떤 것이 있으며, 또 이 지배 성분에 결합되어 있는 의존 성분에는 어떤 것이 있는가를 기술한 문법이다. 이러한 의존 문법은 두 성분들 간의 방향성을 갖는 이진 관계로 표현되며, 종속 성분들 간의 순서가 무시된다는 특징을 가지고 있다.

본 연구에서는 한국어 특성에 맞는 의존 문법을 사용하는 변환 규칙 학습기를 포함한 한국어 의존 구조 분석기를 구현하여 사용하였다.

## 4.3 의미 태깅

대화 인식 시스템에서 중요한 작업 중에 하나는 발화 문장의 핵심어를 찾아내는 것이다. 특히, 대화 인식에서는 이러한 핵심어의 대부분은 인명, 지명, 조직명, 상품명, 날짜, 시간 등의 개체명이다. 이와 같은 개체명 인식은 발화 문장에서 개체명을 추출하고 이의 범주를 결정하는 것을 말한다.

대화 인식 시스템과 같이 영역이 제한되는 경우에는 나타나는 개체명의 단어의 종류나 그 양상이 규칙성을 띄게 된다. 따라서 여러 영역이나 타 언어에 대한 고려가 있기 전의 개체명 연구들은 규칙 기반이 대부분이었으며 가장 좋은 성능을 보여주고 있었다. 또한 요즘에 주로 이용되는 확률기반 접근 방법은 대규모의 수동으로 구축된 코퍼스를 요구한다는 점에서 목표가 되는 반자동 태깅 툴과 같이 코퍼스 구축을 위한 도구를 만드는 데는 적절하지 못하다.

따라서, 본 개체명 인식기[4]는 두 명의 화자의 발화를 대상으로 그곳에서 발견된 개체명 단어들을 기반으로 하여 규칙기반 개체명 반자동 태깅 툴을 사용하였

다. 여기에는 Regular expression을 가장 쉽게 표현할 수 있다고 알려진 언어 중 하나인 Perl을 사용하였으며 개체명 수정을 위한 툴은 구축된 태깅 시스템의 결과를 사람이 수정할 수 있도록 구성되었다.

## 4.4 담화 분석

### 가. 기본 화행 집합

발화 속에는 대화 목적을 수행하기 위해 화자에 의해 의도된 행위가 있게 되는데, 발화 속에 포함된 이러한 행위를 화행이라고 한다. 대화 상황들을 올바르게 분석하고 이에 대한 적절한 응답을 하기 위해서는 대화 속에 포함된 해당 발화의 정확한 화행을 파악해 내야한다.

이러한 화행의 종류를 일반적인 모든 한국어 문장에 대해서 결정하는 것은 매우 어려운 일이다. 그러므로, 본 연구에서는 예약 영역(호텔예약, 항공예약)에서 연구되어온 화행의 종류를 바탕으로 17개의 화행을 정의하고 사용하였다.

나. 지지 벡터 기계(SVM)을 이용한 화행 분석 시스템의 개발[5]

본 연구에서는 화행 분석을 위해서 현재 분류 영역에서 가장 좋은 성능을 보이고 있는 지지 벡터 기계(SVM)을 이용하여 화행을 분석하는 시스템을 개발하였다.

본 시스템은 2 단계로 구성되는데, 첫 번째 단계는 대화 코퍼스로부터 대량의 자질을 추출하고, 추출된 자질 중 효과적인 자질을 선택하는 단계이다. 먼저 우리는 형태소 분석 시스템을 이용해 문장 자질을 생성하고 대화 코퍼스의 각 발화에 부착된 정답 화행을 이용해 문맥 자질을 생성한다. 그리고, 자질 선택 시스템을 통해 정보량이 많은 자질들이 선택된다. 자질 선택을 위해서 카이 정보 통계량을 이용하여 자질의 정보량을 계산하고 정보량이 많은 자질을 선택하였다.

두 번째 단계는 선택된 자질을 바탕으로 화행을 결정하는 단계이다. 화행 결정 시스템은 선택된 자질과 대화 코퍼스를 바탕으로 지지 벡터 기계를 학습하고, 사용자의 입력이 들어왔을 때 학습된 데이터를 바탕으로 화행을 결정하게 된다.

## V. Database 수집결과 및 예제

### 5.1 대화음성 DB 수집결과

1,2 차년도 수집결과 300명 화자의 마이크 및 전화 녹취 대화음성을 각각 65시간씩 수집했다. 이는 약 7만 5천 발화로 구성되어 있으며 전체 발화에 대해서 전사를 수행 하였다.

추가적인 언어정보 태깅은 69명 화자의 20,139발화

를 대상으로 수행하였다.

## 5.2 예제

**■ 시스템 스크립트**  
환율 정보 시스템입니다. 궁금한 점을 질문하세요.  
원화 100원 당 162원 92전입니다.  
고객님은 학생이십니까?  
총 환전 수수료의 50%를 할인해 드립니다.  
16032달러입니다.

**■ 발화자 스크립트**  
현재 홍콩 달러의 환율 조회(반말로)  
환전 수수료 가격을 물어본다.(반말로)  
궁경의 대답(반말로)  
100만원을 홍콩 달러로 환전하면 얼마인지 묻는다.(반말로)  
처음으로

**■ 전사결과**  
p042-s4-2-u1-00.pcm : 현재 홍콩 달러의/달러의 환율을 알고 싶어.  
p042-s4-2-u1-01.pcm : 환전 수수료 가격은 얼마니?  
p042-s4-2-u1-02.pcm : 응.  
p042-s4-2-u1-03.pcm : 백만 원을 홍콩 달러로/달러로 환전한다면 얼마니?  
p042-s4-2-u1-04.pcm : 처음으로 가자.

**■ 발음열 태깅**  
2. [현재 홍콩 달러의 환율을 알고 시퍼]  
H JX N ZH EH HI OW NX KH OW NX TT AA L L  
AX WI HI WA N JU R WW R AA L G OW S IY PH  
AX

**■ 형태소 태깅**

2. 현재	현재/ma	
홍콩	홍콩/nq_loc	
달러의	달러/nbu	의/jcm
환율을	환율/ncn	을/jco
알고	알/pv	고/ecx
싫어	싫/px	어/ef /sf

**■ 구문 태깅**

2. mod	1	2	현재/ma
mod	2	3	홍콩/nq_loc
mod	3	4	달러/nbu+의/jcm
obj	4	5	환율/ncn+을/jco
aux	5	6	알/pv+고/ecx
root	6	0	싫/px+어/ef+./sf

**■ 의미(개체명) 태깅**

2. 현재	현재/ma	
홍콩	홍콩/nq_loc	
달러의	달러/nbu	의/jcm
환율을	환율/ncn	을/jco
알고	알/pv	고/ecx
싫어. 싫/px	어/ef	./sf
#		
# LOCATION :	홍콩	

**■ 화행 태깅**  
2. UT : 현재 홍콩 달러의 환율을 알고 싶어.  
SA : Ask-ref

## VI. 결론

본 연구에서는 대화음성 DB 및 언어정보가 태깅된 말뭉치 구축을 하였으며 그 과정에서 발음열, 형태소, 구문 및 의미정보 태깅 도구를 개발하고 개선하였다.

대화음성인식 시스템의 개발을 위해서는 좀 더 다양한 영역별로 데이터가 수집되어야 할 필요성이 있다. 본 연구에서는 대용량의 대화음성 DB의 수집을 수행하면서도 발전된 대화 시스템 개발에 필요한 언어적 지식을 체계적으로 태깅하기 위한 표준안을 제안함으로써 향후 다양한 영역의 한국어 대화음성 DB구축에 참고가 되고자 한다.

향후 대화음성시스템의 개발을 위한 DB구축에서는 사용자의 컴퓨터에 대한 자연스러운 음성대화과 반응을 이끌어내기 위한 보다 체계적인 시나리오 구축이 필요하고 언어지식 태깅 방법의 표준화 정착을 위한 연구가 추가적으로 수행되어야 한다.

## 감사의 글

본 연구는 산자부 산업기초기술연구개발사업의 지원으로 수행되었습니다.

## 참고문헌

- [1] Oviatt, "User-centered design of spoken language and multimodal interfaces", IEEE Multimedia, 1996, vol. 3, no. 4, pp 26-35
- [2] 이경남, 정민화, 연속음성 인식에서의 발음사전 최적화를 위한 변이음 규칙 적용 방법, 음성통신 및 신호처리 학술대회, 2002년, pp. 149-152,
- [3] 이상호, 오영환, 서정연, KTS: 미등록어를 고려한 한국어 품사 태깅 시스템", 제 12회 음성통신 및 신호처리 워크샵 논문집, p.195-199
- [4] Choong-Nyoung Seon, Youngjoong Ko, Jeong-Seok Kim, and Jungyun Seo, Named Entity Recognition using Machine Learning Methods and Pattern-Selection Rules", Proc. of NLPRS'2001, 2001, pp 229-236
- [5] Kyongsun Kim, Jungyun Seo, Feature Selection in Automatic Speech Act Tagging, Proc. of ICCPOL'2001, 2001, pp 379-383