특 집

대화체 음성 자동통역 기술

박 준

한국전자통신연구원 컴퓨터소프트웨어연구소

I. 서 론

자동통역 기술이란 언어장벽을 허물어 서로 다른 언어를 사용하는 사람 간에 대화가 가능하게 하는 기술이다. 자동통역 기술이 실현되면 현재 진행되고 있는 세계화가 가속되어 개인 생활이나 기업활동, 사회 전반에 지대한 영향을 미칠 것이다. 이러한 자동통역을 실현하기 위해서는 먼저 사람이 발성한 소리를 문자로 나타내는 음성인식 기술과, 이를 같은 의미를 갖도록 상대 언어의 문장으로 변환하는 언어번역기술, 그리고 문자로 표기된 문장을 음성으로 읽어 주는 음성합성기술 등의 요소기술을 확보하여야 한다. 또한, 적어도 두개 이상의 언어를 처리하여야 하며 각각의 언어에 대하여 요소기술들을 개발하여야 한다. 지난 수십년에 걸친 기술 축적과 최근 컴퓨터 및 반도체 기술의 급속한 발전을 바탕으로, 그간 상상 속에서만 존재하던 자동통역 기술을 실현하기 위하여 많은 노력이 경주되고 있다.

다국어간 자동통역에 필요한 기술을 효율적으로 개발하기 위하여 국제 자동통역 공동연구 컨소시엄(Consortium for Speech Translation Advanced Research, 이하 C-STAR라 칭함)이 결성되어 관련연구를 진행하고 있다[1]. 이 C-STAR에는 일본의 ATR 연구소, 미국의 카네기멜런대학교, 이탈리아의 ITC-IRST 연구소, 독일의 Karlsruhe 대학교, 프랑스의 CLIPS 그룹, 중국의 NLPR 등 자동통역 분야의 첨단 연구기관이 대거 참여하고 있으며, 한국전자통신연구원(이하 ETRI라 칭함)도 핵심그룹의 일원으로 참여하고 있다. 1995년부터 시작한 C-STAR 2단계 공동연구에서는 해외 여행을 가고자 하는 고객이 외국의 여행사 직원과 여행계획을 수립하는 상황을 설정하여, 대화체 음성을 대상으로 다국어간 공통 번역 기술을 개발하였으며, 1999년 7월 22일에 국제간 자동통역 실시간 시연을 통하여 그 연구결과를 공개하였다. 이어서 2000년부터 시작한 3단계 연구에서는 휴대폰을 이용하여 여행자를 위한 자동통역 서비스의 상용화를 위한 소요 기술들을 개발하고 있다.

국내에서도 삼성전자가 한국어를 일본어로 통역하는 기술을 개발하여 2001년 시범 서비스를 선보였다[2]. 이 서비스는 약 1,500개의 정해진 문장에 대하여 전화로 전달된 한국어 문장 음성을 인식하고, 이에 해당하는 일본어 문장을 합성하여 들려주는 방식으로 동작한다. 앞으로 예문의 수를 늘려가고, 일본어에서 한국어로 통역하는 서비스를 추가하여 양방향 통역기능을 제공함으로써 상용 서비스로서 자리잡을 것으로 기대된다.

한편, (주)언어과학이나[3] 창신소프트(주)[4]에서는 PDA를 기반으로 주어진 문장을 선택하거나 필기 문자인식을 통하여 문장을 입력으로 받아, 이에 대응하는 상대 언어의 문장을 출력하며, 선택적으로 번역결과를 음성으로 출력할 수 있는 제품을 출시하고 있다. 이러한 제품들에 음성인식 기술을 접목시킨다면 제한적 기능에 대하여 보다 편리한 방식으로 자동통역의 혜택을 제공하게 될 것이다.

본 고에서는 C-STAR 회원기관으로서 ETRI에서 수행하고 있는 자동통역 기술 개발 내용을

중심으로 대화체 음성 자동통역에서의 고려사항과 함께 기술적인 요소를 소개한다. 먼저, 대화체 발화의 특징을 살펴본 후, 음성인식, 언어번역, 음성합성의 각 요소기술의 개발에서 고려사항 및 접근방법을 기술한다. 이어서 자동통역 기술의 상용화와 관련된 고려사항을 고찰하고, 끝으로 결언으로서 마무리한다.

II. 대화체 발화의 특징

대화체 음성은 매우 다양한 발화 자유도를 갖는다. 일반 신문 기사나 소설 등에 나타나는 문장들은 의도하는 목적에 맞도록 시간을 들여 의미나 문법적 검토를 거쳐 작성된다. 이와 달리 대화체 발화의 경우, 대화 흐름에 따라서 발화자의 머리 속에서 실시간으로 생성되어 발성을 하게 되므로 문법뿐 아니라 의미 자체도 불분명한 경우가 많다. 이러한 이유로 대화체 발화에서는 반복, 도치, 수정, 생략 현상이 자주 발생하며, 간투사도 수시로 삽입된다. 또한, '하실래요', '할께요', '한가요' 등과 같이 대화체에 고유한 어미가 다양하게 나타난다. 특히, 실제 음성으로 발화할 때 일관성있게 맞춤법 표기와 다르게 발성되는 경우가 많이 나타난다. 예를 들어, 발화자에 관계없이 어절의 끝부분에서 '로'가 '루'로, '고'가 '구'로, '요'가 '여'로, '도'가 '두'로, '죠'가 '져'로 대부분 바뀌어서 발성되며, 이에 따라 '하고요'는 '하구여'로, '그리고'는 '그리구'로, '호텔도'는 '호텔두' 등으로 발성된다. 그리고, 상대방에게 새로운 정보를 말하는 경우에는 상대적으로 정확하게, 천천히 발음을 하며, 정보가 적은 부분은 빠르고 약하게 발성하여 애매한 발음들이 많이 나타난다.

대화체 발화에 대한 음성신호와 문장을 수집하기 위하여 ETRI에서는 여러가지 상황을 설정하여 모의 대화를 수집하였는데, 위와 같은 발화의 자유도는 대화 상황에 따라 큰 영향을 받는 것으로 관찰되었다. 먼저 맞대면 상황에서 서로 우리말로 대화하는 경우, 발화 패턴은 매우 자유롭게 나타난다. 발화 길이도 대부분 매우 짧고, 문형도 비정형문이 많이 나타난다. 상대방이 한 말을 5초 후에 듣는 방식으로[1]전화를 통하여 서로 우리말로 대화하는 경우에는 발화의 길이가 길어졌으며, 동일한 상황에서 한 사람이 ARS 상황의 응답을 흉내내는 경우 상대방의 발화도 보다 간단명료해졌다. 두 대화자가 서로 다른 방에서 영상회의 시스템을 통하여 상대방의 얼굴을 보면서 다른 방에 있는 통역사가 통역을 하는 상황에서 서로 다른 언어를 사용하여[2] 대화를 나누는 경우에는 발화의 길이는 길었으나, 명료한 발음으로 발성하며, 발화 자유도가 상당히 줄어드는 현상을 보였다.

위의 내용을 정리하면, 얼굴을 맞대고 대화하는 경우 손짓, 몸짓, 표정 등 다른 채널을 함께 사용할 수 있어 발화에서 문형에 대한 제약이 상대적으로 완화되며, 따라서 발화의 자유도 자체는 높아지게 된다. 또한, 발화 사이에 지연시간이 삽입되는 경우에는 기다림 끝에 말하는 기회를 갖게 되므로, 한번에 보다 많은 정보를 전달하고자 함으로써 발화의 길이가 길어지게 된다. 그리고, 기계에게 말을 한다거나 통역사가 중간에서 통역하는 경우에는 자신의 의도를 제대로 전달하고자 발음이나 문형을 보다 분명하게 발성하는 것으로 풀이된다.

이러한 관찰 결과를 바탕으로 판단할 때, 실제 자동통역 서비스를 이용하는 상황에서, 특히 대화가 절실한 상황에서는 사용자가 자유롭게 발화하기 보다는 본인이 발성한 내용이 제대로 통역될 수 있도록 발음을 보다 정확하게 할 것이며, 문장 유형도 보다 정형적인 문장을 사용하는 등 협조적인 자세로 임할 것으로 예상된다.

1) 실제 자동통역 상황에서는 발성한 내용을 통역하는데 시간이 걸리므로, 말이 끝난 후 수 초의 지연 후에 상대방에게 전달된다. 이러한 상황을 모사하기 위하여 5초의 지연시간을 삽입하였다.

2) 한국어와 영어, 한국어와 일본어

Ⅲ. 대화체 음성인식

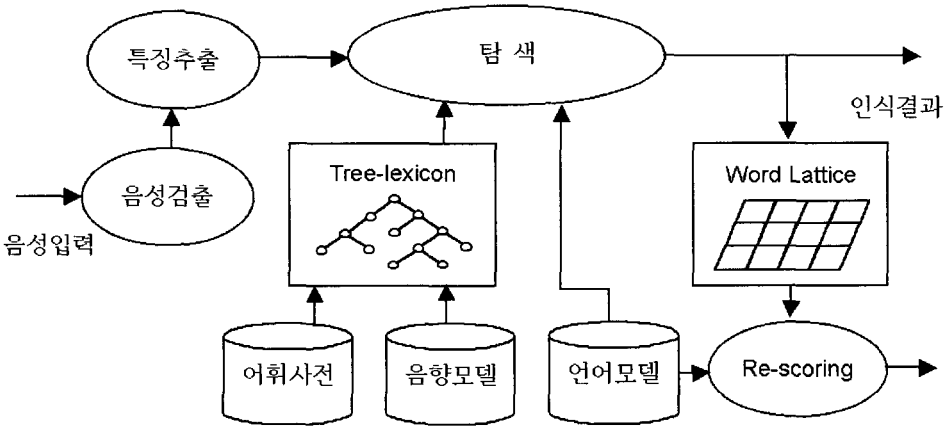
현재 ETRI에서 개발하고 있는 대화체 음성에 대한 음성인식 시스템의 기본적인 구조는 낭독체 음성에 대한 연속음성 인식 방법의 구조와 같다[5]. 〈그림 1〉에서 보는 바와 같이 대량의 음성 데이터를 이용하여 HMM 기반의 음향모델을 훈련하고[6], 가용한 대화체 문장을 수집하여 이를 기반으로 Katz의 백오프 방식[7]을 적용한 트라이그램 언어모델을 구축한다. 그리고, 어휘사전에 등록되어 있는 어휘에 대하여 음향모델을 참조하여 tree-lexicon을 구성한다. 음성이 입력되면 먼저 음성 구간을 검출한 후, 10 msec 주기로 16 msec 길이의 프레임 데이터에 대하여 특징을 추출하고 이를 기반으로 tree-lexicon 상에서 탐색을 수행한다. 이때, 각 어휘의 끝부분을 나타내는 노드에서 언어모델의 확률을 적용한다. 이러한 기본 구조 위에 선택적으로, 탐색 과정에서 최적의 결과와 아울러 확률이 높은 후보 문장들도 함께 취합하여 어휘 격자(word lattice)를 구성하고, 이를 대상으로 4-gram 등 추가 언어모델 정보를 활용하여 재검색하는 모듈로 구성된다. 본 장에서는 대화체 음성 인식에서의 고려사항과 그에 대한 접근방법을 간략히 기술한다.

1. 잡음 처리

마이크 음성과 달리, 전화 음성 특히, 휴대폰 음성을 인식하려면 잡음에 대비한 처리가 필요하다. 잡음은 크게 부가 잡음과 채널 잡음으로 구분할 수 있다. 채널잡음은 음성에 비하여 그 특성이 시간적으로 천천히 변하는 성질을 활용하여 기본적인 특징추출 작업을 한 후 특징벡터의 각 요소별로 낮은 주파수 성분을 제거하는 방식의 RASTA 필터링[8]을 적용한다. 부가잡음에 대해서는 대역별로 신호대비 잡음의 크기에 따라 신호에 대한 가중치를 부여하는 방식의 Wiener 필터를 적용하여 잡음을 제거한다[9]. 즉, 신호에 비하여 잡음이 큰 경우 해당 대역의 신호를 감소시키고, 신호에 비하여 잡음이 미미한 경우 원래 신호를 그대로 유지하는 방법으로 잡음을 제거한다. 여기에 부가하여 인간의 청각특성을 반영하는 방식도 함께 고려하고 있다.

2. 의사형태소

연속음성 인식의 경우 어휘사전에 등록될 인식 단위를 결정하여야 한다. 띄어쓰기 단위인 어절을 사용할 수 있으나, 하나의 명사에 많은 종류의 조사가 따라올 수 있고, 같은 용언에 대해서도 다양한 형태의 어미가 존재하여 어절의 수는 텍스트 양이 늘어감에 따라 급격하게 증가하게 된다. 다시 말하여, 일정 규모의 어휘사전을 사용할 때



〈그림 2〉 대화체 연속음성 인식시스템의 구조도

미등록 어휘 문제가 심각하게 발생하게 된다. 또 다른 선택은 어절을 구성하는 형태소로 어휘 사전을 구성하는 것이다. 그러나, 불규칙 용언의 경우 형태소로 분리하였을 때 원형이 복원되어 그 소리값이 유지되지 않는 경우가 발생한다. 예를 들어 '써서'가 '쓰+어서'로 분리되면 그 소리값이 달라져서 인식단위로 사용할 수 없다. 또한, 관형사형 전성어미인 'ㄴ'이나 'ㄹ'과 같이 한 두 음소로 이루어진 형태소의 경우 타 형태소와의 음향적 변별력이 떨어져 인식률이 떨어지게 된다. 이러한 문제를 해결하기 위하여 ETRI에서는 의사형태소를 정의하여 사용하고 있다[10]. 의사형태소는 형태소와 유사하나, 불규칙용언에서 그 소리값을 유지하는 형태소를 새로 도입하여 분리하고, 짧은 길이의 어미는 서로 결합하여 하나의 어절이 두 개의 의사형태소로 분리되도록 수정한 것이다. 예를 들어, '써서'의 경우는 '써+서'로 분리하여 이를 사용한다.

3. 다중발음 처리

앞서 살펴본 바와 같이 대화체에서는 맞춤법 표기와 달리 발음 변이 현상이 자주 발생한다. 예를 들어, '어떻게'가 '어트케'로 '그리고'가 '그리구'로, '하고요'가 '하구여'로 발성된 경우가 많이 나타난다. 이러한 발음변이된 어휘를 그대로 어휘사전에 등록할 경우, 어휘 수가 불필요하게 많아져서 인식 시간이 늘어나며 성능도 떨어지게 된다. 또한, 언어모델에서도 같은 어휘에 대하여 복수의 표기가 존재하여 모델의 성능을 떨어뜨리게 된다. 이러한 문제를 해결하기 위하여 다중발음 처리를 도입한다. 즉, 하나의 어휘에 대하여 표준 표기에 따라 대표 어휘를 정하고, 이에 대하여 나타날 수 있는 모든 발음에 대하여 대표 어휘를 가리키도록 한다. 또한 대표 어휘를 사용하여 언어모델을 구축함으로써 주어진 양의 코퍼스로부터 보다 정제된 언어모델을 얻을 수 있다.

4. 광역음운환경 기반 음향 모델링

연속음성 인식에서는 고립단어 인식에 비하여 음소간에 음운변이 현상이 많이 나타나며, 이러한 현상에 대처하기 위하여 일반적으로 음소 전후의 음운환경을 고려한 트라이폰을 사용한다. 그런데, 대화체 음성에서는 인접한 음소간의 영향을 미치는 범위가 낭독체 음성보다 더 넓다고 가정할 수 있다. 따라서, 음운환경을 최대 전후 3개 음소까지 고려하여 음향모델 단위를 설정한다[11]. 그런데, 고려 대상 음운환경의 범위를 넓히는 경우 독립적인 음소의 개수가 크기 늘어, 주어진 양의 훈련용 음성 데이터 내에 해당 음소에 할당되는 음성 데이터의 양이 적어지므로 음향모델의 성능이 저하된다. 따라서, 유사한 음운환경을 갖는 음소들은 다시 병합하여 최종 음향모델 단위를 결정한다.

5. 언어모델링

일반적으로 언어모델을 구축함에 있어 방대한 양의 텍스트 데이터를 필요로 하는데 반하여 대화체 문장의 경우 가용한 텍스트 데이터의 양이 제한되어 있어, 고품질의 언어모델을 구축하기가 어렵다. 그리고, 아래에서 기술하는 바와 같이 작업상황을 세분하여 접근하는 경우 각각의 상황에 대한 텍스트 데이터의 양도 따라서 크게 줄어들게 되어 어려움이 가중된다. 이러한 문제에 대하여 먼저, 모든 가용한 대화 문장을 이용하여 언어모델을 구축한 다음 세부 상황별로 구축한 언어모델과 융합하여 사용하는 방법에 대하여 연구를 진행하고 있다. 보다 근본적으로는 대화 흐름에 대한 모델을 정립하여 문맥에 적절한 세부 언어모델을 적용하는 방식이 바람직할 것이며, 앞으로 해결해야 할 과제로 남아 있다.

Ⅳ. 대화체 언어번역

대화체 발화에서는 생략, 반복, 오발성, 간투사 등이 자주 발생하므로, 언어번역에 있어서 이러한 특성에 대처하는 접근 방식이 필요하다. 또한, 여러 언어간 자동통역을 효율적으로 실현할 수 있는 방식이 요구된다, 이러한 점을 고려하여,

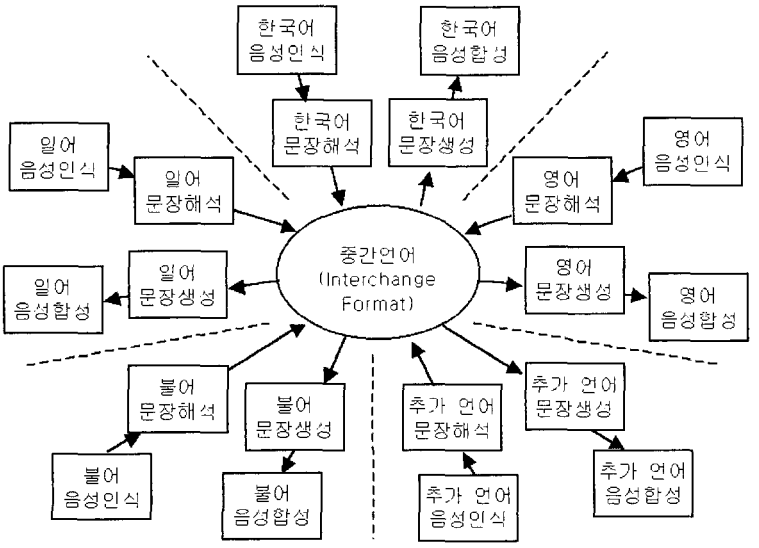
C-STAR에서는 여러 언어를 효율적으로 수용할 수 있는 방법으로서 개념 기반의 중간언어를 통한 번역 방식에 대한 연구를 2단계 공동연구에서 수행하였으며[12,13], 2000년부터 시작한 3단계에서는 대규모 대역코퍼스를 공동으로 구축하고 이를 기반으로 통계적 번역방식을 연구하고 있다. 본 장에서는 이와 같은 개념 기반 중간언어를 통한 언어번역과 대역코퍼스 기반 통계적 접근방식을 소개한다.

1. 개념 기반의 중간언어 (Interchange Format)를 통한 언어번역

두 언어간 통역을 하려면, 양 방향으로 통역이 이루어져야 한다. 예를 들어 한국어와 중국어 간 통역을 하고자 한다면, 한국어를 중국어로 통역하는 작업과 중국어를 한국어로 통역하는 작업도 함께 필요하다. 만일 n개의 언어간 통역하고자 한다면 각 언어 쌍마다 양방향 통역이 필요하며, 따라서 모두 n*(n-1) 만큼의 통역기술이 필요하게 된다. 그리고, 자동통역을 구성하는 요소 기술 중 음성인식기술과 음성합성기술은 한 언어에 대하여 개발하면 상대 언어에 관계없이 사용할 수 있는 반면, 언어번역 기술은 번역하고자 하는 모든 상대 언어에 대하여 기술이 개발되어야 한다.

C-STAR에서는 다국어간 언어번역에 효율적으로 방식으로서 중간언어 기반 번역방식을 연구하여 왔다. 중간언어 기반의 번역이란 각 언어 사이를 매개하는 중간언어를 정의하고, 이 중간언어를 통하여 언어번역을 수행하는 것이다. 즉, 발화자의 음성을 인식하고, 인식된 결과를 중간언어로 변환한 후, 이에 해당하는 상대 언어의 문장을 생성하고 이를 음성합성기를 통하여 음성으로 출력하는 것이다. 이러한 중간언어 방식은 〈그림 2〉에서 보는 바와 같이 각 언어에 대하여 음성인식 기술, 음성합성 기술, 인식결과를 해석하여 중간언어를 생성하는 문장해석 기술, 마지막으로 중간언어로부터 문장을 생성하는 문장생성 기술을 개발하면, 중간언어를 지원하는 다른 어느 언어와도 번역이 가능하다는 장점을 갖는다. 실제 시스템을 구축하는 관점에서도 인터넷을 통하여 중간언어로 표기된 결과를 용이하게 전송할 수 있으므로 각 언어를 처리하는 시스템을 분산하여 설치할 수 있는 장점을 제공한다.

이러한 중간언어를 모든 작업 영역에 대하여 정의하는 것은 실로 방대한 작업이 될 것이다. 그



〈그림 2〉 중간언어 기반의 다국어간 번역 방식

러나, 실제 자동통역기술을 적용함에 있어서 초기단계에서는 특정 상황을 설정하여 통역하는 방식이 될 것으로 예상된다. 예를 들어, 호텔 예약의 경우 나타날 수 있는 대화가 한정되어 있으며, 이러한 작업에 대하여 중간언어를 정의하고 사용하는 것이 가능하다. 중간언어를 설계함에 있어 특기할 점은 발화자의 의도가 잘 표현되도록 개념을 기반으로 그 구조를 정의한 것이다. 이는 대화체 스타일의 문장에서 나타나는 비 정형성뿐만 아니라, 음성인식 단계에서 의도 전달에 문제가 되지 않는 오류도 발생할 수 있으므로 정형 문법에 의한 구문분석 방식은 적용하기 어렵기 때문이다. 또한, 여러 언어 간에 공통으로 사용하므로 각 언어에 고유한 구문 특성을 반영하기 보다는 언어 공통적인 개념의 표현에 주력하는 것이 적절한 접근 방법이다.

중간언어의 구조는 화행(speech act)과 개념(concept), 그리고 변수(argument)와 그 변수에 해당하는 실제 값(value)로 구성된다[14]. 화행이란 정보요청이나, 정보제공, 또는 인사 등 전체 문장의 종류를 나타내며, 개념은 보다 구체적인 발화의 내용을 나타내는 것으로 예를 들면, 객실 이용의 가능성 여부나, 여행상품, 비행기 등에 대해 정보 제공 등의 보다 명확한 의미를 나타내는 것들이다. 변수는 선택사항으로서 개념의 종류에 따라 구체적인 내용이 필요할 경우에만 존재한다. 〈표 1〉에 중간언어를 통한 한국어와 영어간 번역의 일례가 예시되어 있다. 여기서 화행은 request-information, 개념은 availability-flight, 그리고, 개념을 보완하는 변수 destination과 via 및 해당 값으로 중간언어가 구성되어 있다.

〈표 1〉 중간언어 예문

구 분	예 문
한국어	혹시 바로 런던으로 가는 비행기는 없습니까?
중간언어	c : request-information +availability-flight(destination =london, via=nonstop)
영 어	Is there any flight going to London directly?

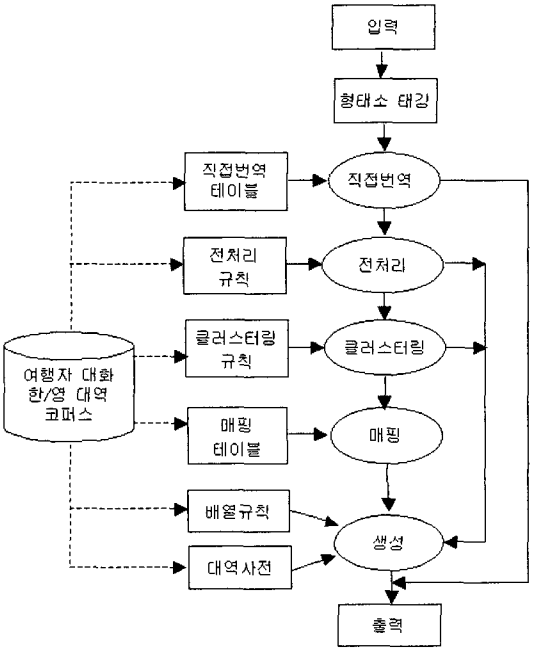
2. PDMT(Phrase-based Direct Machine Translation)

위에서 언급한 개념 기반 중간언어 방식이 다국어간 자동통역 기술 개발에 효율적인 인터페이스를 제공하는 장점에도 불구하고, 1) 문장 해석 및 생성 작업에 필요한 문법의 작성을 수작업으로 수행하여 개발에 시간이 많이 걸리고, 2) 작업영역을 확대하면서 문법을 증강할 경우 기존 문법과 새로 작성한 문법 간에 일관성 유지가 어렵게 되며, 또한, 3) 작업 영역이 바뀌는 경우, 문법을 대부분 재작성하여야 하기 때문에 작업 영역에 대한 이식성이 떨어지는 문제점을 갖고 있다.

이러한 문제를 해소하기 위하여 ETRI에서는 통계적인 접근방법에 기반한 구단위 직접번역(Phrase-based Direct Machine Translation : 이하 PDMT) 방식에 대한 연구를 진행하고 있다. PDMT의 기본 아이디어는 대역 코퍼스로부터 번역단위인 '구'를 통계적으로 자동 추출하고 입력 문장을 '구' 단위로 자동 분할하여 이를 번역한 후, 번역된 '구'를 적절히 재배치하여 상대언어에 해당하는 문장을 생성함으로써 번역을 수행하고자 하는 것이다. 이때, 번역을 하고자 하는 언어간 '구'의 대응관계나, '구'의 재배치 규칙 등을 모두 대역코퍼스로부터 통계적인 방법으로 추출하여 사용한다.

PDMT 방식의 전체 동작 흐름도는 〈그림 2〉와 같으며, 각 단계에서의 동작을 간략히 살펴보면 다음과 같다. 입력된 문장에 대하여 먼저 형태소 태깅을 한 후, 인접한 형태소들을 적당한 크기로 묶어 입력문장을 구단위로 구분하는 클러스터링을 수행한다. 클러스터링에서 사용하는 규칙은 대역코퍼스에 나타나는 형태소 열의 출현 빈도수를 고려하여 클러스터를 정하며, 클러스터의 길이와 발생 확률을 적용한다. 여기서, 제한된 양의 코퍼스로부터 신뢰도 있는 클러스터를 추출하기

위하여 클러스터링을 하기 전에 전처리 단계에서 다음과 같은 작업을 수행한다. 1) 프로텍터의 적용 : 형태소 정보를 기반으로 구분되지 말아야 하는 부분에서는 구분되지 않도록 한다. 예를 들어, '것', '개' 등 의존 명사 앞에서는 분할되지 않도록 한다. 2) 복합성 명사열 추출 : 명사를 수식하는 부분은 명사구로 한 단위로 처리한다. 즉, '재즈 클럽 공연'이나 '산타모니카 행 버스 번호'는 각각 '공연'과 '번호'로서 분할되지 않도록 한 단위로서 처리한다. 3) 변수 치환 : 대화 상에서 나타나는 어휘 중 명사의 종류가 가장 많으며 이를 그대로 처리하는 경우 클러스터도 많아지게 된다. 따라서, 분할된 클러스터의 품질이 떨어질 수 있다. 이러한 문제를 해결하기 위하여 명사를 변수로 치환하여 처리한다. 변수는 명사의 속성을 기반으로 170가지로 분류한 대표명사 분류체계를 이용하여 정한다. 예를 들면, '미숫가루'나 '스테이크'는 '식품'으로, '매표소'나 '놀이공원'은 '장소'로 대체한다[12]. 마지막으로, 4) 같은 의미로서 다양한 표기를 갖는 동사 어미는 이를 은닉함으로써 단순화한다.



〈그림 3〉 PDMT의 동작 흐름도

클러스터링을 수행하면 다음으로 매핑이 수행되는데, 이는 입력 클러스터 각각에 대응하는 상대 언어의 클러스터로 바꾸어 주는 단계이다. 이를 위하여 대역코퍼스로부터 각각 언어에 대한 클러스터링을 수행한 후 문장안에서 동시에 출현한 대응 클러스터의 빈도수를 이용하여 언어간 클러스터의 대응관계를 정하고, 이를 매핑테이블에 수록한다.

생성단계에서는 먼저, 매핑결과로 전달된 클러스터에 대하여 클러스터 bi-gram 정보를 이용하여 순서를 다시 정한다. 또한, 대역사전을 이용하여 복합성 명사열이나 변수 처리된 명사들을 목표 언어의 어휘로 바꾸어 주고, 목표 언어의 특징을 반영하여 보다 자연스러운 문장으로 후처리 작업을 수행한다. 즉, 영어의 경우 수와 시제의 일치, 의문문이나 평서문과 같은 문형처리 등을 수행한다.

순서상으로 입력문장에 대하여 첫번째로 수행하는 단계는 직접번역인데, 이는 자주 쓰이는 문장은 위에서 기술한 과정을 거치지 않고 그대로 목표언어의 문장으로 매핑하여 번역을 수행하는 것이다. 직접번역 작업은 입력문 그대로 사용하는 경우와 동사 어미를 대표 표현으로 바꾸고, 명사를 변수로 처리하여 번역을 수행하는 2가지 세부단계로 구성된다. 직접번역단계는 기술적으로는 간단하지만, 실제 응용에 있어서는 처리되는 문장의 많은 부분을 담당하여 전체 번역 성공율을 높일 것으로 기대되고 있다.

V. 대화체 음성합성

자동통역에서는 단순한 정보 전달뿐만 아니라 대화 상황에 따라 자연스러운 의사 소통이 가능하도록 대화체 스타일의 합성음이 요구된다. 대화체 합성기를 개발하기 위해서는 대화체 음성 DB 구축 및 자동화 기술, 코퍼스 방식에서 합성단위 선정기술, 대화모델링 기술 및 대화체 운율부가 기술, 대화체 언어처리 기술 등이 개발해야

할 주요 기술이 될 것으로 보인다. ETRI에서는 현재 대화체 음성 DB 구축은 보다 자연스러운 음성 DB를 수집하기 위해 성우 2명이 서로 대화(dialogue)하는 상황에서 음성을 수집하였으며, 합성단위로 사용되는 트라이폰(triphone) 커버리지를 높이기 위해 일부 모놀로그(monologue) 음성 DB와 낭독체 음성 DB도 함께 수집하였다. 또한 이러한 방대한 크기의 음성 DB로부터 합성 DB를 최대한 효율적으로 구축하고자 음소분할, 피치추출, 운율레이블링 자동화 기술도 아울러 개발하고 있다. 코퍼스 방식에서 합성단위 선정 기술은 합성단위가 최적으로 선정될 수 있도록 최적화 함수 (cost function)의 정의 및 새로운 운율/언어정보를 추출하여 특징으로 활용하고자 하고 있다. 대화모델링 기술 및 대화체 운율부가 기술은 어절 형태가 같으면서 대화 맥락에 따라 선택적으로 운율이 구현될 필요가 있는 문장 내의 언어적 요소에 대해 입력 텍스트 상에 나타난 화맥(speech context) 정보를 이용하여 대화 맥락에 맞는 운율을 표현할 수 있도록 하는 기술을 개발하고 있다. 또한 합성음의 자연성에 주요하게 영향을 미치는 대화체 음성의 끊어읽기 강도 모델링 및 대화체 텍스트에서의 끊어읽기 강도 예측모델도 함께 개발하고 있으며 대화체 언어처리는 대화체 텍스트 처리를 위한 형태소 태깅 성능개선에 주력하고 있다. 이와 같이 주요기술의 개발이 이루어지면 궁극적으로 대화체 음성합성기는 대화형 음성 인터페이스 상황에서 사용자 편의성을 한층 높일 수 있게 해줄 것이다.

VI. 자동통역기술의 실용화에 대한 고찰

C-STAR에서 1993년 학회등록에 대한 시연으로 자동통역 기술의 실현 가능성을 처음으로 열었을 때와 비교하면, 그동안 기술적으로 많은 발전을 하였으나, 일상의 모든 대화를 완벽하게 통역하기까지는 앞으로도 많은 노력과 시간이 필요할 것이다. 그러나, 여느 첨단기술의 응용과 마찬가지로 자동통역 기술도 처음에는 제한된 상황이나 성능에서 상용화가 시작되어 점차 그 적용범위와 수준을 확대하고 개선해 갈 것이다. 이러한 관점에서 ETRI에서는 자동통역 기술 개발의 목표를 해외 여행자를 위한 기본 대화의 통역으로 설정하고 있다. 해외 여행을 할 때 언어가 전혀 안통하는 경우 길 안내, 교통편 이용, 자동차 기름 충전 등 기본 활동에 상당한 불편을 겪게 된다. 이러한 경우 한마디만 통하여도 큰 도움을 얻을 수 있다. 또한, 여행에서 나타나는 여러가지 상황을 호텔, 기차역, 길안내, 공항, 식당, 상점 등 세부적으로 구분하고, 각각의 상황별로 통역을 수행한다. 이와 같이 상황 별로 나누어 접근함으로써 처리 대상 어휘나 문장의 수를 축소시킬 수 있으며, 따라서 자동통역의 성능을 보다 높일 수 있다.

음성번역의 성능을 나타내는 척도로서 문장번역 성공율과 아울러 작업성공율을 함께 고려할 수 있다. 문장 번역 성공율은 발성자가 말한 문장이 제대로 번역되어 상대방에게 전달되는 비율로서 정의할 수 있다. 작업 성공율은 대화의 목적을 달성하는 비율, 즉 호텔예약에 대한 대화인 경우 고객이 원하는 날짜에 원하는 조건의 방을 예약하는 작업을 성공하는 비율로 정의할 수 있다. 자동통역 시스템을 이용하여 대화를 나눌 때 문장번역 성공율이 높으면 당연히 작업 성공율도 높게 된다. 그런데, 특기할 사항은 문장 번역 성공율이 낮아지더라도 작업 성공율이 기계적으로 낮아지지는 않는다는 점이다. NESPOLE 프로젝트에서 실험한 바에 의하면, 문장번역 성공율이 40% 대에 머물더라도 작업성공율은 100%를 유지하는 것으로 나타났다[15]. 이는 오역이 일어나는 경우에도 듣는 사람이 문맥을 파악하고 있으므로, 번역 결과를 선택적으로 받아들여 필요한 경우 확인을 하며, 또한 문법적으로는 잘못 되었다고 하더라도 번역 결과에 나타난 핵심어를 기반으로 상대방의 의도를 파악하고 이에 적절하여 대응하기 때문이다. 즉, 양방향 대화 번역 상황에서는 작업의 성공을 향하여 대화자의 지능이 기여하게 된다. 이는 사람과 기계 사이의 인터페이

스의 경우와는 전혀 다른 상황으로서, 상용화의 관점에서 자동통역 성능은 문장 번역 성공율과 함께 작업성공율이 함께 고려되어야 할 것이다.

자동통역 기술의 상용화를 위하여 '실제 서비스나 제품을 어떠한 형상으로 제공할 수 있는가?'에 대해서도 고려가 필요하다. 쉽게 떠올릴 수 있는 구현 방법으로 서버를 설치하고 이를 인터넷이나 전화 등으로 접속하여 서비스를 받는 형상과 모든 작업을 휴대형 단말에서 처리하는 형상을 생각할 수 있다. 서비스 서버 형상에서는 고성능의 서버를 사용하여 고품질의 서비스 제공에 유리한 반면, 사용자가 서비스 요금을 매번 지불해야 하며, 특히 서비스가 제공되는 지역에서만 사용하여야 하는 제약이 따른다. 이와 반대로, 휴대형 단말 형상에서는 초기 단말기 구입 비용이 부담될 수 있으나 시간, 장소에 제약없이 사용할 수 있는 장점을 갖는다. 단기적으로는 리소스 문제로 자동통역 기능을 PDA 등과 같은 휴대형 단말기에 구현하기 어려우므로, 초기에는 자동통역 서버 형상이 먼저 나타날 것이다. 향후, 포스트PC 등 보다 강력한 단말기의 보급이 확산되면 휴대형 단말 형상의 자동통역 제품이 출현할 것이며, 서버 기반의 서비스와 함께 병존할 것으로 예상된다.

Ⅶ. 결 언

이상과 같이 대화체 음성 자동통역에 대하여 기술적인 요소를 살펴보고, 작업의 특성과 함께 ETRI에서 진행하고 있는 연구내용을 기술하였다. 자동통역이 이루어지기 위해서는 음성인식, 언어번역, 음성합성 분야의 기술이 함께 필요하며, 적어도 2개 이상의 언어에 대한 각각의 요소기술이 개발되어야 한다. 또한, 이러한 요소기술 자체도 완성도를 갖춘 성능을 달성하기 위하여 신호처리, 언어처리, 지식처리부문의 다양한 세부기술이 개발되어야 하므로, 자동통역은 실로 방대한 기술 복합체라 할 수 있다. 그러나, 수십년에 걸친 기술개발 노력과 아울러 최근 반도체, 컴퓨터 및 통신 분야의 비약적인 발전을 기반으로, 자동통역 기술이 실제 상용화 되어 일상생활에서 접할 수 있는 시대가 가까운 미래에 올 것으로 기대된다.

이러한 자동통역 기술의 상용화는 여행자용 기본대화에 대한 통역이나 국제 공항에서의 다국어 안내 등 비교적 단순한 응용 분야로부터 시작하여, 기술적 진화를 거듭하면서 장차는 국제회의 동시통역, 일반 대화에 대한 자동통역 전화 등과 같이 난이도가 매우 높은 응용 분야에까지 적용될 것이다.

자동통역 기술을 보다 다양한 분야에 응용하기 위해서는, 앞으로도 많은 연구개발 노력이 필요하다. 음성인식에서는 현재 초기상태에 있는 언어정보 활용 수준을 단순 문형정보나 n-gram을 활용하는 수준에서 의미분석을 통한 문맥정보를 활용할 수 있는 수준으로 끌어 올려야 하며, 응용분야에 대한 지식을 활용하는 기술도 접목되어야 할 것이다. 언어번역에서는 인식오류에 대하여 강인하게 대처, 수용하는 기술이나 응용 영역이 변경되어도 이를 유연하게 대처할 수 있도록 이식성이 보장되는 기술이 개발되어야 할 것이다. 이와 함께, 사용 리소스를 감축하여 소용량 플랫폼에서도 동작하게 하는 노력도 아울러 지속적으로 기울여야 할 것이다.

참 고 문 헌

〔1〕 C-STAR Home page, *www.c-star.org*

〔2〕 삼성종합기술원 HCI Lab 홈페이지,www.hci-lab.com

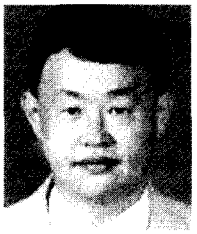
〔3〕 (주)언어과학 홈페이지, *www.eoneo.co.kr*

〔4〕 창신소프트(주) 홈페이지, *www.cssoft.co.kr/kr*

〔5〕 Oh-Wook Kwon and Jun Park, "Korean large vocabulary continuous speech recognition with morpheme-based recognition units," Speech Com-

munication, Vol. 39, Issues 3-4, pp. 287-300, 2003.
〔6〕 L. Rabiner and B.-H. Juang, Fundamentals of Speech Recognition, Prentice Hall, 1998.
〔7〕 S. M. Katz, "Estimation of Probabilisties from Sparse Data for the Language Model Component of a Speech Recognizer," IEEE Trans. Acoustics, Speech and Signal Processing, 3, pp. 400-401, 1987.
〔8〕 Hynek Hermansky and Nelson Morgan, "RASTA Processing of speech," IEEE Trans. on speech and audio processing, Vol. 2, No. 4, pp.587-589, 1994.
〔9〕 S. V. Vaseghi, Advanced Signal Processing and Digital Noise Reduction, Wiley & Teubner, 1996.
〔10〕 권오욱, 박준, 황규웅, "의사형태소 단위 대어휘 연속 음성 인식기 개발," 제15회 음성통신 및 신호처리 워크샵, 한국음향학회, pp.320-323, 1998.
〔11〕 J. Fritsch, M. Finke and A. Waibel, "Context-Dependent Hybrid HME/HMM Speech Recognition Using Polyphone Clustering Decision Trees," ICASSP, vol. 3, pp.1759-1762, 1997.
〔12〕 M. Mayfield, et al. "Concept Based Speech Translation," ICASSP, vol. 1, pp.97-100, 1995.
〔13〕 박준, 이영직, 양재우, "대화체 음성언어번역 시스템 개발," 제15회 음성통신 및 신호처리 워크샵, 한국음향학회, pp.281-286, 1998.
〔14〕 최운천, "다국어 대화체 음성언어번역 시스템을 위한 IF(Interchange Format)와 IF 태깅," 제15회 음성통신 및 신호처리 워크샵, 한국음향학회, pp.409-412, 1998.
〔15〕 C-STAR Partner Meeting, Trento, Italy, December, 2002.

저 자 소 개



박 준

1994년 8월 University of Southern California, Dept. of Electrical Engineering-Systems (Ph. D.), 1983년 2월 서울대학교 대학원 전자공학과 졸업 (석사), 1981년 2월 서울대학교 공과대학 전자공학과 졸업 (학사), 1983년 3월~현재 : 한국전자통신연구원 음성처리연구팀 책임연구원(팀장), 1989년 9월~1991년 6월 : Teaching Assistant, Dept. of EE-Systems, University of Southern California, 1994년 12월~1995년 11월 : Visiting Researcher, Center for Machine Translation, Carnegie-Mellon University, 〈주관심 분야 : 음성인식, 음성합성, 자동통역〉