

연합작전을 위한 군 자동화 시스템에서의 기계번역

정광렬, 김성용, 김철호
국방정보체계연구소

Machine Translation in an Automated Defense System for Combined Operation

Gwang Ryeol Jung, Seong Yong Kim and Cheol Ho Kim
Institute for Defense Information Systems

요약 : 본 논문은 연합 작전을 위한 군 자동화 시스템에서의 기계번역에 대하여 설명한다. 연합작전은 단일한 임무의 수행을 위하여 공동 행동을 취하는 둘 이상의 연합국 부대에 의하여 실시되는 작전이며, 작전 구성원들이 사용하는 언어는 한국어와 영어 등 2개 국어 이상을 동시에 사용하고 있으므로, 상호간의 의사소통을 원활히 하기 위해서는 정확한 번역 시스템이 필요하다. 현재 사용중인 군 자동화 시스템에서 기계번역의 개발과 시험 내용을 사용자의 요구사항, 시험 항목 및 현 기능을 중심으로 설명한다. 또한, 번역 결과의 평가에 대해 설명하고 향후의 기능 향상 요구사항 및 예정사항을 제시한다.

1. 소개. 군 자동화 분야에 있어 정보처리체계(information management systems)는 군조직이 당면하고 있는 문제들에 대하여 적시적인 작전 및 관련정보를 해당 지휘관에게 실시간으로 제공해야 한다. 이는 다시 말하여 작전의 계획, 관리 및 통제에 대한 집중된 수단을 의미한다. 이러한 체계의 핵심은 각각의 사용자들로, 이들은 데이터베이스의 갱신, 정보자료의 처리, 분석업무의 수행 및 타 사용자와의 통신을 일상업무로 하고 있다.

현대 군 자동화 체계의 확장은 컴퓨터 관련기술의 통합에 대한 새로운 요구를 창출하고 있다. 특히, 컴퓨터 및 소프트웨어간의 상호운용성(interoperability)문제는 최신 체계의 개발에 있어 지속적인 관심분야가 되고 있다. 더우기 여러나라의 연합환경하에서는 서로 다른 언어를 통한 정보의 관리 문제가 상호운용성에 대한 또다른 차원의 문제 - 상호 다른 언어를 사용하는 사용자간의 문서 및 정보 교환이라는 -를 내포하고 있다. 국가간의 공동목적 달성을 위한 상호노력의 증대는 군 자동화체계(지휘통제체계 및 정보관리체계)에서의 양국어 관련 요소에 대한 요구사항에 관심을 기울이기 마련이며, 각 사용자들은 서로 다른 언어간의 의사소통을 쉽게 달성토록 요구하기 마련이다.

양국어를 이용하는 지휘통제체계에서는 요구가 있을 시 한 언어로 작성된 문서(source text)를 다른 언어(target text)로 번역할 필요가 있다. 이를 달성하는 방법으로는 첫째, 각 사용 참가자가 자신의 작성 문서를 직접 번역하는 방법이 있다. 또 다른 방법으로는 지정된 번역전문가에게 각 문서를 번역 요청하여 그 결과를 받는 수가 있다. 그러나, 첫번째 방법에서는 각 사용자가 양국언어를 능숙하게 구사할 수 있어야 한다는 단점이 있으며, 두번째 방법에서는 번역에 소요되는 시간 및 경비의 증가와 함께 긴급한 또는 중요

한 문서를 결정권자에게 적시에 제공하기 어렵다는 단점이 있다. 여기에서 나올 수 있는 대안으로는 기계번역을 이용함으로써 신속하고 개략적인 번역 결과를 제공토록 하여 사용자로 하여금 주어진 문제에 관한 정보의 필요 정도를 판단하게 하는 것이다. 만일 번역의 정확도가 필요한 문서라고 판단될 때에는 개략적인 번역문에 대하여 후편집(post-editing)을 통한 기계번역 과정을 적용시킬 수 있을 것이다.

본고에서는 양국언어를 사용하는 군 자동화체계를 위한 기능적인 요구사항을 언급하며, 이에 결들여 한영 양국어를 사용하고 있는 TACCIMS(Theater Automated Command and Control Information Management System)를 예로 제시한다.

2. TACCIMS 소개. TACCIMS는 평시, 훈련시, 위기시 및 전시 동안 한미연합사령관을 지원하기 위한 지휘통제 체계이다 [1]. TACCIMS의 양국가적 요소들은 한측 및 미측 TACCIMS 사용자들로 하여금 군 지휘 정보 자료들을 상호 공유하도록 하고 있다. 즉, 연합군 조직상에서 상호 정보를 교환하는 측면이 TACCIMS의 특징이다. 이러한 연합군의 작전 환경 지원은 세계적으로 다른 여러 연합군 작전시에도 응용될 수 있는 다양한 내용을 포함하고 있다.

TACCIMS는 ISO를 비롯한 각종 국제 규격을 사용하고 있다. 이러한 접근방법은 향후 예상되는 연동 및 확장성과 다양성을 보장하기 위한 것이다. 또한, TACCIMS 소프트웨어는 양국어 데이터베이스 관리체계, 양국어 전자우편, 양국어 업무용 및 지도 그래픽 등을 보유하고 있으며, UNIX SVR 3.2 운영체제 및 ASCII와 한국공업규격 (KSC) 5601 표준 코드 체계를 사용한다. 이러한 양국어간의 각종 표준규격을 포함한 통합체계는 상용 소프트웨어 및 하드웨어의 단순한 통합에 또 다른 차원의 고려사항을 제시하고 있다. TACCIMS의 중요한 요소의 하나는 이러한 한국어 및 영어간의 의사소통을 위한 기계번역 기능이라 할 수 있다.

현재 구축되어 있는 TACCIMS 구조상에서는 번역이 워크스테이션, 화일서버 및 중앙 HESS(Hangul English Support Sytem) 서버의 세가지 수준에서 처리된다. 각각의 차이점은 내장되어 있는 사전 및 번역 규칙들의 크기와 번역 소프트웨어의 수정 및 유지보수 등에 있다. 중앙 HESS 서버는 가장 최신의 버전을 유지하며, 전체 통신망을 통하여 모든 번역 소프트웨어의 수정에 대한 권한 및 책임을 지닌다.

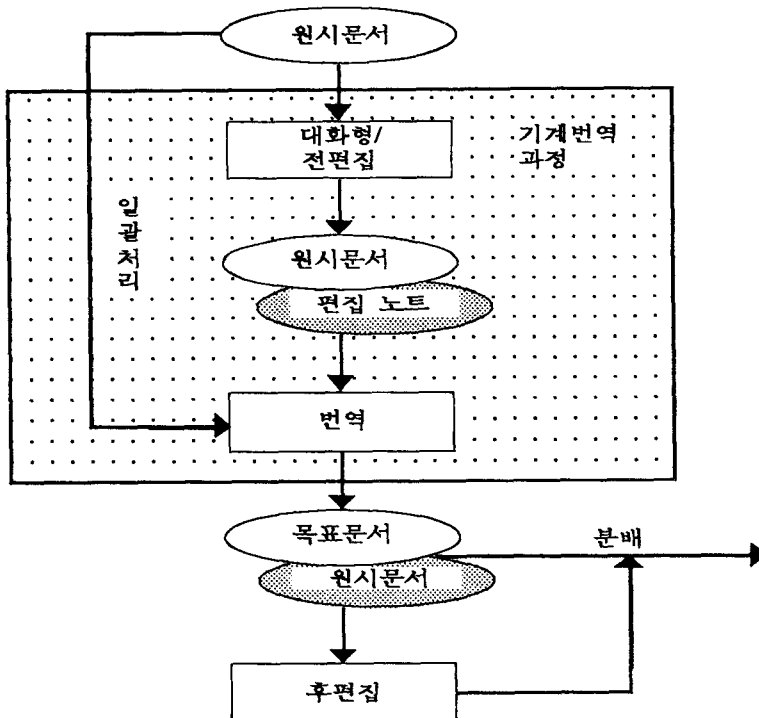
3. 기능적 요구사항. 기계번역 작업에서는 입력문서에 포함되어 있는 양식이나 두문 정보 등의 특정 형태를 그대로 유지시키면서 원시언어를 목표언어로 번역하여야 한다. 이 과정을 수행하기 위한 방법으로는 직접번역 및 간접번역 방식 등의 여러가지가 있다 [2]. 번역방식과는 관계없이 최초 사전편집(pre-editing) 과정에서 선택적인 후편집(post-editing) 과정에 이르기까지 기계번역이 거쳐야 할 처리 흐름은 일정하나, 번역 질의 요구 정도에 따라 각 세부과정이 적용되어야 할 수준도 다양하다.

<그림 1>은 기계번역 체계가 갖추어야 할 주요 운용 요소를 나타내고 있으며, 각 요소

는 다음에 개략적으로 설명되었다.

3.1. 원시문서(Source Text). 번역을 위한 원시문서는 여러 곳으로부터 제출된다. 군 지휘통제 체계의 경우 다음과 같은 정보처리 소프트웨어가 그 입력 수단이 될 것이다.

- 워드프로세서(word processor)
 - 보고서(reports)
 - 브리핑(briefings)
- 전자우편(electronic mail)
- 데이터베이스 질의(database query)
 - 데이터베이스 자료(data records)
 - 스프레드시트(spreadsheets)
- 업무용 및 지도 그래픽(business and map graphics)
- 군 전문양식(military messages)



<그림 1> 기계번역의 기본적 구성요소

3.2. 일괄처리(Batch Mode). 일괄처리시에는 대화형/전편집 과정을 생략하게 되며, 따라서 번역의 정확도가 저하된다. 이 정확도의 저하는 정확한 의미에 기초하기보다는 내장되어 있는 용어의 빈도에 기초한 단어 선택 및 구문 정보 추출을 수행하기 때문이며, 이에 대한 예제는 뒤에서 보이기로 한다.

3.3. 대화형/전편집(Interactive/Pre-editing). 대부분의 한측 및 미측 사용자들은 한 언어만을 알고 있으므로, 기계번역 과정에 참여시키기 위해서는 그들의 언어를 이용한 참여 수단을 제공해야 한다. 즉, 이러한 사용자의 참여는 원문이 작성된 곳에서 일어나야 하며, 원문에 대한 기계번역시 발견된 문제점이나 애매성에 대하여 해당 사용자에게 그 나라 말로 질의하고 응답을 유도하여야 한다. 또한, 원 작성자가 직접 번역에 참여함으로써 번역과정상 발생할 수 있는 문서상의 오류에 대한 책임문제를 해결할 수 있다. 한편, 번역 자체가 잘못되는 경우가 발생할 수 있으므로, 이의 보완을 위하여 번역문과 함께 원문을 해당 수신지로 전송하는 방법을 취하는 것이 좋다.

대화형/전편집기는 작성자에 의해 명확히 되어야 할 단어, 어절, 또는 약어 등을 확인하며, 품사 및 특정 단어가 의미하는 바 등에 대한 모호성을 해결한다. 예를 들어, 한 단어가 여러가지 의미를 가지고 있는 경우("spring"이라는 단어는 도약, 용수철, 봄, 샘, 탄력의 다섯가지 의미를 가지고 있음) 이 단어가 쓰이는 용도에 따라 여러가지 목표언어로 변환될 수 있으므로, 이 편집기를 통하여 사용자가 문맥에 맞는 용어 선택을 할 수 있도록 한다. 한편, 번역 알고리즘 자체는 여러가지 대안 중 하나를 선택하는 정확도와는 무관하기 마련이며, 각 의미의 통계적 분포에 의존하게 된다.

이 편집기는 또한 구문적 모호성을 해결한다. 이 경우, 가능한 구문적 해석 대안들이 각각 제시된 후 사용자에 의해 최적의 해석안이 선택된다. 이 편집기는 번역 핵심부분과 연계되어 번역 과정을 지원하기 위한 도구로 사용된다.

3.4. 번역(Translation). 이 처리과정에서는 원문을 분석하여 중간표현을 생성하고 이를 목표언어로 변환하는 기능을 한다. 이러한 과정에는 여러가지 방식이 있으나, 과정 전체를 결정짓는 것은 사전의 크기와 적용되는 번역 규칙의 완벽성이다. 입력의 종류에 따라 적용 방법을 달리 할 수 있는 번역 알고리즘은 유용할 것이다. 예를 들어 번역 대상이 스프레드 시트의 문구일 경우 번역은 단순한 단어 변환의 수준으로 충분하다. 또한 브리핑 슬라이드의 내용을 번역하고자 할 때 완전한 문장을 번역하려는 시도는 오히려 불분명한 해석과 번역의 오류만 초래할 뿐이며, 단어 변환의 경우가 훨씬 정확하고 효과적이라는 것을 알 수 있다. 따라서, 각 입력 대상에 대한 감지와 이에 따른 적절한 번역방법을 취할 수 있는 번역 알고리즘이 필요할 것이다.

TACCIMS에서의 HESS 번역 처리는 Lexical Functional Grammar (LFG) [3, 4]를 사용하고 있다. LFG는 각 단어의 의미에 중요성을 부여한 문법으로서[5], 각 문장의 표층요소 구조와 해당 의미요소간의 연계과정을 표현하는 방식이다. 이러한 문법적 관계는 f-구조(f-structures)로 표현된다.

다음 문장에 대한 f-구조의 예를 <그림 2>에 도시하였다.

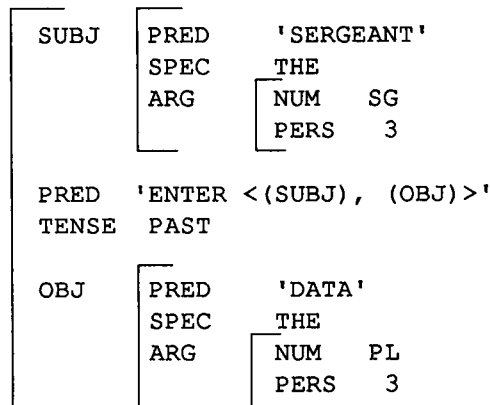
(1.a) "The sergeant entered the data."

(1.b) "그 하사관이 그 자료를 입력했다."

3.5. 번역의 예(Translation Example). 다음 문장은 대화형/전편집기의 유용성을 보여 주기 위한 예이다.

(2.a) "The data has been entered into the computer in the office."

(2.b) "그 자료가 그 사무실에서 그 컴퓨터 안으로 들어가고 있었다."



<그림 2> 문장 (1.a) 에 대한 f-구조

위의 문장은 HESS에서 일괄처리에 의해 수행된 번역의 결과이다. 여기에는 두 가지 오류가 보인다. 첫번째 오류는 조사의 선택 오류로서, "에서"에 대한 올바른 번역은 "에 있는"이 되어야 한다. 이 번역 오류는 변환상의 오류로 볼 수 있다. 한편, 번역된 단어 "들어감" (방에 들어간다는 의미)의 선택은 문맥에 맞는 단어의 선택이라기보다는 사전에 등록된 통계적 빈도에 기초한 선택이다. 이러한 오류는 대화형/전편집기에 의해 해결될 수 있다. 각 대안 중 사용자에게 의해 선택된 최적 대안이 사용자 인터페이스를 통해 선택될 수 있으며, 이는 변환과정에서 올바른 의미를 가진 목표언어 "입력되"로 생성되는 결과를 가져오게 된다.

3.6. 목표문서(Target Text). 이는 번역과정에서 생산된 초기 번역결과물로서, 번역결과의 품질은 번역 알고리즘 자체의 성능 이외에도 입력 문장의 복잡성, 대화형/전편집 과정의 적용 수준, 그리고 사용자가 요구하는 번역물의 정확도에 따라 그 평가가 달라질 수 있다. 번역물에 대한 보완이 필요한 경우에는 후편집을 통하여 보다 정확한 번역을 할 수 있다.

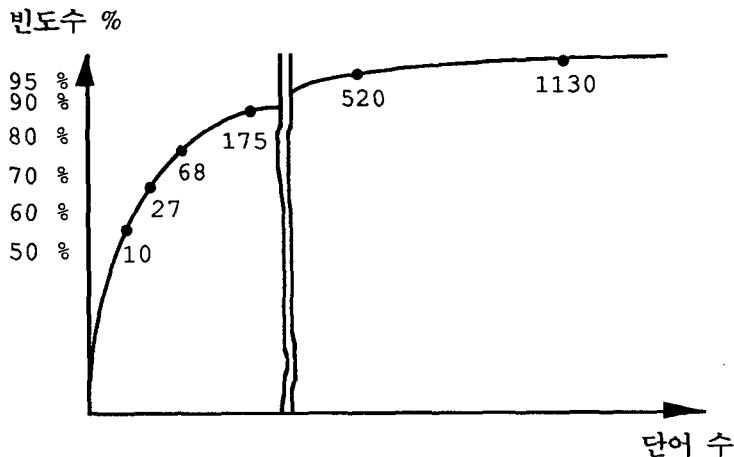
3.7. 후편집(Post-editing). 원시언어, 목표언어 및 번역 대상분야에 정통한 번역전문가는 초기 번역물에 대한 문법 교정이나 어색한 문구의 첨삭, 올바른 의미의 전달 및 문장 전체의 통일성 등을 제고시킬 수 있다. 즉, 후편집은 문서의 전반적인 이해가능성을 향상시키며, 번역의 정확성을 보장해야 할 주요 문서에 대해서 꼭 필요한 과정이다. HESS에서는 후

편집이 중앙 서버에 위치한 "Linguist Mode"에서 수행되며, 여기에서는 원시문서와 목표 문서를 동시에 보면서 워드프로세서의 기능을 이용하여 목표문서에 대한 수정 보완을 수행하게 된다.

4. 사전

4.1. 사전 항목(Dictionary Entries). 사전에는 입력문서에 나타날 가능성이 있는 모든 단어, 구문, 약어 및 속어들이 등록되어 있어야 한다. 한 언어의 번역에 효과적이고 유용한 사전 항목을 수집하기 위해서는 여러가지 분야에 걸쳐 적어도 백만 단어 이상의 표본 문치를 분석하는 것이 필요하다. 세계적으로 이에 대한 노력이 진행되어 왔는데, 그중에는 Thronkide-Lodge Word Frequency Count (1944), Brown Corpus (1967), Lancaster Oslo/Bergen (LOB) Corpus (1978), Birmingham Collection (1987), 문교부 어휘문치(Korean Ministry of Education (MOE) Corpus) (1955, 1956), 그리고 연세대학에서 수행한 어휘문치(Korean Corpus by Yonsei university) (1988) 등이 있다 [6]. HESS에서는 일반용어사전, 군 약어사전 및 연합사령부 번역부서에서 수행한 분석결과 등을 토대로 70,000 여 단어를 수용하고 있다.

한편, 단어의 사용빈도 측면에서 보면 약 5,000 개의 단어가 전체 사용빈도의 90 % 이상을 점유하는 것을 알 수 있다. <그림 3>은 한국어에 대해 이러한 관계를 보여주는 것으로, 2백만 단어 이상의 어휘문치를 분석한 문교부 어휘문치[7]에 대해 조사한 결과이다.



<그림 3> 사전 크기와 처리범위와의 관계

사전은 원문 작성과정과 연계되어야 한다. 이는 번역에 앞서 사전에 등록되지 않은 문서상의 단어를 미리 확인할 수 있도록 한다. 또한, 단어의 사용 패턴을 추적해감으로써 분산된 구조에서 각 워크스테이션에 상주할 사전의 크기가 제한되어 있을 때 가장 발생빈도

가 높은 단어를 우선적으로 상주시킬 수 있도록 한다.

4.2. 사전의 유지관리(Dictionary Maintenance). 사전의 등록/편집 도구는 번역체계를 이용하기 위해서는 관리자가 필수적으로 운용해야 할 소프트웨어이다. 사전은 필연적으로 그 추가 및 수정이 일어나기 마련이므로, 사전의 유지 및 관리는 중요한 부분 중의 하나이다. 사전의 유지를 위해서는 양국언어에 정통한 언어학자가 필요하다. HESS에서의 사전 유지관리 유틸리티(Dictionary Maintenance Utility (DMU))는 풀다운, 팝업 메뉴방식을 이용하여 사전 항목의 추가, 변경, 확인 및 사전의 추출, 혼합 등의 기능을 수행한다. HESS에서는 크게 두 종류의 항목 구분이 있는데, 특정항목(reserved words)은 개발언어학자에 의해 갱신될 수 있으며, 일반항목(open class words)은 HESS 관리자에 의해 유지관리되도록 되어 있다.

또한, HESS 관리자가 사전을 손쉽게 관리할 수 있도록 번역기에서는 다음 세가지 기록 화일을 유지한다: 1) 대화내용 기록 화일(interaction log file)은 대화형/전편집 과정에서 사용자와 질의/응답을 거친 일련의 내용을 원문과 함께 기록하고, 2) 통계 기록 화일(statistics log file)은 번역에 사용된 각 단어 및 번역규칙의 통계적 정보를 유지하며, 3) 미등록어 화일(unfound words file)은 번역 알고리즘이 해결 불가능하였던 미등록어들을 해당 원문과 함께 유지한다. 이러한 기록 화일들은 모호성을 가진 단어에 대한 대안 제시 메뉴에 관한 추가 정보를 제공할 뿐만아니라, HESS 관리자와 개발 언어학자로 하여금 사전 항목 및 번역규칙의 보완/향상을 지원함으로써, 궁극적으로 전체 체계의 성능 향상을 지원하게 된다.

5. 성능 평가

5.1. 번역속도(Translation Speed). 단어의 단순한 나열이나 복잡한 문장과 같은 경우 사람이 이해하기 힘들뿐만 아니라 컴퓨터로 처리하기도 힘들기 마련이다. 기계번역을 위한 입력문서의 작성시에는 되도록이면 간단명료한 문장을 위주로 작성한다는 생각을 가지고 있어야 할 것이다. 문장의 길이는 번역기가 탐색해 나가야 할 분석 갈래를 보다 많이 발생시키게 되며, 길이가 길수록 문장의 해석은 어려워지게 된다. 그 상관관계는 기하급수적이며, 예를 들어, 짧은 길이의 문장은 몇초 내에 번역될 수 있는 반면, 긴 문장은 몇분이 걸릴 수도 있는 것이다. HESS에서 사용된 평가기준은 6개 단어로 이루어진 문장에 대한 번역속도로 6초를 요구하고 있다.

5.2. 번역의 정확도(Translation Accuracy). HESS에서의 번역의 정확도는 편집비율(edit ratio (ER))에 의해 평가하고 있다. 편집비율이란 초기 번역물에 대한 후편집의 비율을 말한 것으로서 [8], 일반적으로 20 % 내외의 결과는 매우 좋은 것으로 이해된다. 물론,

규정의 번역과 같이 매우 정확한 번역을 요하는 경우에는 해당분야의 번역전문가가 후편집을 수행하여야 할 것이며, 이와 달리 문서의 개요만을 사용자가 이해하고자 할 때에는 기계번역을 통한 최초의 번역물 자체로도 충분할 것이다. 예를 들어 연료공급건에 관한 문서는 참모 기능 축소에 관심을 두고 있는 결정권자에게 긴급한 필요문서가 아니며, 이 경우 그 문서가 의사결정과 직접 관계되지 않는다는 내용만을 파악함으로써 더이상의 정확도를 필요로 하는 후편집을 거칠 필요가 없는 것이다. 따라서, 아주 정확한 번역보다는 적절한 정확도를 가지면서 신속하게 번역을 수행할 수 있는 체계의 필요성은 그 의미를 가진다고 할 것이다.

기계번역 체계에 대한 성능을 평가하기 위해서는 ER을 얼마만큼 객관적으로 적용하는가가 중요하다. ER의 예를 들면 다음과 같다.

$$(3) ER = \frac{\sum_{i=1}^n \mu(\Gamma_i) * \rho(\Gamma_i)}{\sum_{i=1}^n \rho(\Gamma_i)}$$

ER: 편집 빈도(Edit Ratio), $0 \leq ER \leq 1$

$\Gamma = \{\Gamma_1, \Gamma_2, \dots, \Gamma_n \mid \Gamma_i = i\text{번째 문장}\}$

$\mu(\Gamma_i)$: Γ_i 의 정확한 의미와 목표문장의 의미간의 차이

$$\mu(\Gamma_i) = \begin{pmatrix} 0.0: \text{정확} \\ 0.2: \text{단어 삭제} \\ 0.4: \text{단어 추가} \\ 0.6: \text{단어 변환} \\ 0.8: \text{단어 순서} \\ 1.0: \text{수락불가능} \end{pmatrix} + \{\text{대화 빈도} * \text{각 대화의 영향치}\}$$

$$0 \leq \mu(\Gamma_i) \leq 1$$

$\rho(\Gamma_i)$: Γ_i 의 중요치. 각 Γ_i 에 대한 $\rho(\Gamma_i)$ 는 시험 문장군의 분석에서 Γ_i 의 빈도수로 주어진다. $\rho(\Gamma_i)$ 는 3차원 행렬로 나타나는데 각 차원은 다음과 같다.

- 문장 길이
- 문장 구조
- 서술 형식

6. 평가를 위한 자료. 성능을 평가한다는 것은 번역된 결과가 얼마나 유용한가를 파악하기 위한 것이다. 이를 위해 개발과정에서는 신중하게 선정된 문장들을 이용하여 개발시스템의 각 부분들을 체계적으로 시험할 수 있도록 하여야 할 것이다. 이러한 시험 문장군 및 평가기준을 위한 연구가 지속적으로 진행되고 있는데, 그 대표적인 것들로는 ALPAC 보고서에 기술된 평가내용 (1965), 유럽공동체(EEC)에서 Systran 프로젝트에 대해 수행한 평가 (1978), 그리고 일본에서 Mu 프로젝트에 대해 수행한 평가(1985) 등을 들 수 있다[9].

그러나, 한 체계에 대해 성공적으로 시험할 수 있는 평가 문장군 및 기준이 다른 체계에 똑같이 성공적으로 적용되지는 않는다 [10]. TACCIMS에서는 번역체계의 성능을 평가하기 위하여 간단한 문법에 의한 단순, 평서문을 주어-동사-목적어(subject-verb-object (SVO)) 형태로 작성함으로써 시험용 문장군을 만들었다. 예를 들어, 주어는 "The soldier," 또는 "The American soldier," 또는 "The American soldier near the computer" 등과 같이 복잡도가 증가하는 형태로 생성되었으며, 각각의 성분에 대한 복잡도를 같은 방식으로 증가시키면서 이의 결합을 변화시킴에 의해 시험 문장군을 작성하였고, 이를 토대로 기계번역 알고리즘의 속도 및 정확도에 대한 변화를 추적하여, 각 결합 부분에 대한 확인과 향상 가능성을 검토할 수 있도록 하였다.

이러한 방식으로 700 여 문장이 생성되었으며, 각 문장에 대하여 번역의 속도와 ER이 평가되었다. 또한, 이 평가 문장군은 번역체계의 향후 확장성(extensibility)을 검사하기 위한 벤치마크 시험의 베이스라인 역할을 수행한다.

7. 향후 고려사항. 실시간 기계번역 체계의 향상은 번역의 속도와 정확성에 초점이 주어질 것이다. 이중 속도의 향상은 보다 빠른 컴퓨터의 사용으로 이루어질 수 있으며, 정확도의 점진적인 향상을 이루기 위하여 전산언어학 분야에서 계속적인 연구가 진행되고 있다. 그러나, 한가지 짚고 넘어가야 할 사항은 자연언어 처리분야가 결코 단순한 분야가 아니라는 사실이다. 따라서, 언어 자체에 내재되어 있는 여러가지 난점들을 극복해 나가면서 체계를 개발하고 시험하기 위해서는 많은 시간이 소요될 것이라는 점이다. 말하자면, 몇주간의 단기간동안 사람들이 외국어를 배울 수 없는 것과 마찬가지로 언어에 내재된 복잡성을 파악하기 위해 걸리는 시간 또한 이에 버금가거나 이를 능가하는 것이다. 또다른 측면으로, 번역체계의 복잡성은 그 체계가 처리할 수 있는 언어적 능력에 기하급수적으로 비례하게 된다. 즉, 자연언어 처리분야의 선형적 발전을 위해서는 컴퓨터와 소프트웨어 능력의 기하급수적 발전이 뒤따라야 한다는 것이다 [5].

정확도를 향상시키기 위해 단기적으로 예측 가능한 분야로는 구문지식을 갖춘 워드프로세서의 개발이라고 할 수 있다. 이는 온라인 입력상에서 문장을 미리 분석해감으로써 차후 번역에 어려운 문장의 확인 및 해당 문장의 보완을 유도할 수 있으며, 결과적으로 번역시에 대화형/전편집 과정의 부담을 줄이게 될 것이다. 또다른 향후 연구분야로는 자연언어의 다양한 측면을 다룰 수 있는 번역체계 시험 방법론과 시험 문장군의 개발을 들 수 있겠다. 단문과 평서문 외의 여러 문장 형태를 다룰 수 있는 시험 문장군의 개발과, 적절한 어휘목록에 근거한 ER의 어휘별 중요도를 찾아내는 일이 필요할 것이다.

TACCIMS는 앞으로 UNIX SVR 4로의 갱신, 하드웨어의 갱신, 사전 갱신, 개발 도구를 사용한 번역 엔진의 향상 등을 통하여 단문은 물론, 중문, 복문까지도 번역가능한 시스템을 구축할 계획이다.

8. 결론. 군 자동화체계에서의 기계번역이 지니는 의미는 장래 구축될 체계에 여러가지 도움을 줄 것이다. 본문에 예시된 한영번역 지원체계는 장래 구축될 양국어틀 이용한 지휘 통제체계에 대한 프로토타입이라 할 수 있다. 이 분야에서 중요한 것은 현재 기계번역 기술의 수준에 근거하여, 개발될 특정 번역체계에 요구되는 복잡성과 정확성 등에 대한 현실적인 예측이 필요하다는 점이다. 어쨌든, 군자동화체계에 이러한 분야의 기술을 도입한다는 것 자체가 유사한 체계에 대한 흥미로운 발전의 일보라고 할 수 있을 것이다.

참고문헌

- [1] Project Manager TACCIMS, DAAB07-88-R-M023: TACCIMS Request for Proposal (RFP), AS-PEN-TA, December 29, 1987.
- [2] Pentheroudakis, J. E., You Can Get There From Here: Design and Implementation Issues in Machine Translation Systems, Proceedings of the Topical Meeting on Advances in Human Factors Research on Man/Computer Interactions: Nuclear and Beyond, American Nuclear Society, June 1990.
- [3] Bresnan, J. (Ed.), The Mental Representation of Grammatical Relations, Cambridge, Mass: MIT Press, 1982.
- [4] Frey, W., Noun Phrases in Lexical Functional Grammar, International Workshop on Natural Language Understanding and Logic Programming, Elsevier Science Publisher B.V., 1985.
- [5] Johnson, T., Natural Language Computing: the Commercial Applications, Ovum Ltd., 1985.
- [6] 정찬섭, 한국어 어휘문치의 표본 선정 기준, 한글 및 한국어 정보처리 학술 발표 논문집, 한국인지과학회, 1989.10.
- [7] 문교부, 우리말 말수 사용의 잣기 조사, 태성출판사, 1956.
- [8] Guida, G. and Mauri, G., Evaluation of Natural Language Processing Systems: Issues and Approaches, Proceedings of the IEEE, 74(7), July 1986.
- [9] Nagao, M., Evaluation of the Quality of Machine-Translated Sentences and the Control of Language, Information Processing, 26(10), Information Processing Society of Japan, October 1985.
- [10] Flickinger, D., Nerbonne, J., Sag, I., and Wasow, T., Toward Evaluation of NLP Systems, Hewlett-Packard Laboratories, Palo Alto, CA, June 1987.