ETRI

전자통신동향분석 제26권 제5호 2011년 10월

다국어 자동통역 기술동향 및 응용

The Trends and Application of Multilingual Automatic Speech Translation Technology

소프트웨어 기술의 미래전망 특집

김승희 (S.H. Kim) 자동통역연구팀 선임연구원
윤 승 (S. Yun) 자동통역연구팀 선임연구원
조훈영 (H.Y. Cho) 자동통역연구팀 선임연구원
최승권 (S.K. Choi) 언어처리연구팀 책임연구원
김상훈 (S.H. Kim) 자동통역연구팀 책임연구원

목 차

근래에 국가 간 교류가 한층 활발해지고 스마트폰이 급속히 보급됨에 따라 일반인들이 쉽게 자동통역 제품을 접할 수 있게 되었다. 자동통역 기술은 1990년대부터 세계 각국에서 정부차원의 연구개발 단계, 시범 서비스 및 실용화 연구개발 단계를 거쳐, 이제는 수십 개의 언어를 지원하는 스마트폰용 자동통역 앱이 소개되는 단계에 이르렀다. 본 고에서는 다국어 자동통역의 핵심 요소기술에 대해 설명하고, 최근 자동통역 기술의 개발 동향 및 응용 사례에 대해 기술한다.

Ⅰ. 서론

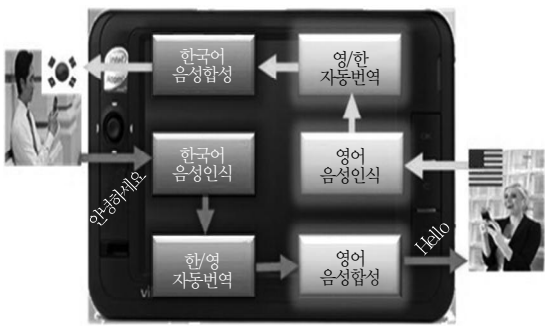
자동통역(automatic speech translation) 기술은 언어장벽을 허물어 서로 다른 언어를 사용하는 사람 간에 대화가 가능하게 하는 기술이다.

현재까지는 두 사람 간의 대화를 통역하는 형상이나, 조만간 서로 다른 언어를 사용하는 다수의 사람들 간의 대화가 가능해 질 것이다.

자동통역을 실현하기 위해서는 여러 요소기술들을 개발하여야 한다. 그리고, 적어도 두 개 이상의 언어를 처리해야 하므로 각각의 언어에 대하여 요소기술들을 개발하여야 한다. 다국어 간 자동통역에 필요한 기술을 효율적으로 개발하기 위하여 국제 자동통역 공동연구 컨소시엄, 이른바 C-TAR(Consortium for Speech Translation Advanced Research)가 1995년 결성되어 관련 연구를 진행하고 있기도 하다.

(그림 1)은 양방향 자동통역 시스템의 일반적인 구조를 나타내고 있다. 양방향 자동통역 시스템은 6개의 주요 모듈로 구성되는데, 각각의 언어를 인식하는 음성인식 모듈, 자동번역 모듈, 음성 합성 모듈 등이 있다.

본 고에서는 다국어 지원을 염두에 두고 자동통역 기술을 구성하는 각 요소기술 및 핵심기술 개발 동향 및 응용 사례에 대해 알아본다.



(그림 1) 한-영 양방향 자동통역 시스템의 구조

Ⅱ. 다국어 자동통역 기술동향

1. 개요

자동통역 기술은 크게 음성인식(automatic speech recognition), 자동번역(spoken language translation) 및 음성합성(text-to-speech synthesis)의 세 가지 요소기술로 구성된다. 그리고, 응용 분야에 따라서는 통역 품질을 높이기 위하여 자연언어 이해(natural language understanding) 기술이 추가되기도 한다.

여러 가지 방식의 요소기술들이 개발되어 왔으나, 현재까지 대체로 확률통계 방식의 기술들이 우수한 성능을 보이고 있다. 다만, 확률통계 방식에서 성능을 높이기 위해서는 많은 양의 양질의 데이터를 수집/가공해야 하는 단점이 있다.

자동통역 시스템은 다수의 주요 요소로 구성된다. 지원하는 언어가 많아질수록 구성 요소의 수는 급속히 늘어난다. 단방향 자동통역 시스템의 경우, 한-영 자동통역 시스템을 예로 들면, 한국어 음성인식 모듈, 한→영 자동번역 모듈, 영어 음성합성 모듈 등 3개의 주요 요소로 구성된다. 양방향일 경우 앞에서의 예처럼 6개로 구성된다. 지원하는 언어에 일본어가 추가될 경우 일본어 음성인식/합성, 한→일/일→한/영→일/일→영 자동번역 등 6개의 모듈이 추가되어 총 12개로 구성된다. 따라서, 다국어를 지원하는 자동통역 시스템을 개발하는 경우에는 개발 비용 및 기간을 줄이면서 고성능을 낼 수 있는 기술들이 필요하다.

자동통역이 주로 사용되는 응용 분야가 두 사람 간의 대화를 통역하는 것이므로, 하드웨어 측면에서 휴대성이 중요한 요소 중 하나다. PDA나 PDA 형태의 전용 단말기에 탑재한 시스템이 출시되고 사용되기도 하였다[1]. 그러나, 단말에서는 제한된 컴퓨팅

자원으로 인해 고품질의 통역 서비스를 제공하기가 쉽지 않다. 최근에는 스마트폰의 급속한 대중화, 와이파이 접속 지역 확대 및 다양한 요금제와 같은 환경 변화와 맞물려 서버-클라이언트 기반의 통역 시스템들이 출시되고 있으며, 예전보다 훨씬 개선된 통역 성능을 체감할 수 있다.

2. 음성인식

음성인식 시스템은 크게 음성인식 엔진과 음향 모델, 언어 모델로 구성된다.

음성인식 기술 자체는 언어 독립적인 성격이 강하다. 현재 적용되는 음성인식 기술은 거의 확률통계 방식인 은닉 마르코프 모델(HMM: Hidden Markov Model) 기반이다. 개별언어에서 인식 성능을 높이기 위해서 언어 종속적인 정보들을 활용하지만, 기반이 되는 알고리즘이 특정 언어를 고려하지는 않는다. 특히 음성인식 엔진에 사용되는 기술은 거의 언어와 무관하다고 할 수 있다.

반면 확률통계 모델을 만들기 위해서는 많은 양의 언어(음향, 텍스트) 데이터가 필요하다. 음성인식 성능과 데이터베이스의 크기 사이에는 양의 상관관계가 있으며, 데이터베이스의 품질에 따라 성능이 달라진다.

최근의 음성인식 성능은 프로세서의 고속화, 메모리 양의 증가, 병렬처리 기법, 음성언어 자원의 증가 등으로 인해 지속적으로 향상되고 있다. 음성인식 시스템은 서버급 컴퓨터로부터 소형 휴대 단말기 또는 가전기기 등과 같이 다양한 하드웨어 플랫폼 상에 탑재가 되고 있다.

자동통역용 음성인식 시스템은 크게 서버형과 단말형으로 구별되어 연구개발되고 있으며, 기술 수준이 점차로 발전함에 따라 보다 짧은 개발 기간에 제한된 양의 음성언어 자원을 이용하여 다국어 확장을 할 수 있는 음성인식 기반 기술을 필요로 하고 있다[2].

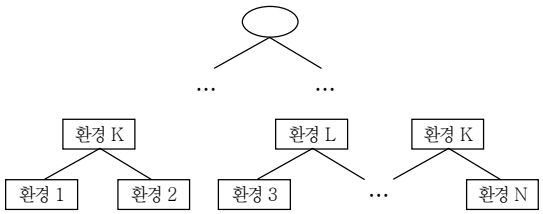
본 절에서는 서버형 및 단말형 음향 모델링, 언어 모델링 기법들과 다국어 확장 연구방법을 순서대로 기술한다.

가. 서버형 및 단말형 음향 모델링

서버형 음성인식은 기본적으로 사용 가능한 음성언어 및 컴퓨팅 자원이 무제한이라고 가정하며, 단말기에서 사용자의 발화에 대해 음성 끝점 검출(EPD: End point detection)을 수행하여 녹음된 음성신호 혹은 음성신호에서 추출한 특징 벡터들을 서버로 전송하고, 그 이후의 단계를 서버에서 처리하여 인식 결과를 다시 단말기로 전송하는 방식이다. 구글 음성인식 서비스의 경우, 수백만 내지 수천만 발화 이상의 방대한 규모의 음성 데이터베이스를 음향 모델링에 사용하고 있으며, 서비스가 진행됨에 따라 그 규모는 지속적으로 증가하고 있다.

자동통역에서 이처럼 방대한 음성 데이터를 필요로 하는 이유는 자동통역 단말기가 사용되는 음향 환경에 존재하는 수많은 종류의 잡음에 강한 음향 모델을 생성하기 위해서이다. 서버형 음성인식에서는 학습용 음성 데이터에 부과된 녹음 환경 레이블링 정보를 사용하여 다수 개의 소음 환경별로 음향 모델을 각각 학습하고, 음성인식 단계에서 이들 중 입력신호와 가장 가까운 모델을 자동 선택하여 인식 성능을 높이는 방식이 매우 효과적으로 알려졌다.

최근에는 음성인식 서비스를 통해 자동 수집되는 사용자 로그 데이터를 활용하여 인식기의 성능을 높이는 기법들이 개발 및 적용되고 있다. 이 경우에는 음성 데이터가 녹음된 환경에 대한 정보가 제공되지 않을 수 있으며, 이 경우 비교사 군집화(unsupervised clustering) 알고리즘 등을 통해 환경별 음향 모델을



(그림 2) 환경 군집화 서버형 음향 모델링 기법

구축할 수 있다[3]. (그림 2)는 구글 등에서 제안한 비교사 군집화 기반의 음향 모델링을 나타낸다. 그림의 종단 노드(terminal node)들은 각각의 독립적인 음향 모델을 이루며, 음성인식 입력신호는 이 중에서 가장 유사한 모델을 찾아 최적의 인식 성능을 얻는다.

단말형 음성인식은 스마트폰 등의 단말기에서 음성인식의 전 과정이 수행되는 방식으로, 사용 가능한 메모리, CPU 속도 등에 어느 정도의 제약이 있다. 이 방법은 서버형 음성인식을 위한 음성언어 리소스 자체 또는 모델링 단계에 제약을 가하여 단말기의 컴퓨팅 능력을 최대한 활용하는 방향으로 최적화된다. 음향 모델링 단계에서는 서버형 음성인식용 음향 모델을 구축한 뒤, 이 모델의 규모를 축소하면서도 성능을 최대한 유지하는 기법을 사용한다.

HMM 기반의 음성인식 방법에서 음향 모델의 규모를 변경하는 것은 음향 모델을 구성하는 모든 HMM들의 모든 평균벡터 및 공분산 행렬 요소들의 총 개수를 늘리거나 줄이는 것을 의미한다. 음성인식 과정의 전체 연산량 중에서 음향적 우도 값(acoustic likelihood score) 계산은 절반 이상을 차지하기도 하므로, 음향 모델의 규모를 변경시키는 것은 모델을 저장하는 저장공간의 크기뿐만 아니라, 음성인식 속도와도 밀접한 연관성을 갖게 된다.

VM-GMM 모델링(variable mixture Gaussian mixture model) 방법은 HMM 기반의 음성인식에서 음향 모델의 파라미터 수를 조절하기 위한 방법으로 주어진 음향 모델 학습용 데이터에 대해 우선 충분히 많은 수의 모델 파라미터를 이용하여 음향 모델을 학습한 뒤, 각 HMM 상태의 가우시안 성분(Gaussian mixture component)들을 점차적으로 줄여나가는 방식이다. 이 방식은 확률적 분포가 가장 유사한 두 개의 가우시안 성분들을 점진적으로 통합하여 이진 트리를 구성한 뒤, 트리를 적정 수준에서 가지치기(pruning)하여 최적의 음향 모델을 생성해 낸다. 확률적으로 유사한 두 가우시안 성분들을 찾기 위한 거리 척도는 구축되는 이진트리의 품질을 결정하는 중요한 단계로서, 가중 K-L 거리(weighted Kullback-Leibler divergence), Bhattacharyya 거리, 가우시안 성분들의 가중치(mixture weight)의 합, DL(delta-likelihood) 거리 척도 등이 연구되어 왔다. 이진 트리를 가지치기하는 방법으로는 MDL(minimum description length) 기반의 최적화 기법이 많이 사용되고 있다[4],[5].

나. 서버형 및 단말형 언어 모델링

서버형 음성인식에서 언어 모델 학습 코퍼스의 규모는 구글 음성검색용 음성인식기의 경우 2300억 어절에 육박한다. 이 코퍼스에 대한 n-gram을 기준으로 할 때, 3-gram이 77억 개 정도이며, 음성인식이기에서 처리 가능하도록 이를 1500만 개로 줄여서 음성인식의 첫 번째 단계에서 단어 후보들을 격자(lattice) 형태로 출력하며, 이차적 인식에서 4-gram 혹은 5-gram 언어 모델을 사용하여 최종 인식 결과의 정확도를 극대화한다. 여행영역으로 제한된 음성인식기의 경우, 무제한 영역인 음성검색에 비해서는 상대적으로 적은 규모의 텍스트 코퍼스를 언어 모델링에 사용하고 있으나, 점차적으로 그 규모는 증가하고 있어 언어 모델링에 대한 접근 방식이 음성검색과 유사하다고 볼 수 있다[6],[7].

단말형 음성인식의 언어 모델은 서버형 언어 모델

용 자원을 그대로 활용하되, Stolcke, Seymore-Rosenfeld 등의 다양한 가지치기 기법과 Katz/Good-Turing, Witten-Bell, Kneser-Ney 등의 평활화(smoothing) 방법을 이용하여 n-gram 규모를 크게 축소시킨다. n-gram의 개수는 음성인식기의 물리적 크기와 밀접한 연관이 있어, 단말기의 메모리 크기에 적합한 수준으로 n-gram 개수를 조정한다.

이외에도 호텔예약 상황 등과 같이 음성인식의 영역이 제한적일 경우, 언어 모델 자체를 세부 영역으로 적응하는 언어 모델 적응(language model adaptation) 기법들을 활용하여 단말형 음성인식기의 크기와 성능을 최적화할 수 있다.

다. 다국어 확장 기술

음성인식기술의 발전에 따라 인식 성능이 크게 개선되고 성공적인 응용 서비스들이 나오기 시작하면서, 점차 전 세계 주요 언어들로의 확장에 대한 필요성이 대두되고 있다[8]. 이 문제는 새로운 언어에 대한 음성인식 시스템을 저비용, 고속 및 고성능으로 구축하는 것을 포함한다. 보다 자세히는 신규 언어에 대한 음향 모델 구축용 음성 데이터베이스, 언어 모델 구축을 위한 텍스트 코퍼스, 발음 사전 생성기, 형태소 분석기 및 언어별 발음 특성에 맞는 특징 추출기 등이 필요하고, 이러한 자원들을 음성인식기에 반영하기 위한 다양한 가공 도구들이 있어야 한다.

최근에 음성인식 서비스가 스마트폰과 인터넷 환경에서 널리 사용되어짐에 따라, 음성인식기 개발에 필요한 음성언어 자원들을 온라인 상에서 다수의 사용자들을 통해 직접 수집하는 방식이 많이 시도되고 있다[9]. 로그 데이터 형태로 저장된 사용자의 음성 데이터는 수작업 전사(transcription) 작업 또는 음성인식을 통한 자동 전사 및 신뢰 정보를 통한 필터링 등을 통해 음성인식기 학습에 유용한 형태로 확보된다. 따라서, 신규 언어에 대해 일정 수준 이상의 인식 성능을 갖는 초기 인식 시스템을 구축할 수 있다면, 여러 형태의 음성인식 서비스를 통해 보다 많은 음성언어 자원을 획득할 수 있다.

ETRI의 자동통역연구팀에서 2008년부터 2011년까지 4년간 수행되고 있는 휴대형 한영 자동통역기 개발과제에서는 여행영역에 대해서 서버형 및 단말형 한영 음성인식 시스템이 개발되었으며, 점차로 일본어, 중국어 등으로 언어를 확장해 나가는 연구개발을 수행할 예정이다[5],[10].

3. 다국어 자동번역

다국어 자동번역은 그 자체로 하나의 완전한 기술일뿐만 아니라 다국어 자동통역 기술의 하부기술이기도 하다. 다국어 자동번역은 시기적으로 1980년대까지는 중간언어 방식과 간접변환방식에 의해, 1990년대에는 통계기반방식에 의해 개발되다가 현재는 간접변환방식과 통계기반방식을 혼합한 하이브리드 방식으로 발전하고 있다[11]. 다국어 자동번역을 위한 중간언어방식, 간접변환방식, 통계기반방식에 대해 살펴보면 다음과 같다[12].

가. 중간언어방식

중간언어방식(Interlingua approach)은 개별언어 독립적인 의미표상(language-independent semantic representation)을 도입하여, 입력문을 분석단계를 거쳐 이 언어독립적인 의미표상으로 매핑하여 번역하는 방식이다. 따라서 n개의 언어를 번역하기 위해서는 각 언어로부터 언어독립적인 의미표상을 도출해낼 수 있는 n개의 분석 모듈과, 이 언어독립적인 의미표상으로부터 목표 언어를 생성해 낼 수 있는 n개의 생성 모듈만이 필요하다.

중간언어방식에서 사용하는 의미표상은 자연언어에서 가능한 모든 의미를 표상할 수 있는 구조적인 장치와 어휘들을 가지고 있어야 한다. 이 방법론의 성공 여부는 자연언어의 모든 의미들을 어떠한 방법으로 표상할 것인가에 달려있다고 볼 수 있다. 일반적으로 의미를 표상하는 방법에는 다음과 같은 것들이 있다.

- 자연언어의 의미를 표현하기 위해 인공언어를 사용(예: TITUS-System)[13]
- 영어와 같은 자연언어를 차용하여 의미를 표현(예: UNL-System)[14]
- 자연언어의 의미를 더 이상 쪼갤 수 없는 단위(의미소)로 나누어 표현(예: UNITRAN-System)[15]

이 방식의 가장 큰 단점은 언어독립적인 의미표상을 정의하는 것이 매우 어렵다는 점이다. 예를 들어 한국어의 '김장'이라는 단어는 한국과 다른 문화권의 언어, 예를 들어 영어나 독일어 등과 같은 다른 외국어에서는 정확하게 대응되는 개념이 없다. 따라서 언어보편적인 의미표상을 구축하려고 할 때 이러한 어휘를 누락하게 되는 경우가 발생하는 문제가 있다.

나. 간접변환방식

간접변환방식(Indirect transfer approach)은 형태소 분석을 거쳐 통사구조(syn-tactic structure), 의미구조(semantic structure) 분석을 한 후 목표언어로 변환(transfer)을 하며, 이 변환된 구조로부터 대역문장을 생성하게 된다. 변환을 통사 단계에서 하느냐 의미구조에서 하느냐에 따라 변환 규칙의 복잡성도 달라지게 된다. 일반적으로 통사구조는 사용하는 문법에 따라 다르게 표상화(repre-sentation)되는데, 주로 구구조 문법(phrase structure grammar)이나 의존 문법(dependency grammar) 등이 통사표상을 위해 사용된다. 의미구조는 통사적으로 분석된 문장의 의미구조를 나타내는 구조로서, 일반적으로 술어논리구조(predicate argument structure)가 표상으로 사용된다.

이 방식에서 번역지식은 변환규칙의 형태로 표현된다. 변환단계에서 입력문장의 인터페이스 구조는 목표 문장의 인터페이스 구조로 변환되게 된다. 통사구조기반 변환 방식은 의미구조기반 변환방식에 비해 분석을 위한 비용과 시간은 덜 소모되나, 변환규칙이 복잡해지는 단점이 있다. 이에 반해 의미구조기반 변환 방식은 통사구조기반 변환 방식에 비해 변환 규칙은 비교적 단순하나, 변환을 위한 인터페이스 구조를 도출해내기까지의 비용과 노력이 많이 들 수 있으며, 그만큼 에러 발생의 가능성도 높아지는 단점이 있다.

그러나 이 방식은 비교적 개발이 용이하고, 소수의 규칙만을 구축하더라도 비교적 높은 성능을 낼 수 있으므로[16], 현재 국내외에서 상용화되어 판매되고 있는 대부분의 다국어 자동번역 시스템에 채택되고 있다. 간접변환방식을 채택한 대표적인 기계번역시스템으로는 캐나다 몬트리올 대학의 TAUM-시스템, 독일 자르브뤼켄 대학의 CAT2 시스템[17], 미국 텍사스 대학의 METAL 시스템 등을 들 수 있다.

이 방식의 대표적인 문제점으로 들 수 있는 것은 분석, 변환, 생성에서 많은 수의 규칙에 의존하므로 규칙 수가 많아질 수록 규칙 간의 충돌이 생기는 경우가 많다. 또한 많은 수의 규칙을 효과적으로 관리하는 문제도 발생할 수 있다.

다. 통계기반방식

통계기반방식(Statistics-based approach)은 이중언어 말뭉치로부터 통계적 분석을 통해 모델의 파라미터를 학습하고 그 모델에 근거하여 입력된 문장을 번역하는 방식이다[18]. 이 방식은 1949년 War-ren Weaver[19]에 의해 소개된 이후, 1991년 IBM

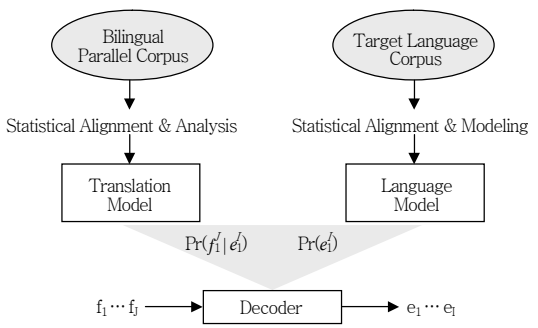
의 Thomas J. Watson 연구소 연구원들에 의해 다시 소개되면서부터 연구가 부활하여[20], 현재 가장 활발하게 연구되는 기계번역 기술이다.

통계기반방식이 활발히 연구되는 배경에는 ① 모델 파라미터를 학습할 수 있는 대용량의 가용 말뭉치가 구축되고 있으며 ② 특정 언어 쌍에 제한 받지 않고 모델을 자동으로 학습해 낼 수 있으며 ③ 규칙기반자동번역이 번역지식을 구축하는 데 상당한 비용을 요구하고, 다른 언어들에 일반화시켜 적용하기 어렵다는 문제가 있다는 것이다.

통계기반방식의 기본 요소는 통계적 번역 모델(translation model)과 언어 모델(language model), 이중언어 말뭉치(bilingual parallel corpus)로부터 은닉된 번역지식 파라미터를 찾아내는 학습 알고리듬, 그리고 학습된 번역 모델에 기반하여 최적의 번역결과를 탐색하는 디코딩 알고리즘(decoder)으로 구성된다((그림 3) 참조).

기본적으로 원시언어 문장 f를 목적언어 문장 e로 번역하는 확률 모델은 p(e|f)인데, 자연스러운 번역결과를 얻기 위해 Bayes Theorem을 적용하여 번역모델 p(f|e)과 언어 모델 p(e)의 결합으로 유도된 생성 모델을 만든다

통계기반방식의 기본 모델인 단어단위 모델(IBM model 1-5)은 문장 길이에 따라 어순 재배열 계산



(그림 3) 통계기반방식 기본 구성도

복잡도가 너무 높기 때문에 계산 복잡도를 낮추기 위한 많은 연구들이 시도되었고, 특히 2003년 어순 재배열에 따른 계산복잡도 감소 및 번역 효율을 고려한 구문단위 번역 모델[21]이 소개되면서 기술이 급격히 발전하여 현재 state-of-the-art를 이루었다. 그리고 최근에는 통사적 언어 구조를 모델에 접합시키기 위한 연구[22],[23], 미등록어 및 관용적 숙어 표현의 효과적 처리를 위한 paraphrasing 기법 또는 번역 지식 일반화 등에 대한 연구가 진행 중이다[24],[25]. 통계기반방식은 중국어-영어, 아랍어-영어 뉴스 자동번역 분야에 적용한 결과 Google, IBM, ISI의 자동번역 시스템들은 실용적 수준에까지 이르고 있는 것으로 보고되고 있다[26].

라. 국가별 기술동향

미국에서는 미국방위 고등연구 계획국(DARPA: Defense Advanced Research Agency)과 미국국립과학재단(NSF: National Science Foundation)을 중심으로 개발 위험도가 높더라도 전망이 큰 기술을 대상으로 통계기반방식의 다국어 통번역 기술을 개발하고 있다[27],[28].

일본에서는 일본 정부의 국제무역산업부 지원 하에 CICC(Center of the International Cooperation for Computerization) 프로젝트에서 1987년~1997년까지 일본어, 중국어, 말레이시아어, 태국어, 인도네시아어와 같은 동양권 언어에 대한 간접변환방식의 다국어 자동번역 시스템을 개발한 바 있다[29].

유럽은 유럽 공동체의 특수성으로 인해 가장 다국어 자동번역이 발달되어 있는 곳으로 모든 EU 언어 쌍을 자동번역하기 위해 도구, 언어자원, 평가 등의 개방형 연구 환경을 만들고 통계적 기술, 언어학적 지식 자원, 통계기반과 규칙기반 자동번역을 모두 통합하는 하이브리드 자동번역 아키텍처를 개발하여 사

용자 지향적 자동번역 기술을 목표로 FP7(FP7 Seventh(7th) Framework Programme)의 지원 하에 2012년까지 간접변환방식과 통계기반방식을 혼합한 하이브리드 방식의 EuroMatrixPlus Project를 추진 중에 있다[30].

이밖에 유럽에서는 온라인 협업 플랫폼을 제공함으로써 통계기반 자동번역을 위해 대량의 학습데이터를 온라인으로 공유할 수 있도록 하는 LetsMT![31] 프로젝트를 추진 중에 있다.

4. 음성합성

자동통역의 마지막 단계인 음성합성은 목적 언어(target language)로 번역된 텍스트를 사람들이 들을 수 있는 말의 형태로 제공하는 기술이다.

서버형의 시스템에서는 대용량의 녹음 데이터를 가공한 후 음소 단위를 반자동으로 추출하여 합성단위로 사용하는 코퍼스 기반 음편 조합방식의 음성합성 기술이 널리 사용된다. 계산량이나 메모리 사용량이 많지만, 고품질의 합성음을 만들 수 있다.

일반적으로 고품질의 합성음을 생성하기 위해서는 30시간 이상의 녹음 데이터가 필요하다. 이 때문에 다국어 합성기나 언어 확장을 위한 여러 방법들이 연구개발되어 왔다. Multilingual voice는 다수의 언어를 수용할 수 있는 합성음을 의미하는데, 다국어에 유창한 발화자의 음성을 녹음하여 사용한다. 새로운 언어로 확장할 때 녹음 데이터를 필요로 하지 않거나 많지 않은 데이터를 사용하는 방법도 있다[32],[33]. 이런 방법들은 언어 확장성을 염두에 둔 것으로, 각 언어별로 별도의 녹음 데이터를 확보하여 개발한 합성음에 비해 음질이 좋지 않지만, 짧은 개발 기간 및 비용 측면에서 장점이 있다.

그러나 데이터 규모나 확보 방법 측면에서 음성합성 시스템의 다국어 확장은, 음성인식이나 자동번역의 경우보다는 상대적으로 용이하다.

단말기 내장형 시스템의 경우에는 계산량, 메모리 등의 제약 조건의 중요성이 훨씬 커진다. 휴대폰 단말기의 경우 합성기가 사용할 수 있는 가용자원은 매우 제한적이다. 특히 유럽지역에서 판대되는 휴대폰의 경우에는 다국어 지원이 필수적이며, 하나의 단말기에서 수 개에서 십 여 개의 언어를 지원하는 경우도 있다.

이 때문에 단말기의 특성에 따라서는 편집합성 방식이나 서버형의 시스템에서 널리 사용되는 음편조합 방식보다는 통계적 모델링을 이용한 파라미터 기반의 음성합성 기술이 이용되기도 한다[34].

Ⅲ. 다국어 자동통역 기술 응용

다국어 자동통역 기술의 최근 응용 사례는 스마트폰의 급속한 대중화와 함께 대부분 스마트폰을 위주로 보고되고 있다. 이에 따라 여기에서는 특징적인 대표적 스마트폰 기반 다국어 자동통역기에 대해 소개하도록 하겠다.

1. Jibbigo

Jibbigo는 미국 카네기 멜론 대학과 독일 칼수루에 연구소의 연구자들에 의해 설립된 Mobile Technologies에 의해 개발된 다국어 자동통역기이다. 이들은 20여 년간의 자동통역 기술 연구 경험을 기반으로 휴대형 자동 통역기가 비교적 대중화되지 않았던 시기인 2009년 10월에 시장을 선도하여 Jibbigo를 출시하였다.

Jibbigo의 경우 아이폰과 안드로이드 기반 플랫폼 양쪽 모두에서 동작한다. 여행, 의료분야의 약 40,000 단어에 대해 인식이 가능하며 번역 기술로는 SMT

<자료>: http://www.jibbigo.com

(그림 4) Jibbigo

(Statistical Machine Translation) 방식을 채택하고 있다. 초기에는 영어-스페인어 번역만을 지원하였으나 지원하는 언어 쌍을 점차 확대하여 최근에는 영어와 이에 대응하여 일본어, 중국어, 이라크어, 타갈로그어, 독일어, 프랑스어, 한국어 등의 다국어 번역을 지원하고 있다. 영어 기반이 아닌 언어 쌍으로는 독일어-스페인어 번역을 지원한다. (그림 4)는 Jibbigo에서 다국어 자동통역을 수행한 모습이다.

Jibbigo의 경우 대표적 특징을 살펴 보면 최근의 경향인 서버-클라이언트 통역 방식이 아니라 단말기에서 음성인식 및 번역 과정을 수행한다는 특징을 지니고 있다. 이러한 방식은 데이터 통신에 걸리는 시간을 줄일 수 있고 또한 네트워크에 연결되지 않아도 어디에서나 자동통역이 가능하다는 장점이 있다. 다만 올해 7월부터는 일부 언어 쌍에 대해서 서버 접속 방식의 자동통역도 지원하고 있다. 이를 통해서 언어 학습용으로도 활용할 수 있도록 하는 동시에 무료 배포를 함으로써 사용자들의 구매 전 평가가 가능하도록 하고 있다.

2. VoiceTra

VoiceTra는 일본의 NICT(the National Institute of Information and Communications Technology)가 개발한 다국어 자동통역기이다. 일본 정부가 언어 장벽 극복을 목표로 추진하는 '언어의 벽을 뛰어넘는 음성 커뮤니케이션 기술의 실현' 프로젝트의 일환으로 개발되었으며 2010년 6월에 아이폰 기반으로 최초 발표되었다.

<자료>: VoiceTra 화면 캡처

(그림 5) Voicetra

VoiceTra는 (그림 5)에서 알 수 있듯이 다양한 언어를 지원한다. 발표 당시부터 음성인식의 경우 일본어, 영어, 중국어, 베트남어, 인도네시아어, 말레이시아어를 지원하였고, 번역의 경우 이들 언어를 포함하여 대만어, 독일어, 프랑스어, 덴마크어, 네덜란드어, 이탈리아어, 스페인어, 포르투갈어, 포르투갈어, 러시아어, 아랍어, 힌디어, 태국어, 타갈로그어, 한국어 등의 21가지 언어를 지원하는 등 본격적인 다국어 자동통역기로서의 면모를 보였다. 최근에는 안드로이드 기반 플랫폼으로 서비스를 확장하고 한국어 음성인식도 지원하는 등 점차 그 영역을 확대하고 있다.

VoiceTra의 경우 서버 접속 방식의 다국어 자동통역기이며 번역 방식은 SMT 방식을 채택하고 있다. 자체적으로 밝힌 바에 따르면 번역 성능은 토익

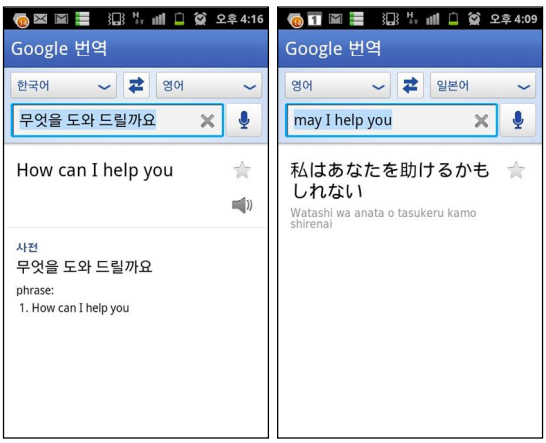
600점급에 해당한다고 말하고 있다. 또한 NICT에서는 VoiceTra 외에도 텍스트 기반 번역기인 TexTra, 음성 대화에 기반한 관광 도우미 AssisTra 등과 같은 앱도 개발 중이다.

3. Google

다국어 자동 통역기를 말하는데 있어 빼놓을 수 없는 것이 Google 번역이다. Google 번역은 일반인들에게도 널리 알려진 Google에서 올해 초 출시하였다.

PC 기반의 웹 버전 구글 번역을 모바일 환경에서도 사용 가능하도록 아이폰과 안드로이드용 앱으로 내놓았으며 모바일 출시와 함께 15가지 이상의 언어에 대해 음성인식 기능을 지원하여 단순한 텍스트 입력 방식의 번역만이 아니라 실제적인 음성 입력 방식의 다국어 자동통역이 이루어질 수 있도록 하였다. (그림 6)을 보면 Google 번역에서 자동통역을 수행하는 모습을 참고할 수 있다.

Google 번역의 경우 출시는 비교적 늦었지만 오랜 기간의 웹기반 Google 번역 서비스를 그 근간으로 하기에 번역에 있어서는 현존하는 다국어 자동 통역기 중 가장 많은 언어인 50가지 이상의 언어간 상호 번역을 지원한다. 음성인식의 경우에도 비록 대상 언어는 제한되지만 서버 기반으로 동작하기에 무제한 급에 가까운 인식 대상 어휘를 지니고 있다는 점이 특징이다. 다만 이러한 장점이 오히려 성능 면에서 일부 한계를 가져온다는 점은 Google 번역이 해결해야 할 과제일 것이다.



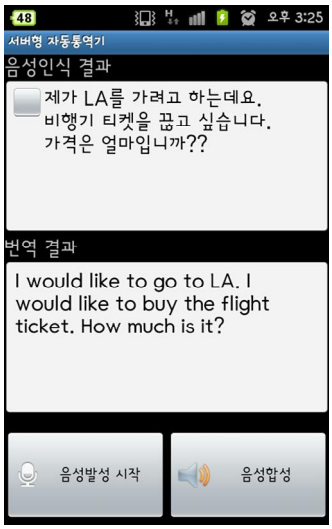
<자료>: Google 번역 화면 캡처

(그림 6) Google 번역

또한 Google 번역의 경우 서버 기반으로 동작하는데 이는 단말에 탑재되는 다국어 자동 통역기와 비교할 때 경량화에 대한 부담 없이 가능한 모든 지식을 활용할 수 있다는 데서 강점을 갖는다.

4. SayTran

한국어의 경우 앞서 언급한 다국어 자동 통역기들에서 통역 대상 언어로 지원하고 있기는 하나 아무래도 주요 언어가 아닌 탓에 그 성능은 영어, 일본어 등에 비해 떨어지는 편이다. 국내의 경우 한국어 기반의 일부 실험적인 자동통역 관련 제품이 출시되기는 하였으나 문장만을 인식하거나 Google의 API(Application Program Interface)를 이용하여 통역하는 등 본격적인 한국어 지원 다국어 자동 통역기라 할만한 제품은 아직 출시되지 않고 있다. 그래도 그 중 가장 관심을 가져볼 만한 것은 ETRI에서 개발 중인 SayTran(가칭, (그림 7) 참조)이다. SayTran의 경우 ETRI에서 한국어 및 영어를 대상으로 개발 중인 제품이며 현재 개발 완료 단계에 이르러 관련 업체와 함께 상용화를 준비하고 있다. SayTran의 가장 큰 특징은 한국어에 대해 강점을 보인다는 점이다. SayTran의 경우 주력하는 여행 및 일상 분야의 경우 한국어에 대해서 다른 자동 통역기와 비교했을 때 월등히 높은 음성인식 성능을 보이며 번역 성능 또한 SMT 기반의 다른 통역기들에 비해 현저히 높은 성능을 나타낸다. 다만 현재 지원하는 언어가 한국어와 영어뿐이라는 점에서 다국어 자동 통역기로서는 부족한 점이 있으나 현재



<자료>: SayTran 화면 캡처

(그림 7) SayTran

"WBS(World Best Software)" 사업에 참여하면서 일본어를 지원하는 동시에 관련 기술을 보강해 고성능 다국어 자동 통역기로의 한 단계 도약을 준비하고 있고, 또한 향후에는 중국어까지 확장할 계획을 갖고 있어 계획대로 진행이 된다면 한국어를 지원하는 다국어 자동 통역기를 만나게 되는 것도 머지 않은 일일 것으로 보인다.

Ⅳ. 결론

본 고에서는 다국어 확장에 중점을 두어 자동통역의 각 요소기술 및 응용 사례에 대해 살펴보았다. 자동통역 기술은 세계화의 시대에서 언어 장벽 문제를 해결할 수 있는 기술로서, 선진 각국에서는 일찍부터 연구 개발에 지원을 아끼지 않았다. 그 결과, 1990년대 초 C-STAR I 프로젝트의 일환으로 이루어진 시연에서 자동통역의 가능성을 볼 수 있었으며, 7~8년 전부터는 제한된 범위이긴 하나 실제 사용되는 상용 제품들이 출시되었다. 최근에는 방대한 양의 사용자 로그 데이터와 컴퓨팅 자원을 기반으로 예전보다 훨씬 뛰어난 통역 성능을 가진 제품들이 출시되고 있으며, 지원되는 언어도 확장되고 있다. 또한, 스마트폰의 대중화를 포함한 환경 변화로 인해 일반인들이 자동통역 제품을 쉽게 접할 수 있게 되었다. 지금은 자동통역 분야의 시장 선점을 위한 치열한 경쟁이 시작되는 중요한 시기로서, 국내 산/학/연 관련 기관들의 적극적인 대처가 필요하다.

● 용 어 해 설 ●

이중언어말뭉치: 같은 뜻을 가진 용례가 두 언어로 되어 있는 일정한 규모 이상의 크기를 갖추고 내용적으로 다양성과 균형성이 확보된 자료의 집합체

의미표상: 표상은 실세계와 대응되거나 그 대응을 가능하게 만들어주는 표현으로써 의미표상은 의미의 언어학적 표현을 말함.

약어 정리

API	Application Program Interface
CICC	Center of the International Cooperation for Computerization
C-STAR	Consortium for Speech Translation Advanced Research
DARPA	Defense Advanced Research Agency
DL	delta-likelihood
EPD	End Point Detection
FP7	Seventh(7th) Framework Programme
HMM	Hidden Markov Model
MDL	minimum description length
NICT	the National Institute of Information and Communications Technology
NSF	National Science Foundation
SMT	Statistical Machine Translation
VM-GMM	variable mixture Gaussian mixture model
WBS	World Best Software
가중K-L 거리	weighted Kullback-Leibler divergence

참 고 문 헌

[1] 김승희 외, 훤히 보이는 음성언어기술, 전자신문사, 2009.

[2] A. Waibel and C. Fugen, "Spoken Language Translation," *IEEE Signal Proc. Mag.*, May, 2008, pp. 70-79.
[3] F. Beaufays, V. Vanhoucke, and B. Strope, "Unsupervised Discovery and Training of Maximally Dissimilar Cluster Models," *Proc. INTERSPEECH*, 2010, pp. 66-69.
[4] K. Shinoda and K. Iso, "Efficient Reduction of Gaussian Components Using MDL Criterion for HMM-Based Speech Recognition," *Proc. ICASSP*, 2002, pp. 869-872.
[5] H.Y. Cho and S.H. Kim, "A New Distance Measure for a Variable-Sized Acoustic Model Based on MDL Technique," *ETRI J.*, vol. 32, no. 5, 2010, pp. 795-800.
[6] C. Chelba et al., "Query Language Modeling for Voice Search," *Proc. IEEE Workshop Spoken Language Technol.*, 2010, pp. 115-120.
[7] C. Chelba et al., "Study on Interaction between Entropy Pruning and Kneser-Ney Smoothing," *Proc. INTERSPEECH*, 2010, pp. 2422-2425.
[8] P. Fung and T. Schultz, "Multilingual Spoken Language Processing," *IEEE Signal Proc. Mag.*, May, 2008, pp. 89-97.
[9] T. Hughes et al., "Building Transcribed Speech Corpora Quickly and Cheaply for Many Languages," *Proc. INTERSPEECH*, 2010, pp. 1914-1917.
[10] Ilbin Lee et al., "An Overview of Korean-English Speech-to-Speech Translation System," *Proc. TCAST Workshop*, 2009, pp. 6-9.
[11] Hans Uszkoreit, "Hybrid Machine Translation," *Translingual Eur.*, 2009. http://ufal.mff.cuni.cz/tle2009/program.html.
[12] 최승권, 홍문표, 박상규, "다국어 자동번역 기술," 전자통신동향분석, 제20권 제5호, 2005, pp. 16-27.
[13] W.J. Hutchins, Machine Translation-past, present, future, Chichester: Ellis Horwood, New York: Hatsted Press, 1986.
[14] Hong, M.P. and O. Streiter, "Overcoming the Language Barriers in the Web: The UNL-Approach," *Tagungsband der 11 Jahrestagung der Gesellschaft fuer Linguistische Datenverarbeitung(GLDV)*, Frankfurt am Main, Germany, 1999.
[15] B.J. Dorr, A View from the Lexicon, Mass.: MIT Press, London: Cambridge, 1993.
[16] Sung-Kwon Choi, Tae-Wan Kim, and Dong-In Park. "Common and Constraint Grammar in Transfer-based Multilingual Machine Translation," *Nat. Language Proc. Pacific Rim Symp.*, 1997, pp. 517-520.
[17] Sung-Kwon Choi, Unification-based machine translation with Korean as source language, Dissertation, 1995.
[18] 김운 외, "자동번역 기술 동향 및 응용 사례," 전자통신동향분석, 제23권 제1호, 2008, pp. 89-98.
[19] W. Weaver. "Translation(1949)," In: Machine Translation of Languages, MIT Press, Cambridge, MA, 1995.
[20] P. Brown et al., "The Mathematics of Statistical Machine Translation: Parameter Estimation," *Comput. Linguistics*, vol. 19, no. 2, 1991, pp. 263-311.
[21] P. Koehn, F.J. Och, and D. Marcu, "Statistical Phrase Based Translation," *Proc. HLT/NAACL*, 2003.
[22] Y.S. Hwang, A. Finch, and Y. Sasaki, "Improving Statistical Machine Translation Using Shallow Linguistic Knowledge," *Comput. Speech Language*, vol. 21, no. 2, 2007.
[23] D. Chiang, "A Hierarchical Phrase-Based Model for Statistical Machine Translation," *Proc. ACL*, 2005.
[24] C. Bannard and C.B. Callison, "Paraphrasing withBilingual Parallel Corpora," *Proc. ACL*, 2005.
[25] Y.S. Hwang, Y.K. Kim, and S.K. Park, "Paraphrasing Depending on Bilingual Context Toward Generalization of Translation Knowledge," *Proc. 3rd Int. Joint Conf. Nat. Language Proc.*, 2008.
[26] http://www.nist.gov/speech/tests/mt/doc/mt06-eval_ official_results.html
[27] http://www.osti.gov/fedrnd/
[28] http://www.nsf.gov/awardsearch/
[29] http://www.cicc.or.jp/english/
[30] http://www.euromatrixplus.net/
[31] http://ec.europa.eu/information_society/apps/

projects/factsheet/index.cfm?project_ref=250456
[32] B. Pfister and H. Romsdorfer, "Mixed-lingual text Analysis for Polyglot TTS Synthesis," *Proc. Eurospeech*, Geneva, Switzerland, 2003.
[33] Alan W Black and Kevin A. Lenzo, "Multilingual Text-to-Speech Synthesis," *Proc. ICASSP*, 2004.
[34] 김종진 외, "내장형 음성합성 기술 동향 및 사례", 전자통신동향분석, 제23권 제 1호, 2008, pp. 77-88.