

사례 발표

번역메모리(Translation Memory) 기반 자동번역 및 번역지원(Computer Aided Translation)시스템

강명주\*

목 차

- 1. 서 론
- 2. 사례 연구
- 3. 시스템 활용
- 4. 결 론

1. 서 론

자동 번역에 대한 연구는 1940년대부터 시작되었고, 이 때의 연구는 일반 사람들에게는 거의 알려져 있지 않았으며, 본격적인 연구는 컴퓨터가 나오게 되는 1950년대부터 시작되었다. 이들이 생각했던 자동 번역은 두 언어간의 단어들을 교환하여 대치시키고 단어 쌍으로 이루어진 전자사전을 컴퓨터에 입력하여 번역하는 형태였다. 이 시기에 사용한 언어 쌍은 영어-러시아어였으며 주로 군사적인 목적으로 미국과 구 소련에서 연구되었다[1, 2].

1950년과 60년대의 연구는 주로 우주항공과 군사 분야에서 이루어졌다. 1954년에 IBM과 Georgetown 대학이 핵물리학에 관한 러시아어 문서를 영어로 번역하는 시스템을 공동으로 개발하였다. 이 시기에 직접 번역 방법(Direct Translation Method)과 중간언어 번역 방법(Interlingual Translation Method)이 연구되었다[3]. 그러나 느린 발전 속도와 낮은 번역 결과는 일반인은 물론 자동 번역을 연구하는 사람들까지도 실망시켰다. 특히 ALPAC(Automatic Language Processing

Advisory Committee) 보고서에서 자동 번역은 실질적인 면에서 실현성이 없다라는 결론을 내린 후 컴퓨터가 번역가를 지원할 수 있는 시스템, CAT(Computer Aided Translation)에 대한 필요성이 제기되었다[4].

1980년대에 ALPAC 보고서 이후 침체되었던 연구는 인공 지능 기반 방법(AI-based Method), 지식 기반 방법(Knowledge-based Method) 또는 규칙 기반 방법(Rule-based Method)의 제시와 컴퓨터의 성능 발달에 힘입어 전세계적으로 활발하게 진행되었다[7]. 1990년대에는 규칙 기반의 방법을 극복하기 위해 IBM Candide 프로젝트에서 처음 소개된 코퍼스 기반 방법(Corpus-based Method), 예제 기반 방법(Example-based Method), 통계 기반 방법(Statistics-based Method) 등이 연구되었다[8][9][10][11]. 통계 기반 방법은 사람들이 실제로 사용하는 많은 언어 데이터로부터 확률 정보 및 통계 정보를 추출하고, 이를 이용하여 여러 가지 언어 현상을 규정하여 자동 번역에 이용하고자 하는 방법이다. 그러나 통계 기반 방법은 실제로 사용하는 대용량 코퍼스를 모으기가 어렵고, 언어적 활용이 많아 다양한 문장 구성의 연구가 필요하

\* 클릭큐 부설 연구소 연구소장

며, 대용량 데이터를 처리할 수 있는 컴퓨터가 필요하다라는 단점이 있다. 한편 이 시기에 개인용 컴퓨터가 보편화되고 인터넷이 활성화되어 월드 와이드 웹(WWW, World Wide Web)이 등장함에 따라 일반 사람이나 번역가들도 쉽게 컴퓨터를 사용할 수 있는 환경이 되었다. 이런 시대적인 환경에 따라 번역가를 위한 도구인 Trados Translator Workbench, IBM TranslationManager/2, STAR Transit, Eurolang Optimizer와 같은 CAT 소프트웨어들이 시장에 출시되었고 번역메모리(TM, Translation Memory), 용어 관리 도구, 정렬 도구, 워드 프로세서, 탁상용 출판 소프트웨어 등 번역 작업을 지원하는 도구들이 등장하였다. 1990년대 후반에는 기업들을 중심으로 규칙 기반 방법과 통계 기반 방법을 혼합하는 방법도 제시되었고, 자동 번역과 번역 메모리를 혼합한 시스템도 등장하였다.

2000년도에 들어 한국에서도 ETRI와 민간 기업에서 번역메모리 엔진에 대한 연구가 진행되었다 [3]. 특히, 클릭큐에서는 번역메모리 기반 자동번역과 CAT 시스템 개발에 착수하여 TraNew(트래뉴), TraVita(트라비타)와 같은 제품이 국내 최초로 출시되었다. 본 원고에서는 TM 기반 자동번역과 CAT 제품이 산업 현장에서 사용된 한 사례를 집중적으로 설명하고자 한다.

## 2. 사례 연구

본 사례는 2003월 6월부터 진행된 대한민국 공군의 “기술도서 기계번역체계 개발” 프로젝트다. 공군 기술도서 기계번역체계는 기존 기술문서의 관리와 번역지원의 지연과 업무량의 과다로 인한 업무처리의 효율성을 제고하기 위하여 정비 기술문서의 용어의 통일과 전문 번역 업무 부서의 업무지원, 업무시스템 통합을 통한 번역업무의 효율성 극대화과 기술도서의 디지털화하기 위한 지원 소프트웨어 및 번역지원체계를 구축하는 것이다. 이

러한 목적 달성에 기여하기 위하여 기술도서 기계번역체계 구축 사업을 수행함으로써 보다 효과적이인 영한 번역 시스템을 개발하여 활용하는 것이다.

### 2.1 시스템 개발의 필요성

대한민국 공군은 여러 종류의 전투기 및 무기체제를 보유하고 있으며, 이런 무기체제를 정비하는데 필요한 기술지시(T.O: Technical Order)는 한글이 아닌 영문으로 편집되어 있다. 현재 공군은 각 전투기의 기종별 기술지시에 대하여 전문 번역 업무 부서를 운영 중에 있으나 번역 처리 능력과 번역 요구량간의 업무 처리에 많은 애로점이 있다. 항공기 정비 및 기타 무기체제의 기술문서인 경우, 반복되는 표현들이 많고 문장 구조가 규칙성에 입각하여 작성되었기 때문에 자연어 처리 기술과 번역 시스템 및 번역메모리 시스템을 이용하여 기술 문서 번역에 사용할 경우 차후에는 번역 수요를 충족시킬 수 있을 것으로 판단된다. 항공기 정비 및 기타 기술문서에 대하여 기 번역된 번역문서와 용어를 이용하여 자동번역 및 번역메모리를 이용하여 예하 운영기지에 한글 정비 매뉴얼 및 기술문서를 제공함으로써 항공기 정비를 할 수 있고, 항공기에 대한 기술 습득을 원활히 함으로써 공군의 정비 및 항공기 관련기술을 향상시켜 공군을 지식화하고 공군 전체의 비용과 시간을 절감하여 전력을 극대화할 필요성이 있다.

### 2.2 개발 목표

시스템의 개발 목표는 자동번역과 번역메모리 검색 기술을 통합한 번역시스템과 제반 용어집과 TM을 관리, 검색할 수 있는 시스템 및 두 시스템간의 상호 연동하는 시스템이며, ‘사람을 대신하는 번역 시스템’이 아니라 ‘사람의 번역을 지원하는 시스템’을 추구한다. 기계번역은 완벽한 직역을 추구하고, “사람의 번역”은 완벽한 의역을 추구하고, 개

발 목표가 “사람이 번역하는 생산성 향상”이라는 전제에서, 시스템적으로 사람의 번역 결과의 재사용 및 용어의 통일, 자동번역의 번역률 향상을 통하여 최대한 사람의 노력을 절감하는 번역 시스템을 개발함을 목표로 한다.

### 2.3 개발 범위

기술도서 기계번역체계에서 개발된 시스템의 범위는 (그림 1)과 같이 번역메모리를 검색하여 번역할 수 있는 TM 기반 자동번역 시스템, 용어와 번역메모리 구축을 원활하게 할 수 있는 CAT 시스템, 기술 문서가 종이로 되어 있는 경우 전자 문서로 변경할 수 있는 문자인식 시스템, TO와 TTO (한글기술지시, Translated Technical Order) 기술 문서를 번역기술도서 관리 시스템이며, 개발되는 시스템연계, 응용소프트웨어, 지원 부문, 정보인프라 등이 포함된다.



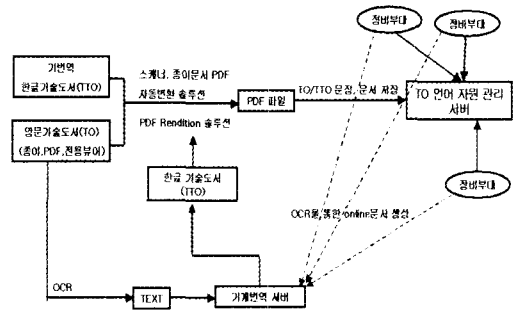
(그림 1) 기계번역 체계 개발 범위

## 3. 시스템 활용

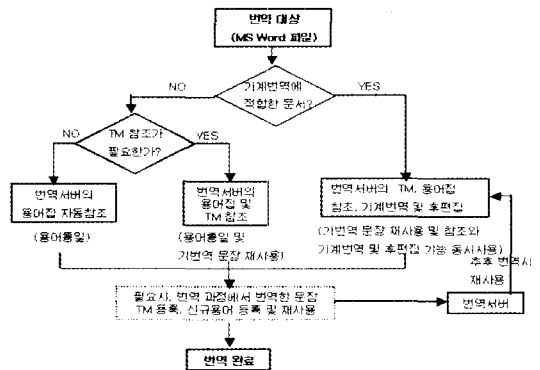
### 3.1 번역 공정

시스템의 전반적인 번역 공정 구성은 (그림 2)와 같다. 정비부대는 공군 내에서 전투기 및 무기체제를 정비하는 부서이며, 번역을 담당한 번역부대는

기계번역 서버와 TO/TTO 문장의 쌍인 번역메모리와 문서를 관리하는 TO 언어 자원 관리 서버를 이용하여 영문 기술문서를 한글 기술문서로 번역한다. 번역 부서는 모든 번역된 문서의 TO/TTO 번역메모리가 TO 언어자원 관리 서버에 저장되어 있기 때문에 번역메모리와 유사한 다른 기술 문서를 번역할 때 재번역 없이 TM 기반 자동번역을 이용하여 번역할 수 있다.



(그림 2) 번역 공정 구성도



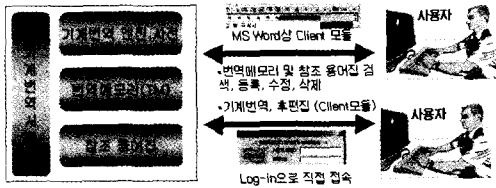
(그림 3) 시스템을 이용한 번역 공정

사용자가 기술문서인 MS-Word 문서를 번역하고자 할 때 모든 문서를 자동번역 시스템을 이용하는 것이 아니라 (그림 3)과 같은 번역 공정에 따라 번역한다. 우선 번역할 문서가 기계번역에 적합한 문서인지를 판단한다. 이 경우 용어와 문장의 패턴이 기계번역에 적합한지를 판단하여 만약 기계번역에 적합한 경우 그 분야에 맞는 번역메모리와 용어집을 선택하여 자동번역을 이용한다. 그렇지 않

은 경우는 기존의 번역메모리를 참조하고 용어집을 이용하여 사람이 직접 번역한다. 자동번역 혹은 사람이 번역하든지 모든 문장은 TO 언어자원 관리 서버에 TO/TTO 문장 번역메모리를 저장하여 다음에 재활용한다.

### 3.2 기계번역 서버

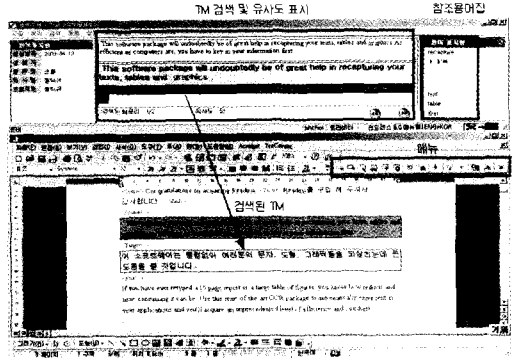
기계번역은 (그림 4)와 같이 서버에서 수행하지만 사용자가 이용하는 것은 클라이언트 모듈을 이용하게 된다. 기계번역 서버를 이용하여 번역하는 사용자는 MS-Word 모듈 및 클라이언트 모듈을 이용하여 서버에 로그인하여 사용자의 종류에 따라 번역메모리/용어집을 검색, 등록, 수정, 삭제하고 후편집을 이용하여 기계번역을 할 수 있다.



(그림 4) 기계번역 서버/클라이언트 시스템

### 3.3 TM 기반 CAT

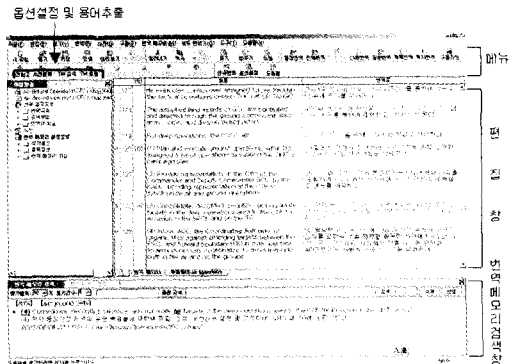
TM 기반 CAT 시스템은 (그림 4)에서 표시한 MS-Word의 클라이언트 모듈이다. 이 모듈의 프로그램은 (그림 5)에 있으며, 사용자는 MS-Word와 연동이 되는 클라이언트를 이용하여 번역 시스템을 이용할 수 있다. 이 프로그램의 기본적인 UI는 MS-Word에 있으며 Trados Workbench와 같은 형식으로 구현된다. Trados는 클라이언트에서만 수행이 가능하지만 본 시스템은 완벽한 서버/클라이언트 환경에서 수행되기 때문에 모든 사용자와 실시간으로 용어, 번역메모리, 문서와 같은 번역에 필요한 언어자원을 공유할 수 있다.



(그림 5) TM 기반 CAT 시스템

### 3.4 TM 기반 자동번역 시스템

TM 기반 자동번역 시스템은 (그림 4)에서 표시한 클라이언트 모듈이다. 이 모듈의 프로그램은 (그림 6)에 표시하고 있다. 사용자가 번역할 문서에서 문장 단위로 정렬하고 각 문장이 TM과 검색하여 일치 혹은 유사한 문장이 있으면 번역메모리 창에 내용을 표시하여 번역하고, 만약 유사한 문장이 없으면 기계번역을 수행하여 번역한다.



(그림 6) TM 기반 자동번역 시스템

## 4. 결론

지금까지 TM 기반 자동 번역과 CAT의 사례인 대한민국 공군의 “기술도서 기계번역체제 개발” 프로젝트에 대한 사례를 살펴보았다. 자동 번역은 아직까지 사람이 번역한 것보다는 질적인 면에서

미흡하다. 지금까지 이런 차이를 극복하기 위해 연구된 많은 방법들 있었으며, 완전 자동 번역이 아닌 번역가의 번역 작업을 지원하는 도구인 CAT 시스템이 어떻게 이용되고 있는지 살펴보았다.

앞으로 컴퓨터가 인터넷과 웹 환경에서 사용됨에 따라 콘텐츠와 지식의 공유가 더욱 활성화되고 그 지식에는 영어, 일어, 중국어 등 외국어로 된 지식들이 더욱 많아져 실시간 자동 번역에 대한 요구가 증가될 것으로 예상된다. 기업이 자사 제품을 여러 나라에 출시하기 위한 방법론이 L10N에서 G11N으로 변화함에 따라 기업의 제품이 출시되면서 여러 나라 언어로 된 제품과 매뉴얼 및 기술 문서들이 동시에 나와야 하는 환경으로 변하고 있다. 이런 변화에 따라 자동 번역 시장에서 질 높은 자동 번역과 CAT 도구에 대한 수요는 기업체 및 관공서 등에서 더욱 절실해지고 있다. 자동 번역의 질을 높이고 매체의 다양성을 추구하기 위해 대학, 연구소 그리고 기업간의 공동 연구도 활발히 이루어지리라라고 전망된다.

### 참고문헌

- [1] Barr, A., Feigenbaum, E.A., The Handbook of Artificial Intelligence, Vol. 1, William Kaufmann Inc., 1981.
- [2] John Hutchins, "Retrospect and Prospect in Computer-based Translation.", MT Summit, 1999.
- [3] 박상규, 강명주 외 12인, "동양 4국 언어에 대한 TM 기반 통합 CAT 시스템에 관한 연구", 연구개발결과보고서, 2003.
- [4] Languages and Machines, Computers in Translation and Linguistics, ALPAC-Report(Automatic Language Processing Advisory Committee Report), Washington, D.C., 1966.
- [5] Bostad, D., "Machine translation in the USAF.", Terminologie et Traduction no.1, pp. 68-72, 1986.
- [6] Pigott, I.M., "Systran machine translation at the EC Commission : present status and history.", Luxembourg CEC, January 1988.
- [7] Nirenburg, S. et al. "Machine translation: a knowledge-based approach." San Mateo, Ca., Morgan Kaufmann, 1992.
- [8] Rimon, M. et al., "Advances in machine translation research in IBM.", MT Summit 3, pp. 11-18, 1991.
- [9] Church, K.W., Miercer, R. L., "Introduction to the Special Issue on Computational Linguistics Using Large Corpora.", Computational Linguistics, Vol. 19, No. 1, pp. 1-24, 1993.
- [10] Sumita, E. et al., "Translating with examples: a new approach to machine translation.", TMI-90, pp. 203-212, 1990.
- [11] Brown, P. et al., "A statistical approach to language translation.", Computational Linguistics 16, pp. 79-85, 1990.
- [12] Hutchins, W.J., "Machine translation: past, present, future.", Chichester (UK): Ellis Horwood, 1986.
- [13] Harris, Z.S. 1954. "Transfer grammar.", International Journal of American Linguistics 20, pp. 259-270, 1954.
- [14] Bourbeau, L. et al., "Bilingual generation of weather forecasts in an operations environment.", COLING 90 (1), pp. 90-92, 1990.
- [15] Isabelle, P. and Bourbeau, L., "TAUM-AVIATION: its technical features and

- some experimental results.", Computational Linguistics 11(1), pp. 18-27, 1985.
- [16] Kaplan, R.M. and Bresnan, J., "Lexical-functional grammar: a formal system for grammatical representations.", Bresnan, J. (ed.), The mental representation of grammatical relations, Cambridge: MIT Press, 1983.
- [17] Boitet, C., "Current state and future outlook of the research at GETA.", MT Summit 1987, pp. 26-35, 1987.
- [18] Arnold, D.J., "Eurotra: a European perspective on MT.", Proceedings of the IEEE 74(7), pp. 979-992, 1986.
- [19] Sadler, V., Working with analogical semantics: disambiguation techniques in DLT., Dordrecht: Foris, 1989.
- [20] Teruko M., The KANT System: Fast, Accurate, High-Quality Translation in Practical Domains, COLING-92, 1992.
- [21] Muraki, K., "PIVOT: two-phase machine translation system.", MT Summit 1987, pp. 81-83, 1987.
- [22] Wheeler, P., "LOGOS.", Sprache und Datenverarbeitung 9(1), pp. 11-21, 1985.
- [23] Mann, J.S., "Get Smart! Industrial strength language processing from Smart Communications.", Language Technology 3, pp. 12-15, September/October 1987.
- [24] Lynn E. W., Advantages and Disadvantages of Translation Memory : Cost/Benefit Analysis, Einfhrung in die Informatik II, Vol 6, 1998
- [25] Hutchins W.J., "Machine translation and human translation: in competition or in complementation?," International Journal of Translation Vol.13, no.1-2, pp. 5-20, Jan-Dec 2001.
- [26] Woszczyna, M. et al., "Recent advances in JANUS: a speech translation system.", TMI-93, pp. 195-200, 1993.
- [27] Wahlster, W., "Verbmobil: translation of face-to-face dialogs.", MT Summit 4, pp. 127-135, 1993.
- [28] Gerald D., "Translation Memory: Concept, products, impact and prospects", Major project report South Bank University School of Electrical, Electronic and Information Engineering, 1995.
- [29] Achim B., "Workflow using linguistic technology at the Translation Service of the European Commission", EAMT Workshop, Geneva, pp. 7-18, April 1998.

## 저자약력



강명주

1988 동국대학교 컴퓨터공학과 학사 졸업  
 1991 동국대학교 컴퓨터공학과 석사 졸업  
 1991~1993 아이비 컴퓨터 시스템  
 1993~1997 제주산업정보대학 컴퓨터 정보계열 전임강사  
 1997~2000 Tyco/Fire & Security/동방전자산업  
 2000~현재 클릭큐 부설 연구소 연구소장  
 관심분야 : 자연어처리, 데이터베이스, 정보 검색, 실시간 시스템

이 메 일 : mjikang@clickq.com