

기계 번역 기술의 개요 및 동향

김 태 완
시스템공학연구소

I. 기계번역의 역사

기계번역(Machine Translation)이란 기계, 즉 컴퓨터를 이용하여 하나의 언어를 다른 언어로 변환 하여 주는 것으로서 자동 번역(Automatic Translation) 또는 컴퓨터 번역이라고도 한다. 기계번역에 대한 연구는 1940년대 후반 컴퓨터가 비수치 데이터의 처리에 응용될 수 있는 전형적인 연구 분야로 제안되어 1950년대에 미국, 러시아, 유럽, 일본 등지에서 많이 연구되었다. 특히, 당시 러시아와 대치 중이던 미국에서 러시아의 정보를 수집하기 위한 방안으로 러시아어-영어 기계번역에 많은 투자를 하였다. 그러나, 1966년 미국의 National Academy of Sciences에서 구성한 자동 언어 처리 자문 위원회(ALPAC; Automatic Language Processing Advisory Committee)가 제출한 보고서에서 “① 기계번역은 실용화를 위해서는 전산언어학적 연구가 선행되어야 하며, ②인간의 번역에 비하여 질이 떨어지고, ③경비도 크게 절감되지 않으며 ④기계번역 시스템의 개발에 상당한 기간이 소요되는 등 실용화가 요원하다[1]”는 내용의 보고서를 낸 이후, 미국에서는 기계번역에 대한 연구가 쇠퇴하게 되었다. ALPAC 보고서가 발표된 이후에도 영어에 대한 번역 수요가 막대한 일본, 회원 국가 언어 간의 번역이 필요한 EC 등에서는 꾸준히 기계번역에 대한 연구가 진행되었으며 1970년대 후반 하드웨어의 발달, 전산언어학의 발달 등 기계번역 기술 연구에 필요한 주변 환경의 발달에 힘입어 기계번역에 대한 연구가 다시 부흥하기 시작하였으며 현재에는 전세계적으로 100여개의 상용 기계번역 시스템이 발매되고 있다. 국내에서도 영한 기계번역 시스템, 일한 기계번역 시스템이 발매되고 있다.

II. 기계번역에 대한 견해와 검토

외국어를 어느 정도 공부한 사람에게는 번역이

라고 하는 작업이 명시적으로나 암시적으로 수행하고 있는 일상 업무이기도 하다. 따라서, 기계번역에 대해 개개인이 나름대로의 기대와 평가를 쉽게 내릴 수 있는 분야이기도 하다. 실제로 기계번역 기술이 제안된 이후부터 현재까지 기계번역에 대한 여러 가지 견해들이 존재하여 왔다. 이것들은 기계번역에 대한 선입견, 편견 또는 과잉 기대에서 비롯된 것들이다. 이러한 견해들에 대한 검토를 하여 봄으로써 기계번역에 대한 올바른 시각을 정립할 필요가 있다.^[2]

견해 1: 기계번역 시스템의 번역 품질이 낮아서 실용적으로 쓸 수 없다.

이러한 견해는 기계번역에 대해 인간과 동일한 수준의 번역을 기대하는 것으로부터 비롯된다. 현재 기술 수준으로 기계번역 시스템이 완벽한 번역 품질을 보장할 수 없다. 그러나 이것이 곧 기계번역 시스템이 쓸 모 없다는 것은 아니다. 그 이유는 첫째, 번역이 반드시 완벽할 필요는 없다. 만약 당신이 일본어를 모르지만 일본어 신문으로부터 유용한 정보를 얻고 싶다면 전체적인 의미만을 파악하는 수준의 번역도 상당히 유용할 것이며 완벽한 번역이 필요한 기사가 무엇인지 파악하는데 도움을 줄 수 있다. 둘째, 인간의 번역 작업에도 3, 4 차에 걸친 공정이 있다. 일단 대략의 번역(1차 번역)을 한 후, 해당 언어 전문가의 수정(2차 번역), 해당 분야 전문가의 보완(3차 번역) 및 검수를 통하여 완벽한 번역을 만들어 낸다. 기계번역 시스템은 여기서 1차 번역에 해당하는 번역 공정을 자동화하는 도구로써 이용될 수 있다. 1차 번역의 자동화에 따라 번역 전문가는 보다 고품질의 번역을 대량으로 생산할 수 있게 될 것이다.

견해 2: 기계번역 시스템이 번역가의 직업을 빼앗을 것이다.

기계번역 기술 연구 초기에 많은 사람들이 기계번역 시스템이 완성되면 많은 번역가가 직업을 잃을 것이라는 생각을 하였다. 그러나, 이것은 기우에 지나지 않는다. 국제적 교류의 증가에 따라 번역에 대한 수요도 번역가가 감당하지 못할 정도로 폭발적으로 증가해 가고 있다. 또한, 기계번역 시스템은 번역가들의 직업을 빼앗기보다는 번역 생

산성을 제고하는 도구로서 자리 잡을 것이다. 즉, 고통스럽고 지루한 1차 번역을 자동화함으로써 번역가들이 보다 고차원적이고 지적인 번역 작업에 몰두하게 하여 번역 생산성은 물론 번역품질의 향상을 기대할 수 있다.

견해 3: 기계번역 시스템은 어떠한 전문분야, 어떠한 문체의 문장도 번역할 수 있어야 한다.

인간이 번역을 행하는 데에는 다음과 같은 지식이 필요하다.

- 번역 대상언어(Source Language)에 대한 지식
- 번역 목표 언어(Target Language)에 대한 지식
- SL과 TL 간의 대조 언어학적 지식
- 일반적인 상식, 전문 분야에 대한 지식
- 문화, 관습, 사고 방식 등에 관한 이해

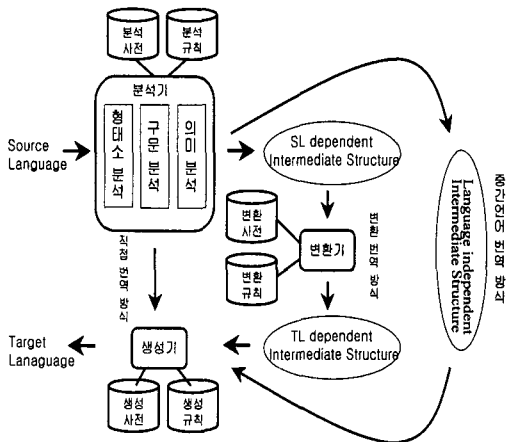
이러한 지식을 모든 전문분야나 장르에 대해 섭렵하는 것은 인간에게도 어려운 일이다. 따라서, 기계번역의 쓰임새는 한정적이고 특수 목적 지향적이 되어야 한다.

III. 기계번역 방법론

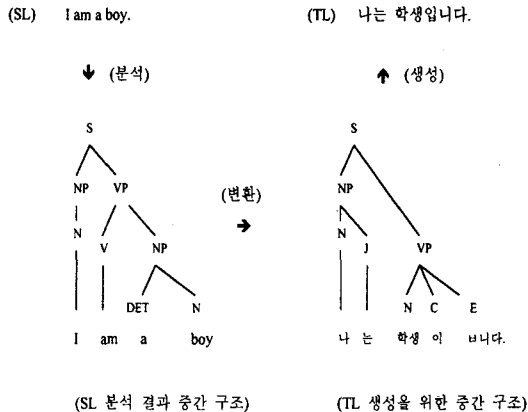
기계번역 기술에는 규칙 기반 기계번역(RBMT: Rule-Based Machine Translation), 예제 기반 기계번역(EBMT: Example-Based Machine Translation), 통계 기반 기계번역(SBMT: Statistical-Based Machine Translation), 코퍼스 기반 기계번역(CBMT: Corpus-Based Machine Translation), 지식 기반 기계번역(KBMT: Knowledge-Based Machine Translation) 과 같은 기술들이 연구되고 있다. 본 장에서는 그 중 대표적인 규칙 기반 기계번역, 예제 기반 기계번역 기술에 대해 간략히 설명한다.

1. 규칙 기반 기계번역

90년대 초반까지 기계번역 기술의 주류를 이루



〈그림 1〉 RBMT의 시스템 구성



〈그림 2〉 변환 방식 기계번역의 처리 과정

어 은 번역 방식으로서 가장 역사가 오래되었으며 따라서 기술적으로 안정되어 있는 방식이 규칙 기반 기계번역(RBMT)이다. RBMT의 기본적인 시스템 구성은 그림 1과 같다. RBMT는 기본적으로 형태소 분석, 구문 분석, 의미 분석, 번역 대상 언어(SL: Source Language) - 번역 목표 언어(TL: Target Language) 간의 변환, TL 생성 단계를 거치게 되며 분석과 변환의 깊이에 따라 직접 번역 방식, 변환 번역 방식, 중간언어(Interlingua) 번역 방식으로 나누어진다. 분석, 변환, 생성에 필요한 지식을 규칙으로 표현하여 이용하므로 규칙 기반 기계번역 방식이라고 한다.^[4,5]

1) 직접 방식 기계 번역

직접 번역 방식은 한국어와 일본어, 영어와 불어 등과 같이 언어 구조가 유사한 언어 간의 번역에 사용될 수 있는 것으로 Source Language를 분석하여 Target Language로 직접 대응될 수 있는 단어나 구를 추출한 후 대응하는 대역어로 대체하여 줌으로써 번역을 수행하는 방식이다.

- (SL) 私は學生です
- (SL 분석) →私 / は / 學生 / です
- (SL-TL 대체) →私-나 / は-는 / 學生-학생 / です-입니다.
- (TL 생성) →나는 학생입니다.

2) 변환 방식 기계 번역

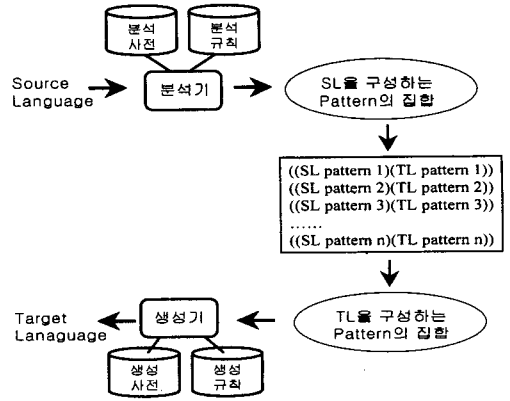
변환 방식은 그림 2와 같이 Source Language

를 분석하여 그 언어적 구조와 의미를 일정한 중간 구조(Source Language Oriented Intermediate Structure)로 표현하고 그것을 번역 목표 언어의 언어적 구조와 의미를 반영한 중간 구조(Target Language Oriented Intermediate Structure)로 변환한 후, 번역 목표 언어를 생성해 내는 3단계 과정을 거치게 된다. 변환 방식은 한국어와 영어, 일본어와 영어와 같이 언어적 구조가 상이한 언어 간의 번역에 많이 채택되고 있는 방식이다.

3) 중간언어 방식 기계번역

변환 방식 기반 기계 번역에서는 SL의 언어적 특성이 반영된 중간 구조와 TL의 언어적 특성이 반영된 중간 구조의 두가지 중간 구조가 필요하다. 이에 비해 중간언어(Interlingua) 방식은 SL이나 TL의 언어적인 특성이 배제된, 순수하게 문장의 의미만을 나타내는 하나의 중간언어(Interlingua)만을 사용한다. 즉, SL를 분석하여 Interlingua를 만들고 그로부터 TL을 직접 생성해내는 방식이다. Interlingua란 한국인이 “나는 학생입니다.”라는 한국어 문장을 보았을 때나 미국인이 “I am a student.”라고 하는 문장을 보았을 때나 인간의 두뇌에 만들어지는 “나는 학생입니다.”라는 의미를 나타내는 의미 표현은 한국인이든 미국인이든 동일하다고 생각하는 것으로 이때 인간의 두뇌에 형성되는 그 의미 표현을 말한다. 이 방식은 기본적으로 인간의 번역 처리를 모사한 것으로 SL - TL

간의 변환 처리가 필요하지 않다. 그러나, 인간의 두뇌에 만들어지는 의미 이해 구조(Interlingua)의 규명, SL로부터 Interlingua를 만들어 내기 위한 분석 처리, Interlingua로부터 TL을 만들어 내는 생성 처리에 필요한 지식에는 어떠한 것들이 있으며, 그 지식들이 어떻게 유기적으로 관계하는지에 대한 연구가 불충분하여 일반적으로 사용되고 있지 않은 방식이다. 단지, 극히 제한된 범위에서의 연구가 일본의 NEC(Pivot System), 미국 카네기 멜론 대학(Knowledge Based Machine Translation System)에서 연구된 바 있다.



<그림 3> EBMT의 시스템 구성

2. 예제 기반 기계번역(Example Based MT)

규칙 기반 기계번역은 번역 지식을 condition과 action으로 구성된 규칙으로 표현하고 처리한다. 이러한 규칙은 규칙간에 간섭을 발생시켜 전혀 다른 의미의 번역 결과를 생성하거나 모든 가능한 규칙을 적용시켜 여러 개의 번역 결과를 생성한다. 또한, 인간이 대량의 규칙을 작성, 수정, 관리하는데에 큰 어려움이 따른다. 이와 같은 문제점을 보완하고자 최근에는 예제 기반 기계번역 기술에 대한 연구가 활성화되고 있다. 예제 기반 기계번역이란 SL에서 나타나는 문구조를 번역에 편리한 구성 단위별로 패턴화하고 각 패턴에 대한 대역 구조를 데이터 베이스화 하여 규칙 기반 기계번역 방식의 변환 단계를 생략하는 방법이다. 예제 기반 기계번역 시스템의 전형적인 구성을 그림 3에 보인다.^[4]

IV. 기계번역 시스템의 기술적 과제

기계번역 기술은 자연언어처리 기술을 포괄하는 기술로서 모든 자연언어처리 기술의 집합체이다. 따라서, 현재 자연언어처리기술이 해결하고자 하는 모든 기술적 과제가 곧 기계번역 기술이 해결하고자 하는 과제이다. 이에 더하여 기계번역 기술은 타 언어로의 변환, 타 언어의 생성이라고 하는 부가적인 기능을 수행하여야 한다. 따라서, 해결해야 할 기술적 과제도 더 광범위하다고 할 수 있다.

이 장에서는 기계번역 기술이 해결하고자 하는 기술적 과제에 대하여 설명한다.^[5]

1. 언어 분석의 문제(Analysis Problem)

1) 어휘적 모호성(Lexical Ambiguity)

하나의 단어가 2가지 이상으로 분석이 가능한 것을 나타낸다. 여기에는 Category ambiguity, Homography, Polysemy, Transfer ambiguity 등이 있다.

Category ambiguity

하나의 단어가 여러 가지 품사를 갖고 있는 경우이다. 분석 단계에서 올바른 품사를 선택하지 못하면 전혀 엉뚱한 번역문이 만들어진다.

예) I can do it : 나는 그것을 할 수 있다.(can - 조동사) 나는 깡통 그것 한다.(can - 명사)

Homography

하나의 단어가 전혀 다른 두가지 의미를 가지고 있을 경우이다. 역시 올바른 의미로 분석하지 못하면 전혀 다른 의미의 번역이 되어 버린다.

예) Bank of Ohio : 오하이오 은행, 오하이오 강둑

Polysemy

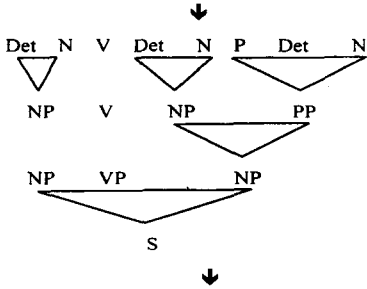
하나의 단어가 의미는 유사하나 전혀 다른 상황에서 쓰여질 수 있는 경우이다.

예) mouth of river : 강의 입 (?), 하구

2) 구조적 모호성(Structural ambiguity)

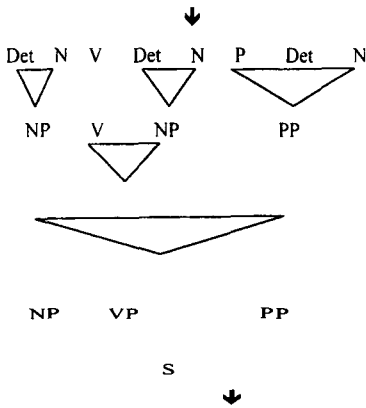
SL 문장을 구문 분석(Parsing) 하였을 때 여러 가지 Parsing Tree가 만들어지는 경우이다. 분석된 Parsing Tree의 구조의 차이가 번역문의 의미에 큰 차이를 발생시킨다. (그림 4)

The man saw the girl with the telescope.



그 사람은 망원경을 가지고 있는 소녀를 보았다.

The man saw the girl with the telescope.



그 사람은 그 소녀를 망원경으로 보았다

(그림 4) 구조적 모호성의 예

2. 언어 변환의 문제(Transfer problem)

기계번역은 4.1에서 설명한 어휘적 모호성, 구조적 모호성 외에 타 언어로의 변환 시에 발생하는 다음과 같은 문제점을 해결할 필요가 있다.

1) Lexical differences

SL에서는 ambiguity를 갖고 있지 않은 단어가 TL에서는 여러 가지 대역어로 번역될 수 있는 경우를 말한다. Stylistic translational ambiguity, Grammmatical translation ambiguity, Conceptual

translational ambiguity가 있다.

Stylistic translational ambiguities

대역어의 선택이 상황에 따라 달라지는 경우이다. 예를 들면 불어에서 'domicile'이라고 하는 단어는 영어에서 문서의 성격에 따라 'home'으로도 'domicile'으로도 번역될 수 있다.

Grammmatical translational ambiguities

SL의 단어가 문법적 특성에 따라 TL에서 다른 대역어를 선택하여야 하는 경우이다. 예를 들면 영어를 불어로 번역할 때, 영어의 'know'라는 단어는 직접 목적어로 명사구를 취할 경우에는 'connaitre'로 종속절이나 부정사를 취할 경우에는 'savoir'로 번역되어야 한다.

I know the right answer. → Je *connais* la bonne response.

I know what the right answer is. → Je *sais* quelle est la bonne response.

Conceptual translational ambiguity

기계번역에서 가장 문제가 되는 것으로서 기계 번역 방법론과 시스템 설계에서 중점적으로 다루어지는 것이다. SL에서는 하나의 개념을 가지고 있으나 TL에서는 해당 되는 개념을 갖는 단어가 존재하지 않아 여러 가지 단어로 분화되는 경우이다. 이것은 한국어와 일본어, 영어와 스페인어와 같은 유사 언어간에도 자주 발생한다. 예를 들면, 영어의 'leg'라는 단어는 스페인어에서는 'pierna' (인간의 다리), 'pata'(동물이나 테이블의 다리), 'pie'(의자의 다리) 로 상황에 따라 대역어가 달라져야 한다.

2) Structural differences

영어에서는 형용사가 일반적으로 수식하는 명사의 앞에 오지만 불어에서는 형용사가 명사의 뒤에 온다. 또한 전체적인 문장 구조에서도 영어와 불어간에 차이가 있다. 예를 들면

Ces livres se lisent facilement

These books read themselves easily.(구조 유지된 번역)

These books are easily read.(같은 의미의 자연스러운 영어)

Ici on parle anglais.

Here one speaks English.(구조 유지된 번역)

English is spoken here.(같은 의미의 자연스러운 영어)

와 같이 유사한 언어 구조를 가지고 있는 언어 간에도 구조적인 차이가 있다. 이러한 차이는 한국어와 영어와 같이 전혀 다른 언어 구조를 갖고 있는 언어 사이에서 더욱 큰 문제를 발생시킨다. 국지적인 구조의 차이 뿐만 아니라 전체 문장의 구조가 판이하게 다르기 때문이다. 간단한 단문(simple sentence) 구조에서도 영어는 “주어 + 술어 + 목적어”의 어순이지만 한국어는 “주어 + 목적어 + 술어”의 형태를 취한다. 이러한 구조의 차이는 여러 개의 절을 포함하는 복문의 경우 더욱 커지며 양 언어간의 번역을 어렵게 하는 주요 요인이 된다.

V. 국내의 기계번역 기술 연구 동향

1. 국외 주요 프로젝트

EUROTRA

EC 회원국 언어간의 기계번역 시스템을 개발하기 위하여 1983년부터 1992년까지 수행되었다. EUROTRA 프로젝트의 결과를 기반으로 하여 각 회원국은 언어 지식의 보강, 사전, 전문용어의 재사용을 위한 연구를 자체적으로 진행하고 있다.

CSTAR

카네기 멜론 대학(미국), ETRI(한국), Automatic Telephone Research(일본), Siemens(독일) 등이 참여하여 수행 중인 국제 공동 연구 컨소시엄으로서 자동 통역 전화의 개발을 목표로 하고 있다. 1995년 제한된 어휘수(약 500 단어)로 구성된 한국어-일본어 간 호텔 전화 예약 대화문 통역 시연회를 KT에서 가진 바 있다. 1999년에 수천 단어 규모의 자동 통역 시스템의 시연을 목표로 추진 중이다.

2. 국내 주요 연구

국내의 주요 연구 활동을 연구 기관을 중심으로 설명하기로 한다.

KAIST

KAIST는 시스템공학연구소와 공동으로 순수 국내 기술에 의한 최초의 영한 기계번역 시스템 MATES/EK를 1992년에 개발하였다. 현재는 예제 기반 기계번역 방식에 따른 한영 기계번역 시스템의 개발을 추진 중이다.

ETRI

자동 통역 전화 개발을 위한 국제 컨소시엄인 CSTAR에 회원국으로 참여하여 음성 언어 번역을 수행 중이다. 인간의 음성으로 발화되는 음성 언어는 정제되어 저장되는 텍스트보다 난이도가 높고, 음성 인식, 음성 합성 기술의 확보도 필요하다. 따라서, 음성 처리 기술의 연구에 보다 많은 비중을 두고 연구 중이다. 음성 처리 기술이 음성 언어 번역의 병목(bottle neck)이라고 할 수 있기 때문이다.

서울대

미국 IBM과 공동으로 영한 기계번역 시스템을 개발하였다. 개발된 영한 기계번역을 상품화 한 것이 시중에서 발매 중인 앙코르 영한 번역 시스템이다. 1997년 SEK에서 한일 기계번역 시스템을 발표하는 등 활발한 연구 활동을 전개하고 있다.

포항공대

일한 기계번역 시스템 COBALT/JK를 개발하였다. 일본어 언어 패턴에 기반한 일본어-한국어 번역 ambiguity 해소에 역점을 두었다. 중한 기계번역 시스템의 개발을 위한 준비를 진행 중이다.

시스템공학연구소

일본 후지쓰와 공동으로 국내 최초의 일한 기계번역 시스템인 ATLAS/JK를 1990년에 개발하였다. ATLAS/JK는 후지쓰의 메인 프레임 컴퓨터인 M 계열 컴퓨터에서 운용되므로 일반 사용자들이 사용하지 못하였으나, 현재도 연구개발정보센터에서 일본의 JOIS DB의 번역 서비스에 사용되고 있다. 또한, KAIST와 공동으로 영한 기계번역 시스템 MATES/EK를 개발한 바 있다. 현재는 영한, 한영 기계번역 기술의 향상을 위한 언어 대

〈표 1〉 국내 기계번역 시스템 제품

제품명	개발 기관	번역 대상 언어
한글 가나	창신 컴퓨터	일한
J-Seoul	고덴샤	일한
오경박사	유니소프트	일한
명품	히다찌 정보 네트	일한
앙코르	서울대	영한
워드체인지	정소프트	영한
Trannie	언어공학연구소	영한

이터의 구축과 웹 페이지 전문 고품질 영한 기계 번역 시스템의 개발을 추진 중이다.

이 밖에도 1995년부터 국내에는 상품화된 일한, 영한 기계번역 시스템이 표 1에서와 같이 다수 발표되었다. 기계번역 시스템 제품이 국내에서도 다수 발표되고 많은 기업체가 기계번역 시스템의 개발에 큰 관심을 갖게 된 것은 상당히 고무적인 일이다. 그러나, 아직 번역 품질에 있어서는 개선되어야 할 점이 많으며 각 기업체의 독자적인 개발력으로는 품질 향상을 위한 대량의 언어 지식 데이터의 구축, 언어 분석/변환/생성 기술의 기반 확보에 한계가 있다는 것이 문제점으로 지적될 수 있다. 이러한 문제점을 극복하기 위한 개발 전략의 수립이 필요하다.

VI. 기계번역 기술의 미래

번역이란 인간에게 있어서도 고도의 지식과 경험이 필요한 지적인 업무이다. 이러한 고도의 지능을 요구하는 업무를 컴퓨터가 수행하게 하기 위해서는 앞으로도 많은 시간과 노력이 필요하다. 그러나, 기계번역이 먼 장래에 겨우 손에 넣을 수 있는 것은 아니다. 실제로 외국의 경우에는 기계번역 기술을 활용하여 번역 업무의 생산성을 제고하는 등의 사례가 있다.^[6,7]

Sublanguage Machine Translation

Sublanguage란 항공 관제 언어, 군사 언어와 같이 특수 분야에서 정형화된 패턴만을 사용하는

언어를 말한다. 이러한 언어는 무제한적인 언어 사용에서 야기되는 ambiguity를 포함하고 있지 않으므로 현재의 기계번역 기술을 즉각 활용할 수 있는 분야이다.

Controlled Input

Controlled language란 의미를 문장으로 표현함에 있어서 제한된 어휘와 문장 패턴만을 사용하는 것을 말한다. Sublanguage보다 덜 제약적이다. Controlled language는 주로 제품 매뉴얼이나 과학 기술 문헌의 저작에 많이 사용된다. Controlled language 개념이 도입되어 저작된 문서는 문장의 복잡도가 낮고 사용 어휘도 적으므로 기계번역이 해결해야 할 문제의 범위도 상당히 줄일 수 있다. 따라서, 기계번역의 효율을 높일 수 있다.

Human Aided Machine Translation/Machine Aided Human Translation

현재의 기계번역 기술은 5장에서 설명한 각종 ambiguity 문제를 완벽하게 해결하지 못하고 있다. 따라서, 이러한 ambiguity를 해소하는 데에 인간의 도움을 얻는 방법이다. 기계번역 시스템은 ambiguity가 발생하면 모든 가능한 경우를 사용자에게 제시하고 사용자가 그 중에서 올바른 것을 선택하여 주는 방법이다. 번역의 주체와 보조 역할의 담당에 따라 Human Aided Machine Translation, Machine Aided Human Translation으로 분류된다.

인간의 언어 생활에 기반한 기계번역 기술의 개발

지금까지의 기계번역 시스템은 번역 대상 언어, 목표 언어의 교과서적인 문법서에 기초하여 SL의

분석, SL - TL 간의 변환, TL의 생성 규칙을 작성하고 그로부터 인간이 실제 사용하는 언어를 번역하고자 하는 연역적 방식을 취하여 왔다. 이러한 연역적 번역 규칙 개발 방법은 문법서에서 제시하는 단순한 언어 현상별 번역 규칙의 작성과 단순한 문장의 번역에는 쓸모가 있다. 그러나, 실제 인간이 사용하는 언어는 하나의 문에 여러 가지 언어 현상이 복합되어 있거나, 문법 요소 간의 경계가 모호하거나, 나아가서 비문법적인 문장이 나타날 수 있다. 따라서, 정통적인 문법서에 기초하여 아무리 정밀하게 번역 규칙을 작성하여도 실제 인간이 작성한 언어를 번역할 수 없는 경우가 상당히 많이 발생한다. 이러한 문제점을 해결하기 위해서는 실제 인간이 사용하는 문장을 대량으로 수집하고 그로부터 번역 규칙을 추출함으로써 기계번역 시스템이 보다 인간의 언어 생활에 가까운 번역 규칙을 갖도록 함으로써 번역 품질을 높일 수 있도록 하는 귀납적 번역 규칙 작성 방법이 필요하다. 또한, 대량의 번역 예문(Bilingual Corpus)를 축적하여 놓고 SL과 가장 유사한 문장 구조, 어휘를 갖는 예문을 선택한 후, 그와 대응되는 번역문으로부터 TL을 생성해내는 예제 기반 기계번역에 대한 연구도 활성화되고 있다.

참 고 문 헌

- [1] W. John Hutchins, Machine Translation : Past, Present, Futures, John Willey & Son, 1986
- [2] Doug Arnold, Lorna Balkan et al, Machine Translation : An Introductory Guide, <http://clwww.essex.ac.uk/~doug/book.html>
- [3] Bonnie Jean Dorr, Machine Translation : A biew from the Lexicon, The MIT Press, 1992
- [4] W. John Hutchins, Harold L. Somers, An Introduction to Machine Translation, Academic Press, 1992
- [5] Sergei Nirenburg, Machine Translation - Theoretical and methodological issues, Cambride University Press, 1987
- [6] Christian Rohrer, The Future of MT Technology, Proceeding of 4th MT Summit, pp 195 - 196, 1993
- [7] Tsujii Junichi, After Linguistics-based MT, Proceeding of 4th MT Summit, pp 197 - 200

저 자 소 개



金 泰 完

1960年 6月 13日生

1983年 2月 한양대학교 전자공학과 졸업

1985年 2月 한양대학교 대학원 전자공학과 졸업(석사)

1991年 현재 한국과학기술원 전산학과 박사 과정

1985年 2月~현재

시스템공학연구소 자연어정보처리연구부

주관심 분야: 기계번역, 한국어 정보 처리, 다국어 정보 검색