《主 題》

자연언어 처리의 동향

김 영 길, 최 병 욱
(한양대학교 전자통신공학과)

□차 례□

Ⅰ. 서 론

자연언어 처리의 역사는 거의 반세기에 가까운 계산기의 역사에 필적할 정도로 오래되었으며 기계번역을 중심으로 발전을 거듭해온 자연언어 처리 기술은 1990년대에 들어서 여러 응용시스템의 형태로 다양하게 실용화의 단계에 접어들고 있다. 앞으로의 정보화 사회는 멀티미디어 분야가 주도하는 사회가 될 것으로 예상되며 여러 분야의 복합 기술인 멀티미디어 분야에서 자연언어 처리 기술은 가장 중요한 기술의 하나로 자리잡고 있다.

현재 기계번역에 관한 연구는 국내는 물론 국외에서도 음성인식 및 음성합성과의 인터페이스에 의한 음성언어 번역 시스템이라고 하는 통합 시스템을 목표로 대화체 번역에 관한 연구가 진행되고 있으며 그 방법론에 있어서도 종래의 규칙 기반의 언어처리에서 벗어나 대량의 말뭉치(Corpus)를 이용한 통계적 방식의 언어처리가 제시되고 있다. 아울러 자연언어 인터페이스를 이용한 데이타베이스 검색 시스템, 지능적 대화 시스템 등의 Man-Machine 인터페이스와 같은 자연언어 처리 응용 시스템에 관한 연구도 활발히 진행중이다.

따라서 본고에서는 음성언어 번역 시스템을 중심으로 국내외의 기계번역 연구 동향을 살펴보고 현재 관심의 대상이 되고 있는 통계론적 자연언어 처리의 개요와 동향을 기술한다. 그리고 응용 시스템으로 DB 인터페이스와 지능형 대화 시스템과 같은 자연언어 인터페이스 시스템을 소개한다.

Ⅱ. 기계번역 시스템의 동향

기계번역(Machine Translation)이란 컴퓨터를 사용하여 하나 또는 그 이상의 대상언어를 다른 대상언어로 번역하는 것이다. 언어는 한국어 또는 영어와 같은 자연언어를 말하는 것이며, 기계번역은 한 자연언어에서 다른 자연언어로의 변환을 말하는 것이다. 기계번역에 관한 연구는 현재 다수의 시스템들이 연구되고 있으며 해외의 경우 여러 시스템이 상용화되어 있는 실정이다. 그러나 자연언어를 완전히 형식화한다는 것이 불가능하고, 거의 무한에 가까운 방대한 언어 데이타를 취급해야 하므로, 기존의 어떤 시스템도 완전한 번역 시스템이라고 말할 수는 없다. 아래에서는 기계번역의 유형과 현재 활발한 연구 대상이 되고 있는 음성언어 번역 시스템에 있어서의 기계번역에 관하여 기술한다.

2.1 기계번역의 유형

기계번역의 유형은 현재 전통적인 규칙에 기반한 기계번역과 현재 새롭게 관심의 대상이 되고 있는 예제 기반의 번역 시스템 그리고 두 가지 방식을 통합

한 시스템으로 크게 나눌 수 있다. 전통적인 규칙에 기반한 번역 방식으로는 직접(Direct)번역 방식과, 변환(Transfer) 방식 그리고 중간언어(Interlingual) 방식의 세 가지로 구별하며 변환방식은 각 언어마다 독립적인 해석결과를 가지며 이들의 변환으로 번역이 이루어지므로 다언어간의 번역에 유리하며 현재 가장 활발히 연구되고 있는 부분이다.

변환방식은 원시언어와 목표언어에 대해 두 가지 형태의 중간언어를 설정하고 번역의 재과정을 진행하는 것을 일컫는다. 일반적으로 원시언어의 분석, 변환, 생성의 단계를 통해 번역이 진행된다. 분석과정은 목표언어와 무관하게 진행될 수 있으며 원시언어를 중간표현으로 재표현시킨다. 변환과정에서는 원시언어의 분석결과를 목표언어의 그것에 알맞게 구조적 차이를 조정하여 생성과정에서 이 결과를 이용하여 최종적으로 번역문을 생성한다.

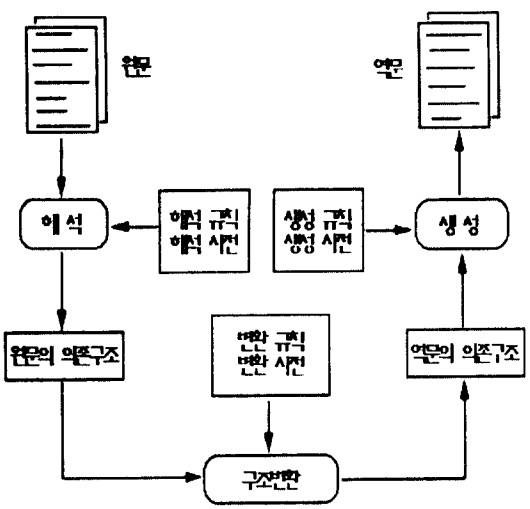


그림 2-1. 변환방식 기계번역 시스템의 구성도

그러나 전통적인 규칙 기반의 기계번역에 관한 연구는 대량의 규칙들에 대한 통제의 어려움과 적용분야의 협소성으로 인하여 실용적인 MT 시스템의 구성에 있어 한계에 도달하였으며 이에 대한 대안책으로 예제 기반의 기계번역이 새롭게 제시되고 있다. 예제 기반 기계번역(Example-based MT)은 비슷한 문장의 번역 예들을 모방하는 방법으로 번역을 수행하는 것이다. 이러한 방식의 한 예로 이단계 예문 기반 기계번역 방법론(two-phase example based machine translation methodology)이 있는데 이것은 예문에서 번역 템플릿(templet)을 개발하고 템플릿 일치(template matching)를 사용하여 번역을 하게 된다.

EBMT는 실제 사용자가 사용하는 구어체 등의 실용적 정보들을 지식으로 구성할 수 있으며 번역 지식의 획득이 인위적인 수단이 아니라 계산기를 이용한 자동적 수법으로 수집이 가능하다는 점에서 장점이 있다. 그러나 입력과 유사한 예문을 검색하기 위하여 대량의 예제들을 검사하고 그 유사도를 계산하는 등의 계산량이 방대하며 한쪽으로 편중되어 수집된 예제들을 이용함으로서 발생되는 해석 오류 등의 문제점들은 지속적으로 보완되어야 할 부분이다.

따라서 현재 전통적인 규칙기반의 방식과 예제 기반의 방식을 하나의 통합된 방식으로 재구성하려는 연구가 서서히 시도되고 있다. 통합 시스템은 2가지 해석 방식을 각각 하나의 하부 시스템으로 구성하고 중앙 제어기(central controller)를 두어 입력되는 문장의 종류에 따라 그 입력문에 장점이 있는 해석방식을 구동시키는 혼합형 방식과 전통적 규칙 기반의 해석기를 중심으로 예제 기반 방식을 특정한 문제 해결을 위해서 부분적으로 사용하는 방식이 제시되고 있다.

최근의 기계번역은 이상과 같은 여러가지 해석 방식의 시도와 아울러 대화체 언어(Spoken Language)를 대상으로 하고 음성인식의 결과를 번역하려는 시도가 진행되고 있다. 또한 대용량의 번역과 자동화를 위해 말뭉치에 기반하거나(Corpus-based Processing), 지식에 기반한(Knowledge-based Processing) 번역 기술을 연구하고 있으며, 문장간의 믿음이나 추론을 통해 대화를 이해하려는 연구도 있다.

2.2 음성언어 번역 시스템

국외의 경우 대화체 음성언어 번역에 관한 연구가 최근 몇 년간에 걸쳐 활발히 진행중이며 제한된 범위내에서 일부 시스템이 상용화를 목표로 하고 있어 그 장래가 주목받고 있다. 예를 들어 미국 Carnegie Mellon 대학을 중심으로 한 JANUS 시스템은 제한적이나마 영/일, 영/독 음성언어 번역이 가능하며 일본의 ATR에서 개발된 ASURA 시스템은 일/영 번역이 가능하고, 유럽의 SRI Cambridge에서는 영/스웨덴 음성언어 번역 시스템을 개발한 바 있다. 그러나 이러한 시스템들은 음성인식 분야의 성능 향상에 관한 연구에 집중하였지만 음성인식의 기술적 한계로 인하여 실용적인 음성언어 번역 시스템의 구축에 있어 어려움을 겪고 있는 실정이다.

그림 2-2는 음성언어 번역 시스템의 일례로 JANUS의 구성도를 나타내고 있다. JANUS는 구어체를 입력으로 받아 음성인식부에서 N개의 인식 후보를 출력하고 이를 LR 파서, NN 파서 및 의미 파서를 동시에 이용하여 각각 중간언어로 해석결과를 내는 혼합형 구문해석 방식을 채택하고 있다.

그리고 현재 국외에서는 음성인식 시스템과 자연언어 처리 시스템의 일관성 있는 상호 보완적인 인터페이스 모듈의 구성으로 전체 시스템의 성능 향상을 이루기 위한 논의가 일부 이루어지고 있지만 구체적인 방법과 정보의 형태가 제시된 바 없다. 국내에서는 현재까지 음성인식에 관한 연구와 자연언어 처리에 관한 연구가 개별적으로 진행되어 왔으며 이에 대한 통합 시스템의 구성을 목표로 전자통신연구소(ETRI)에서 대화체 음성언어번역 통신시스템에 관한 모델을 제시한 바 있다.

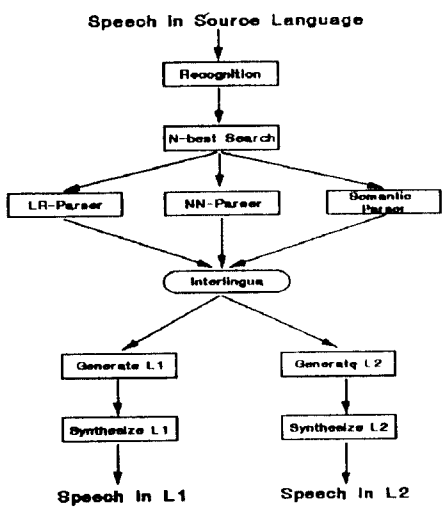


그림 2-2. 음성언어 번역 시스템 JANUS 시스템의 구성도

음성언어 번역 시스템은 그림 2-3과 같이 크게 음성인식부, 기계번역부 및 음성합성부로 이루어진다. 음성인식부에서는 기계번역부로의 입력문장과 함께 화자 식별 정보, 발화 간격, 상세, 억양 등의 부수적인 양상 정보를 전달하여 번역의 데이타로 삼게 하며 역으로 기계 번역부에서는 음성인식부로 다음발화에 대한 예측정보를 전달하여 음성인식에서의 성공률을 높일 수 있도록 한다. 기계번역부에서는 음성합성부로 번역된 목표언어의 문장과 함께 발화 간격, 상세, 억양 등의 부수적인 양상 정보를 전달하여 자연스러운 낭독이 가능하게 한다.

현재 음성인식기로서 실용화되고 있는 것은 주로 단어단위로 발화된 음성을 대상으로 한 것이며 어휘수도 수백 단어 정도이다. 현재 연구의 대상이 되고 있는 것은 어휘수가 수백에서 수천개이며 연속 및 화자독립 음성의 인식이다. 음성언어 번역을 위한 음성인식으로도 연속 및 불특정 화자의 음성인식을 실현할 필요가 있다. 그러나 단어단위로 발성된 음성에 비하여 연속음성의 인식은 조음결합에 의한 음성의 변화에 대처해야 하기 때문에 많은 어려움이 뒤따른다.

그리고 음성언어 번역을 위하여 음성인식과 언어해석의 인터페이스부는 기계번역부에서의 언어처리를 위하여 음성에 수반하는 애매성을 해소하고 가장 확실한 후보를 선택할 필요가 있지만 음성인식과 언어처리를 어떻게 결합할 것인가 하는 문제가 생긴다. 이러한 인터페이스 방식은 음성인식과 언어처리를 각각 독립적으로 수행하는 방식과 동시에 수행하는 방식으로 크게 분류할 수 있다.

독립적 처리 방식에서는 음성인식부에서 먼저 Bi-gram이나 CFG 등의 문법적 제약을 이용하여 인식을 수행한다. 이 때 출력은 가장 확실한 후보 하나가 아니라 다수의 N개를 인식 후부로 출력한다. 이들 후보들은 각각 확실성을 나타내는 확률값들이 부가되어 있다. 그리고 언어처리에서는 이들 N개의 후보들 중에서 음성인식의 인식 확률값이외의 구문적, 의미적 및 문맥적 제약을 이용하여 가장 확실한 인식 후보를 결정한다. 동시 처리 방식에서는 의미 제약 등을 이용

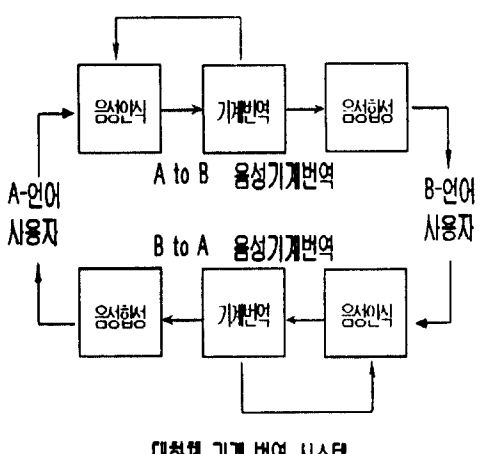


그림 2-3. 대화체 기계번역 시스템 구성도

하여 음성인식을 수행한다. 즉 음성인식이 종료되는 시점에 의미해석 결과도 얻어지게 되는 방식이지만 독립적 수행 방식보다 덜 선호되는 방식이다.

일반적으로 기계번역은 해석, 변환 및 생성의 3가지 단계로 분류된다. 우선, 해석에서는 음성인식의 결과들을 입력받아 구문 및 의미해석을 수행하지만 음성언어번역의 대상인 대화체는 생략이 빈번하고 어순이 비교적 자유롭기 때문에 전통적인 규칙기반의 해석 방식에 의해서는 해석에 많은 어려움이 따르게 된다. 일반적으로 규칙을 사용하는 해석기는 각 규칙들이 독립적이지 못하기 때문에 기존의 규칙들에 새로운 규칙을 추가하고 확장하기가 매우 어려운 것으로 알려져 있다. 따라서 현재 국외에서는 물론 국내에서도 대량의 말뭉치(Corpus)에 의한 통계적 데이타를 이용한 기계번역 방식이 활발히 연구되고 있는 중이다.

대량의 말뭉치를 이용한 통계적 기계번역 방식이란 이미 구문 분석이 된 문장들로 구성된 Corpus내에서 어떤 규칙이 Corpus내의 문장을 분석하기 위해 적용된 것을 하나의 확률 사건으로 간주하고 이 사건들의 발생에 대해 확률 모델을 적용하는 것이다.

음성합성부에서는 기계번역부에서의 출력 문장을 다시 자연스러운 음성으로 합성하여 출력하는 부분이다. 양질의 음성 출력을 위해서는 운율, 강세, 억양, 속도, 음색 등의 측면에서 세심한 배려가 요구되며 이는 출력 문장의 어휘적, 구문적 및 의미적인 특성과 무관하지 않다. 이상적으로는 음성입력에서의 이들 특성까지도 함께 감안하여 출력에서 이를 살릴 수 있다면 내용은 물론 어감까지도 전달하는 좋은 번역이 가능할 것이다.

이러한 처리가 가능하기 위해서는 기계번역부는 단지 최종적으로 얻어진 번역문만을 음성합성부로 전달해서는 곤란하며 음성합성부의 필요에 맞게 처리된 어휘적, 구문적 및 의미적 특성 및 음성 입력 상의 특성 정보를 번역문과 함께 전달하여야 할 것이다. 그러나 아직은 이 부분의 인터페이스에 대한 연구가 부족하여 어떠한 형태의 정보가 어떠한 형식으로 전달되어야 하는지에 관해서 정립되지 못한 상태이다. 다음 표 1은 국외에서 개발한 주요한 음성언어 번역 시스템들의 시스템명, 적용분야, 대상언어, 음성인식 모델, 번역 모델, 음성합성 모델 등을 비교하고 있다.

2.3 국내의 기계번역 현황

근래에 기계번역에 관한 관심이 높아지면서 일부 대학, 연구소 그리고 기업체에서 많은 연구를 하고 있지만 아직 실용화 단계의 시스템은 개발하고 있지 못한 실정이다. 국내의 기계번역에 관한 연구는 주로 영어와 일본어를 중심으로 한국어와 상호 번역하는 것이 주류였고 이는 현재도 활발히 진행되고 있으며 중

표 1. 음성언어 번역 시스템의 현황

개발기관		일본전기	ATR	CMU	Siemens	AT&T	SRI
시스템명		INTERTALKER	SL-TRANS/ ASURA	JANUS		VEST	
분 야		Ticket 예약	국제회의 참가 문의	국제회의 참가 문의	국제회의 참가 문의	은행창구	항공권 예약 안내
언 어		일→영, 스페인	일→영, 독	영→일, 독	독→일	영→스페인	영→스웨덴
규 모		500 단어	1500 단어	500 단어	700 단어		1400 단어
음성 인식	음소 모델	HMM	HMM	Neural Net	HMM	HMM	HMM
	언어 모델	Network 문법	CFG	bi-gram	bi-gram	Network 문법	bi-gram
음성 언어 인터페이스		음성인식 및 언어처리 동시 진행	N-best	N-best	1-best	N-best	N-best
언어 번역	해석		단일화 처리 구구조 문법	일반화 LR 어휘기능 문법	단일화 처리 구구조 문법	단일화 처리 구구조 문법	단일화 처리
	변환 생성	PIVOT 방식	속성구조 변환 의미주도형 생성	프레임 형식의 변환 생성	PLF의 변환 의미주도형 생성		QLF의 변환 의미주도형 생성
음성 합성		비균일합성 단위 (일본어)			diphone, triphone 등의 합성단위 (영어, 스페인어)	비균일 합성단위 (스웨덴어)	

국어와의 번역도 일부에서 시도되고 있다. 특히 한국어와 일본어 상호간의 번역에 관한 연구는 양국의 지리적, 문화적 특징으로 많은 관심의 대상이 되어 왔고, 상품화 단계에 이르는 제품도 일부 등장하고 있다.

국내에서 발표된 대표적인 시스템으로는 한국 과학 기술원의 영한 번역 시스템 MATE-EK(MAchine Translation Environment-English to Korean)와 한양대의 일한 번역 시스템 ATOM(jApanese To kOrean translation Machine) 등이 있으나 현재는 전체적인 시스템 구축보다는 한국어 전자 사전 구축, Tagged Corpus 구성, 코퍼스 분석 도구 구축, 형태소 해석기에서의 해석 모호성 해소, 해석 문법 규칙에 관한 연구, 한국어 생성 시스템에 관한 연구, 대화모델을 이용한 문맥처리에 관한 연구 등이 개별적으로 진행되고 있다.

그리고 국내에서의 기계번역은 대학 및 연구소를 중심으로 종래의 규칙 기반의 처리에서 벗어나 말뭉치(Corpus)를 이용한 통계정보 기반의 처리가 시도되고 있으며 현재 ETRI와 KT를 중심으로 음성언어 번역 시스템의 하부 시스템으로 대화체 기계번역에 관한 연구가 활발히 진행중에 있다.

Ⅲ. 통계론적 자연언어 처리

통계론적 자연 언어 처리는 실제로 사용되는 대량의 말뭉치(Corpus)에 기반한 새로운 처리 방법으로서 구조적 자연 언어 처리의 한계점을 극복할 수 있는 방법으로 최근에 이를 이용한 연구가 국내외에서 활발히 진행되고 있다. 그 동안 자연 언어 처리의 한 줄기를 이루었던 구조적 전산 언어학의 한계점을 극복하고 강건한 시스템 구축의 한 방편으로서 현재의 자연 언어 처리의 연구 방향을 주도하고 있다.

3.1 역사적 배경

통계 정보를 이용한 자연 언어 처리는 30년 전에 시작이 되었지만 처음에는 이에 관심을 둔 사람이 적었고 각광을 받지 못하였는데 최근에 이르러서 대량의 코퍼스 구성이 가능해지고 이를 처리할 수 있는 컴퓨터의 용량이 증대됨에 따라 구조적 자연 언어 처리의 문제점을 극복할 수 있는 새로운 방법으로 대두되고 있다. 따라서 이에 대한 연구가 국내외에서 활발히 진행되고 있다. 통계적 자연 언어 처리의 밑바탕은 코퍼스이며 코퍼스의 발달에 따라 통계적 자연 언어 처리도 발전하게 되었다.

자연 언어 처리는 1950년대에 Chomsky가 제안한 구조적 문법을 중심으로 연구가 진행되어 왔다. 통계 정보를 이용한 자연 언어 처리도 비슷한 시기에 출발을 하였지만, Chomsky는 통계 정보를 이용한 방법은 부적절하다고 보았지만, 그의 이론은 구조적 언어학의 이론적 바탕이 되어 왔다. 한 때 연구가 중단되었으며 1960년대를 전후하여 SEU(Survey of English Usage) 코퍼스와 대량의 자료를 저장하고 검색할 수 있는 컴퓨터의 출현으로 비약적인 발전을 하게 되었는데, SEU는 대화체와 텍스트 문장을 모두 포함하였다. 첫 번째 기계 가독(machine-readable) 코퍼스는 1960년대 초에 Brown University에서 개발된 Brown Corpus이었으며 이는 그 후로 텍스트 미국 영어의 표준 샘플이 되었다.

그 후 Lancaster-Oslo/Bergen은 Brown Corpus와 동일한 양식을 이용하고 영어의 서로 다른 다양성을 비교할 수 있는 LOB Corpus를 만들었다. 1975년에는 Jan Svartvik와 그의 동료들이 SEU Corpus의 대화체 부분을 기계 가독 양식으로 만들었으며 그 결과 LLC(London Lund Corpus)가 출현하였으며 이는 대화체 연구를 크게 자극하였고 많은 연구 프로젝트에 활기를 불어넣었다.

지금은 크기나 연구 목적에 따라 수많은 전산화된 코퍼스가 존재하며 이를 바탕으로 30년 전에는 예견할 수 없었던 코퍼스 기반 연구가 활발히 확장되고 있다. 코퍼스를 확장하는데는 두 가지 방법이 이용되는데 주로 코퍼스 개발 툴을 이용해서 자동 및 반자동으로 코퍼스를 만들고 있다. 국내에는 아직 공식적인 Corpus가 구성이 되어 있지 않고 몇 몇 학교나 연구 기관에서 현재 일부 코퍼스가 구성이 되었거나 구성 중에 있다.

3.2 품사 태깅

언어는 쓰임새에 따라 공통적인 성질이 있는데, 이를 구별하기 위해 표지를 달게 되며 이런 표지를 태그(tag)라고 한다. 품사(part-of-speech ; POS)가 태그의 대표적인 예인데 이는 형태소 해석이나 구문 해석에 있어서 필수적인 정보이다. 그런데, 한 단어는 유일한 품사를 가지는 것이 아니라 여러 가지의 품사를 가질 수 있다. 이와 같이 한 단어가 여러 가지의 품사를 가질 수 있는 경우를 품사의 모호성이라고 한다. 예를 들면 "감기는"이라는 단어를 형태소 해석한 경우 "감기(명사)+는(조사)"와 "감(동사)+기(접사)+는(어미)"와 같이 두 가지의 해석이 가능하다.

이와 같이 한 단어가 여러 가지로 해석이 가능한데,

이런 단어가 하나의 문장 속에서는 유일한 품사만을 가지게 되고, 그 올바른 품사를 결정하는 것이 품사 태깅이다.

3.2.1 품사 태깅을 위한 은닉 마르코프 모델(Hidden Markov Model : HMM)

HMM은 음성 인식에서 많이 이용하는 방법 중의 하나였는데 최근에는 자연 언어 처리에서도 이를 이용하려는 시도가 많이 이루어지고 있다. 그 주요한 이유는 HMM이 자동적 훈련(training)이 가능하고, 입력에 대한 품사 태깅의 예측이 가능하기 때문이다.

HMM에는 세 가지 중요한 패러미터가 있는데 이는 상태, 전이 확률, 관측 확률이다. 이를 품사 태깅에 적용시키기 위해 HMM의 각 패러미터에 대응시키면 상태는 품사 태그에 대응되고 관측 확률은 어휘 확률, 전이 확률은 문맥 확률에 대응된다. 문맥 확률은 확률 모델에 따라 uni-gram, bi-gram, tri-gram 등으로 나뉜다. 이는 확률 모델이 이웃하는 정보를 이용하는 갯수에 따라 분류하는 것이다. 품사 태깅이란 이를 이용해서 최적의 품사열을 구하는 것이며, 최적의 품사열을 효율적으로 구하기 위해 일반적으로 Viterbi Algorithm을 이용한다.

3.2.2 Tri-gram을 이용한 품사 태깅

Tri-gram을 이용한 품사 태깅은 문맥 확률을 구할 때 이웃하는 이전 또는 이후의 단어 중 2개의 단어 정보를 이용하는 확률 모델로서, 한국어의 경우 어절 단위 뿐만 아니라 형태소 단위의 품사 태깅이 연구된 바 있으며 최근에는 확률 정보와 규칙을 혼합함으로써 정확도를 높일 수 있는 연구가 활발히 진행되고 있다. 연구 결과에 의하면 97% 까지 정확도를 높이고 있다. HMM을 품사 태깅에 적용할 경우 시스템의 성능에 가장 영향을 미치는 것이 확률 모델이다. 정확도 측면에서는 더 많은 정보를 이용하는 것이 유리하겠지만, 그 만큼 확률 정보를 얻는데 많은 시간과 인력이 필요하며 확률 정보가 많아지므로 탐색 속도가 느려지며 시스템이 복잡하게 된다. 반면에 정보를 적게 이용하면 시스템은 간단해 지지만 정확도는 떨어지게 된다. 이로부터 정확도와 시스템 효율간에 트레이드오프(trade-off)가 발생하게 된다. 지금까지는 bi-gram이나 tri-gram을 이용한 확률 모델이 많이 실험되고 있다. 확률 모델도 언어에 따라 그 형태가 많이 차이가 나는데 한국어에 적용한 예를 보이면 다음과 같다.

위 그림 3-1은 어절 단위의 태깅을 위한 HMM을

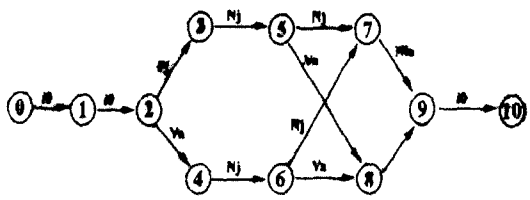


그림 3-1. 어절단위 태깅을 위한 HMM

일부 표현한 그림이다. 각각의 번호는 상태를 나타내며 이는 품사 태그에 대응하고 전이 확률은 이전 단계 두 단계를 조건으로 하여 확률을 구하게 된다.

예를 들어, "나는 학교에 가는 중이다"의 문장은 '나는(Pj, Va)', '학교에(Nj)', '가는(Nj, Va)', '중이다(Nia)'로 분석이 되는데, 이는 아래와 같은 네 가지의 품사 태깅이 가능하다.

① Pj Nj Nj Nia(0, 1, 2, 3, 5, 7, 9, 10)
② Pj Nj Va Nia(0, 1, 2, 3, 5, 8, 9, 10)
③ Va Nj Nj Nia(0, 1, 2, 4, 6, 7, 9, 10)
④ Va Nj Va Nia(0, 1, 2, 4, 6, 8, 9, 10)

각 경로에 대해 $P(I|V) = P(v_i|I_i)P(I_i|I_{i-1}I_{i-2})$를 적용하고 가장 높은 값을 갖는 경로를 선택한다.

3.3 확률적 구문 해석

자연 언어 처리에서 형태소 해석 다음 단계가 구문 해석 단계이다. 구문 해석을 한다는 것은 주어진 문법에 따른 그 문장의 구조를 밝히는 것을 말한다. 실제로 문장을 구문 해석하면 형태소 해석과 마찬가지로 문장 구조의 모호성이 발생하게 된다. 문장 구조는 정의된 문법에 의하여 생성될 수 있는 모든 구조가 발생 가능하지만 실제로 발생되는 문장 구조는 많이 나타나는 문장 구조도 있고 거의 발생되지 않는 문장 구조도 있게 된다. 이와 같은 문장 구조의 발생 빈도는 실제로 코퍼스를 분석하여 얻을 수 있고, 이 확률 정보를 이용하여 문장 구조의 애매성을 어느 정도 해결할 수 있게 된다. 확률적 구문 해석도 확률 모델에 따라 그 성능에 많은 차이가 있다.

확률 정보를 이용하려면 구문 분석이 된 대량의 코퍼스가 필요한데 가장 유명한 것은 PENN TREEBANK이다. 이는 1991년 펜실베니아 대학에서 만들어졌으며 각 단어에 대한 품사 태깅과 각 문장을 단위로 하는 개략적인 구문 구조를 밝혀 놓은 4백만개 이상의 단어로 이루어진 대량의 자연 언어 Corpus이다. 이는

48개의 품사 태그 집합과 14개의 구문 태그 집합을 사용하였다.

3.3.1 확률적 문맥 자유 문법

확률적 문맥 자유 문법은 문맥 자유 문법이 다시쓰기규칙(Rewrite Rule)에 의해 기술될 때 각각의 규칙이 균일하게 기술되는 것이 아니라, 확률 정보를 가지고 기술되는 문법을 말한다. 이 때의 확률 정보는 통계 자료를 분석해서 얻어진다. 그 한 예를 나타내면 아래와 같다.

1. S → NP VP	1.0	6. NP → N N	0.09
2. VP → V	0.386	7. NP → N	0.14
3. VP → V NP	0.393	8. NP → ART N	0.53
4. VP → V NP PP	0.221	9. PP → P NP	1.0
5. NP → NP PP	0.24		

3.3.2 PCFG에 의한 구문 분석의 애매성 해소

문장 확률은 그 문장 구조에 대한 확률이므로 구문 분석에 애매성이 발생했을 때 각각의 구조에 대한 문장 확률을 구하고, 보다 확률값이 큰 구조를 택함으로써 어느 정도 그 애매성을 해소할 수 있음을 예견할 수 있다. 영어 문장의 경우 가장 흔히 발생하는 구조적 애매성 중의 하나가 PP-Attachment 문제인데, 이에 대한 적용은 다음과 같다.

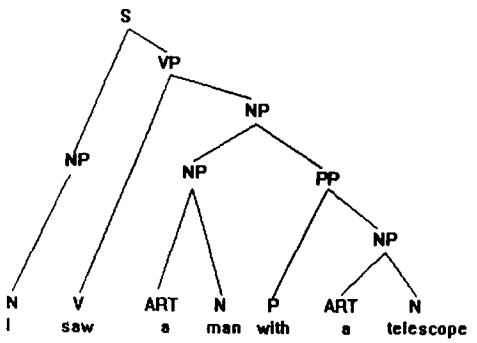
예문 1) I saw a man with a telescope.
예문 2) I put the ball in the house.
예문 3) I like the bird in the house.

예문 1)의 각 어절에 대한 품사는 다음과 같다.

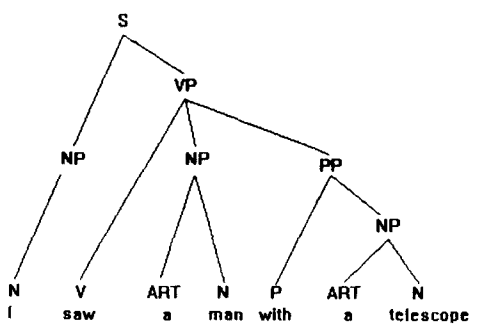
I	saw	a	man	with	a	telescope.
N	V	ART	N	P	ART	N

위 문장은 다음과 같은 두 가지의 분석이 가능하다. [그림 A]는 PP가 NP를 수식하는 경우이고 [그림 B]는 PP가 VP를 수식하는 경우이다. 각각의 문장 확률을 구해 보면 다음과 같다.

P(TS〈[그림 1]〉|[규칙 1]) = P(S → NP VP)
·P(NP → N)·P(VP → V NP)·P(NP → NP PP)
·P(NP → ART N)·P(PP → P NP)·P(NP → ART N)
= 1.0 · 0.14 · 0.393 · 0.24 · 0.53 · 1.0 · 0.53
≒ 0.0037



[그림 A]



[그림 B]

그림 3-2. 구조적 애매성의 일례

P(TS〈[그림 2]〉|[규칙 1]) = P(S → NP VP)
·P(NP → N)·P(VP → V NP PP)·P(NP → ART N)
·P(PP → P NP)·P(NP → ART N)
= 1.0 · 0.14 · 0.22 · 0.53 · 1.0 · 0.53
≒ 0.0087

여기서 어휘 확률은 공통이므로 계산에서 제거하였다. 계산 결과 P(TS〈[그림 A]〉|[규칙 1])이 P(TS〈[그림 A]〉|[규칙 1])보다 더 크므로 [그림 B]를 선택한다. 이는 코퍼스를 분석했을 때 PP가 NP 보다는 VP를 수식하는 경우가 더 많았음을 말해 준다. 그런데 이와 같은 방법을 다른 문장에 적용했을 경우 예문 2)의 경우는 올바른 결과를 내지만 예문 3)의 경우는 오히려 확률 값이 높은 구조를 택하는 것이 틀린 결과를 내게 된다. 서로 확률적으로 뚜렷이 구별되는 문장 구조의 경우 문장 확률이 높은 것을 택할 때 오류의 가능성은 적어지지만 PP-Attachment 문제와 같이 발생 확률이 비슷한 경우는 그만큼 오류의 가능

성은 커지게 된다. 이는 문맥을 고려하지 않고 일률적으로 규칙을 적용한 결과이며 정확성을 높이기 위해서는 문맥을 고려할 필요가 있음을 말해 준다.

3.3.3 PP-Attachment 문제

PP-Attachment 문제를 해결하기 위한 많은 연구가 있었는데, 이 중에서 선택적 제한(Selectional Restriction)이나 의미망(Semantic Network)을 이용한 방법이 주로 시도가 되었지만 의미망 구축의 어려움으로 좋은 결과는 얻지 못하였다. PP-Attachment 문제는 다시 말해 문장 구조가 "verb np1 (prep np2)"로 이루어져 있을 때 pp(prep np2)가 verb를 수식하는지, np1을 수식하는지를 밝히는 문제이다. 이를 확률 모델로 표현하면 다음과 같다.

$P(A \mid prep, verb, np1, np2)$

여기서 A는 verb 또는 np1을 나타내는 random 변수이다. 최근에 통계 정보를 이용한 해결 방법이 많이 제시되고 있는데 몇 가지 연구 결과를 소개하면 다음과 같다.

3.3.4 Hindle-Rooth의 실험

위 네 가지 정보를 조건으로 줄 경우 그 경우의 수는 명사와 동사의 어휘 수를 생각할 경우 천문학적인 숫자가 되고 이를 모두 구한다는 것은 불가능하다. 그래서 Hindle-Rooth는 전치사구(prepositional phrase ; pp)의 머리어 명사는 확률에 영향을 끼치는 않는다고 가정하고 다음과 같은 확률 모델로 실험을 하였다.

$P(A \mid prep, verb, noun_1, noun_2) = P(A \mid prep, verb, noun_1)$

$P(A = noun \mid prep, verb, noun_1) > P(A = verb \mid prep, verb, noun_1)$

$\Leftrightarrow f(prep, noun_1) > f(prep, verb)$

$$P(prep, verb) = \frac{Count(prep\ attached\ to\ verb)}{Count(verb)}$$

$$P(prep, noun) = \frac{Count(prep\ attached\ to\ noun)}{Count(noun)}$$

실험 결과 정확도는 실험 데이터가 적을 경우 약 78.1%, 실험 데이터가 충분히 많을 경우 85.0%의 정확률을 보였다.

3.3.5 Basili의 실험

Hindle-Rooth의 실험 결과는 더욱 높은 정확도를 요구하는 시스템에는 부적절함을 알 수 있다. 이를 해결하기 위해 머리어 명사까지 확률 모델에 넣는다면 실험 결과는 더욱 높은 정확도를 나타낼 지 모르겠지만 모든 확률 데이터를 구하면 약 10^9 개의 데이터를 구해야 되며 이는 실현 불가능하다. 그래서 Basili는 이를 해결하기 위해 품사 대신 의미 태그(semantic tag)를 이용하여 실험하였다. 예를 들어 "artist", "Jane", "plumber"를 human 이라는 의미 태그를 붙이고 "Friday", "June", "yesterday"를 time이라는 의미 태그를 붙였다.

Basili의 확률 모델은 다음과 같다.

$P(A = noun \mid prep, verb, noun_1, noun_2)$

$> P(A = verb \mid prep, verb, noun_1, noun_2)$

$\Leftrightarrow P(S(noun_2), S(noun_1) \mid prep)$

$> P(S(noun_2), S(verb) \mid prep)$

여기서 S(x)는 x의 의미 태그를 말한다. Basili는 임계치를 정의해서 그 임계치를 넘는 경우에만 결정을 하도록 하였는데 그 경우 85%의 정확도를 보여주었다.

3.4 국내의 연구 동향

영어의 경우는 국외에서 많은 연구가 활발히 되고 있고 국내에서도 한 두 차례 연구가 된 바 있다. 한국어의 경우는 실험의 바탕이 되는 코퍼스의 구축도 아직 되지 않은 상태이고 코퍼스가 구축이 되더라도 구문 태그가 된 코퍼스가 필요한데 이를 구축하는데는 많은 사람의 인력과 시간이 필요하므로 좀 더 시간이 지나야 구문 분석에도 통계 정보를 이용한 연구가 진행되리라 보여진다. 한국어에 적용한 예는 거의 없으며 최근에 형식 형태소와 구문 태그가 혼합이 된 확률 모델로 실험이 된 바가 있지만 코퍼스의 부족으로 만족할 만한 결과는 나오고 있지 않다.

Ⅳ. 자연언어 인터페이스에 관한 동향

Man-Machine Interface로 고려되는 방식들은 초보자를 위한 메뉴(menu)방식이나 전문가를 대상으로 하는 형식 언어(formal language)방식이 현재로서는 대표적이며, 그래픽과 같이 시각적인 요소를 중심으로 하는 방식들도 점차 주류를 이루어 가고 있다. 그러나 실제의 경우 인간과 가장 친밀한 인터페이스 방

식은, 일상 생활에서 인간 사이의 대표적 정보 교환 매개인 자연 언어를 사용하는 것으로 판단된다. 인간의 사고 작용이나 판단, 추론 등의 지적활동 결과들은 결국 자연 언어를 통해 다른 사람에게 전달되며, 이것은 음성이나 문자 등을 포함한다.

계산기에 자연언어의 기능을 부가하고자 하는 연구가 인공지능 분야에서 오랫동안 진행되어 왔다. 그러나 자연언어는 형식언어와는 달리 어휘의 방대함이나 문법의 복잡성, 자의적인 표현 양태, 주변 지식이나 상식에 의존하는 의미표현, 그리고 그 표현의 중의나 모호성 등이 실용적 시스템 개발에 난점이 되어 왔다. 그러나 이러한 문제들의 해결은 현대 과학 기술의 진보와 함께 다소 진척되고 있으며, 제한된 영역의 문제를 처리하는 자연 언어 시스템이 상품화되는 단계에 이르고 있다.

현재 자연언어 인터페이스 시스템은 이식성과 모듈성, 시스템의 자연언어 수용능력의 확대, 처리 이론의 최신성 등을 고려하여 연구가 활발히 진행되고 있으며 데이타베이스 검색 시스템, 대화체 음성번역 시스템, 관광 안내 대화시스템 등 기타 여러 응용 시스템의 기반 기술로 확장 가능하다. 또한 단순한 Text 정보뿐 만 아니라 그림과 음성 등의 정보를 포함하여 인터페이스하는 방법론도 고려되고 있다.

4.1 데이타베이스 검색을 위한 자연언어 인터페이스

계산기 과학의 입장에서는 다양한 형식의 많은 정보들을 관리하는 방법으로 데이터베이스 관리 시스템에 관한 연구가 중심이 되고 있다. 그러나 실제 일반 사용자들이 이러한 데이터베이스를 부담없이 효율적으로 사용하도록 유도하는 것이 보다 정보화 사회를 위한 근본적인 바탕이 된다고 하겠다. 이러한 측면에서 보면, 사용자가 얼마나 용이하게 데이터베이스에 접근할 수 있는가 하는 것이 중요한 요소가 되며, 보다 인간에 친밀한 접근 방법을 제공하는 인터페이스의 연구가 주목을 받고 있다.

따라서 자연언어 인터페이스는 장차 데이터베이스는 물론 기타 소프트웨어 패키지들의 주된 인터페이스 방식이 될 것으로 판단된다. 인터페이스의 기능은 사용자와 DBMS와 같은 응용 프로그램의 중재 역할이 우선적이라 할 수 있으며, 특히 자연언어 인터페이스(NLI, Natural Language Interface)의 경우, DB 검색을 위해 사용자가 반드시 배워야하는 질의어 등을 대치할 수 있다는 점에서 고부적인 역할이 기대되고 있다.

자연언어 DB 인터페이스 구축을 위해서는 자연언어 처리 관련 기술과 지식베이스 구축 기술, DBMS 질의어에 대한 처리 기술 등이 선행되어야 한다. 현재 자연언어 인터페이스의 핵심 부분인 자연언어 처리에 있어서는 해외의 경우 많은 연구 결과들이 발표되고 있으나, 한국어를 대상으로 한 처리 기술은 실용적인 수준에 이르지 못한다고 생각되며 기계번역 시스템, 인공지능 기법에 의한 안내 시스템 등의 관련 부분의 연구가 국내의 일부 연구기관에서 수행되고 있지만 정보검색을 위한 자연언어 처리 분야에 관한 연

표 2. DBMS 자연언어 인터페이스의 현황

Vendor	Product	Hardware Required
Artificial Intelligence Corp.	Intellect	IBM Mainframes DEC VAX, VMS
Battelle	Natural Language Query(NLQ)	IBM PC and Compatibles (For Most Mainframes)
Bolt, Beranek and Newman	Parlance	VAX, SUN LISPmachine SUN
Carnegie Group, Inc.	Language Craft	DEC VAXstation, TEXAS Instruments EXPLORER, Symbolics 3600
Dynamics Research	SPOCK	SUN, VAX
IBM	MBNLQ	IBM PC and Compatibles
Intelligent Business Systems, Inc.	EasyTalk	DEC MicroVAXII, VAX
McDonnell Douglas Computer System, Co.	Natural Language	Minicomputers
Natural Language, Inc.	Natural Language(DataTalker)	SUN, VAX
Online Software	English(Works with RAMISII)	IBM Mainframes
Programmed Intelligence, Corp	Intelligent Query	IBM, SUN
Stanford Research Institute	TEAM(only licensed to University)	Symbolics

구는 현재까지 부족한 실정이다.

선진 외국에서 상용화된 자연 언어 인터페이스 시스템들은 1980년대에 이르러서 개발된 것들이 대부분이다. AIC(Artificial Intelligence Corporation)의 Intellect는 1970년대에 개발된 이후로 10여년간의 꾸준한 확장으로 현재 다양한 시스템에 폭넓게 적용되고 있으며 그 이후 개발된 대표적인 DB 검색을 위한 자연언어 인터페이스 시스템으로는 IRUS, EUFID, TEAM 등이 있다.

IRUS는 영국 Cambridge 대학의 BBN Lab에서 만든 데이터베이스 검색을 위한 인터페이스 시스템으로 사용자의 질문에 대한 대답보다는 질문의 의미를 파악하고자 하였다. EUFID(End User Friendly Interface to Database Management Systems)는 미국의 J. Burger에 의해 개발된 Table 주도형의 데이터베이스 검색 시스템이다. 그리고 TEAM(Transportable English Database Access Medium)은 특정 영역의 질문만을 처리하고 하나의 데이터베이스 내에 정보를 유지하는 것을 가능하게 하는 등의 다른 자연언어 인터페이스의 단점들을 극복하기 위한 시도로 구현되었다.

언어 이해와 질의어 변환에 사용되는 지식 베이스 구축 기술에 관하여는, 지식 공학의 실용화 시스템인 전문가 시스템을 중심으로 전문 지식을 수집하여 지식베이스를 구축하여 이용하는 경우에 대하여 많은 연구와 함께 실용화된 시스템도 발표되어 있다. 그러나, 대부분의 경우 고장 진단 등의 분야에 한정되어 있고, 자연언어 인터페이스와 관련되는 담화이해 또는 대화시스템은 이론적 연구를 수행하고 있는 정도이다. DBMS 질의어에 대한 처리 기술은 논리적 표현으로부터 질의어를 생성하는 분야의 기술로서 최적화 등의 연구가 국내외에서 활발히 진행되고 있다.

4.2 지능형 대화 시스템

인간과 기계 사이에 대화를 통한 의사전달을 위한 대화 시스템에 관한 연구가 계속되어 왔지만 현재 대부분의 자연언어 인터페이스는 정보 요구자의 각 질의문을 그 대화 문맥에 대한 고려없이 하나의 고립된 정보 요구문으로 처리하는 단편적 질의응답 시스템의 수준을 벗어나지 못하고 있다.

자연언어에 의한 대화 시스템은 사용자가 그 내용을 형식에 구애받지 않는 일상의 회화문을 입력하면 의미를 파악하여 원하는 정보를 제공하고 자연스러운 대화를 유도하는 시스템이다. 이와 같이 자연스러운 대화가 이루어지기 위해서는 화자의 의도나 대화의 배경이 되는 상식 등을 시스템내에 형식화하여야 하고 대화의 전후 관계나 지시사 문제, 화제의 관리 및 일관성의 유지 등이 필요하다.

그림 4-1은 델라웨어 지능형 조언 시스템인 DIALS의 전체적인 개요도로서 일반적인 지능적 자연언어 인터페이스의 구성도를 보여주고 있다. 여기에서 사용자 모델(User Model)은 시스템의 대화 이해를 나타내며 사용자와 효율적인 정보 교환의 근거가 될 수 있는 현재의 대화 문맥에 대한 정보를 보유하고 있다. 사용자와 시스템 사이의 모든 발화문들은 사용자 인터페이스를 통해 전달되며 사용자 인터페이스는 사용자로부터 발화문을 입력받을 때 발화문 해석기를 가동시킨다. 그리고 발화문 해석기는 발화문과 사용자 모델에 대한 추론을 전개하며 사용자가 의도한 의미를 파악하고 그 파악한 의미는 이 후의 처리를 위하여 사용자 인터페이스에 전달한다.

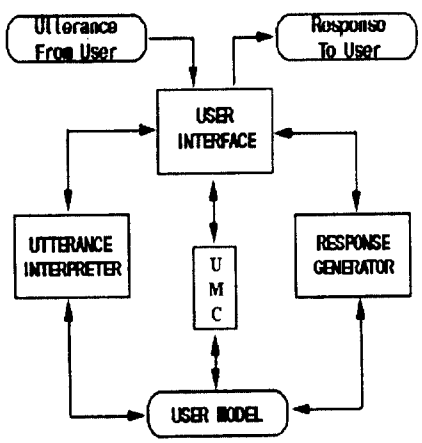


그림 4-1. 지능적 자연언어 인터페이스

일단 사용자의 발화문이 이해되면 사용자 모델링 성분 UMC는 사용자 모델을 갱신하고 사용자와 현 대화 문맥에 대한 시스템의 믿음에 대한 새로운 발화문을 반영한다. 그리고 시스템 응답기를 통해 사용자의 발화문의 의미와 사용자 모델에 대한 추론을 행함으로서 사용자가 획득하고자 한 필요한 정보를 적절히 제공해 준다. 사용자 모델링 성분은 시스템의 응답을 반영하도록 사용자 모델을 갱신하며 사용자 모델은 사용자가 이 후 대화상에서 추구하고자 하는 대화의 방향을 시스템이 이해하는 지식으로서의 역할을 한다.

그림 4-2는 지능적 대화 시스템의 일례로 1993년에

한양대에서 ETRI와 공동으로 개발한 바 있는 호텔예약을 위한 대화 시스템의 구성도를 보여주고 있다. 사용자가 입력한 한국어 문장은 언어 해석부에서 전처리 과정, 형태소 해석, 구문 및 의미 해석을 거쳐 의미표현식이 생성되며 생략 및 대용(Anaphora) 현상과 같은 문맥현상을 고려하는 문맥처리부는 결핍된 정보를 보완하고 이 해석정보들을 이용하여 대화처리에 필요한 정보를 대화정보 추출부에서 추출한다. 이후 사용자의 목표와 계획의 추론을 통한 응답을 수행하고 대화구조 모델을 갱신함으로서 시스템은 대화를 유도 및 관리할 수 있음을 알 수 있다.

그리고 현재 국외에서는 키보드를 통한 입력이 아닌 인간과 기계 사이에 음성을 통해 의사소통을 시도하는 음성대화 시스템에 관한 연구가 상당히 진척되고 있다. MIT의 항공여행 정보제공 시스템인 PEGASUS, Toshiba사에서 개발한 음성언어 대화 시스템 TOSBURG 그리고 국내에서는 KAIST에서 열차예매 영역에서 단어추출을 기반으로 한 음성 대화처리 시스템을 개발한 바 있다.

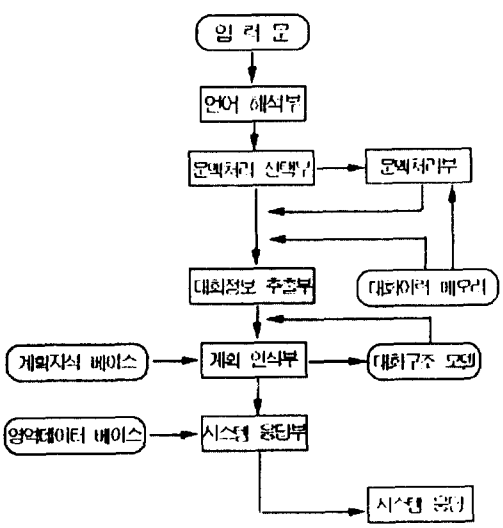


그림 4-2. 지능적 대화 시스템의 구성도

4.3 기타 자연언어 응용 시스템

자연언어 DB 인터페이스와 지능적 대화 시스템과 아울러 자연언어 응용 시스템으로서 널리 알려진 시스템으로는 문서교정 시스템과 정보검색 시스템을 들 수 있다.

문서교정 시스템은 문서 작성에 있어 틀린 철자나 맞춤법에 맞지 않는 단어를 비롯하여 어색한 문장을 자동적으로 교정해 주는 맞춤법 및 문법 검사 시스템을 말한다. 자연언어 처리를 이용한 문서교정 시스템은 워드프로세서에서 제공하는 철자 및 띄어쓰기 검사 기능에만 국한되는 것이 아니라 문장의 구조가 언어의 구문(syntax)에 맞지 않아 발생하는 구문 오류(syntatic error), 단어나 문장의 의미가 틀리거나 어색한 의미 오류 등을 교정할 수 있는 기능을 제공한다. 이 문서 교정 시스템은 그 자체로도 사용되지만 문자 및 음성인식 시스템의 후처리시에 유용한 기능을 발휘한다.

정보검색 시스템은 검색되는 정보의 유형에 따라 데이타 검색, 본문 검색, 참조 정보 검색 등 다양한 형태의 시스템이 개발되고 있는 실정이며 현재 인공적인 형식언어를 사용하여 질의어를 입력하는 시스템이 주를 이루고 있지만 자연언어 인터페이스를 부가하고자 하는 연구가 활발히 진행중이다. 그리고 지능형 CAI(Computer aided instruction), 지능형 교수 시스템, 무인 판매 시스템, 무인 전화번호 안내 시스템 등 자연언어를 이용한 Man-machine 인터페이스에 관한 연구가 다방면에 걸쳐 진행되고 있으며 곧 상품화 단계에 이르리라 예상된다.

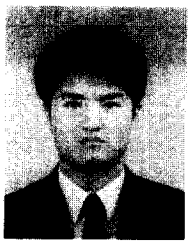
V. 결 론

이상에서 음성언어 번역 시스템의 연구를 중심으로 한 대화체 기계번역 시스템의 연구 동향을 살펴보았고 새로운 언어처리 방식으로 관심의 대상이 되고 있는 통계론적 처리 방식에 대한 개요와 동향을 기술하였다. 아울러 자연언어 처리의 기타 응용 시스템으로 자연언어 데이타베이스 검색 시스템, 지능형 대화 시스템, 문서교정 시스템 및 정보 검색 시스템 등 여러 다양한 분야에서 활발하게 연구가 진행되고 있음을 알 수 있었다.

자연언어 처리 기술은 선진 외국에서는 거의 실용화 단계에 이를 정도로 진전되어 있어 외국의 관련 기술 도입도 감토해 볼 수 있지만 한국어는 타언어와 구별되는 문법적/구문적 특성을 가지므로 결국은 국내의 기술로 개발되어야만 한다. 아울러 자연언어 처리에 관한 기술이 21세기의 정보처리 산업을 주도하고 멀티미디어 기술의 핵심기술로서 자리잡을 수 있도록 장기적인 안목에서 지속적인 투자가 이루어져야 할 것이다.

참 고 문 헌

1. HMM을 이용한 한국어 품사 태깅 시스템 구현, 임철수, 한국 과학기술원 석사학위 논문, 1993.
2. 통계 정보를 이용한 영어 구문 분석, 조 정미, 한국 과학기술원 석사학위 논문, 1992.
3. English Corpus Linguistics, Jan Svartvik, LONGMAN Inc., 1991.
4. Statistical Language Learning, Eugene Charniak, The MIT Press, 1993.
5. Natural Language Understading, James Allen, The Benjamin/Cummings Publishing Company Inc., 1995.
6. 애매성 해소를 위한 확률적 한국어 구문 분석기의 설계, 이 상윤, 대한 전자 공학회 하계 학술대회 논문집, 1995.
7. 대화체 일한 번역 시스템의 설계 및 구현, 강 석훈, 한양대학교 박사학위 논문, 1995.
8. 자연 언어 처리, 김 영택, 교학사, 1994.
9. "自然言語處理技術の應用", 情報處理, 1993. 10.
10. 대화체 및 문어체 기계번역을 위한 한국어 구문/의미 해석시스템 개발, 전자통신연구소 최종연구 보고서, 1995.
11. 지능형 질의응답 시스템의 설계에 관한 연구, 전자통신연구소 최종연구 보고서, 1993.
12. Plan Recognition in Natural Language Dialogue, Sandra Carberry, A Bradford Book, 1990.
14. "데이타베이스를 위한 자연언어 인터페이스 NAULI 이 설계 및 구현 (1) - 언어 해석 과정을 중심으로 - ", 우 요섭, 최 병욱, 대한전자공학회 논문지, vol. 28-B, no. 4, pp1-12, Apr. 1991.



김 영 길

- 1991. 2 : 한양대학교 전자통신공학과 (학사)
- 1993. 2 : 한양대학교 전자통신공학과 (석사)
- 1993. 3~현재 : 한양대학교 전자통신공학과 박사 과정

최 병 욱

- 1973. 2 : 한양대학교 전자공학과 (학사)
- 1981. 3 : 일본 KEIO 대학교 전기공학과 공학박사
- 1981. 8~현재 : 한양대학교 전자통신공학과 교수