

불-한 번역도우미 FK-Tranassis의 FKT-DynaDic사전

서래원*·최민옥**

〈 차례 〉

I. 서론	IV. FKT-DynaDic 사전
II. 불한 번역도우미 FK-Tranassis 구성	V. FKT-DynaDic 사전의 논리적 구조
III. FK-Tranassis의 번역 흐름도	VI. 결론

I. 서론

1.1 연구배경

정보사회를 맞이하면서 인터넷과 같은 정보 기반시설에 대한 지속적인 투자가 이루어지고, 정보 창출 및 소비를 위한 정보 기기 보급의 끊임없는 가속화가 오늘날 전문가들뿐만 아니라 일반 사용자들까지도 국경과 지리적인 한계를 극복하고 다양한 정보들을 신속히 획득할 수 있게 만들었다. 그러나 그와 같은 풍성한 물리적 조건에도 불구하고, 유통되고 있는 정보들이 표현된 언어의 이질성으로 인해, 사용자들은 원하는 정보에 대한 검색에서뿐만 아니라 획득한 정보의 이용에서도 여전히 많은 어려움을 겪고 있는 것이 현실이다¹⁾.

1994년 4월 한국고속철도 건설공단과 프랑스 알스톰사가 고속철을 TGV로 계약 유치하게 됨으로써, 프랑스로 되어있는 TGV에 관한 광대한 기술 및 운영 시스템 자료에 대한 처리가 요구되었다. 뿐만 아니라, 나날이 빈번해지는 한불 두 국가 간의 교류에 보다 효율적인 정보 검색과 활용을 위해서는 물리적 기반들만큼, 지식산업의 핵심요소라고 할 수 있는 한불 번역 기술 개발에 대한 많은 관심과 투자가 시급히 요구되고 있다.

영어 혹은 일어에 관한 자동번역시스템 기술이 꾸준히 연구되어 왔으며, 최근에는 중국의 경제적 발달과 함께 중한 자동번역에 관한 많은 연구가 국책 및 기업 연구소에서 이루어지고 있으나, 한불에 관한 기초기술연구는 상업성의 미비로 인해 활발히

* 배재대학교 IT공학부 교수

** 배재대학교 NLIP 연구원

1) 서래원, 박세영, 유성준, 『다국어 정보처리를 위한 불어 전처리에 관한 연구』, 정보처리학회 Vo.17, No.4 1999.

이루어지고 있지 못한 형편이다.

일부에서 자동번역시스템을 이용하여 방대한 양의 자료들을 번역한다고 하지만 기계번역에 대한 한계성으로 완벽한 번역이 이루어지지 못하고 있다. 따라서 자동번역시스템으로 번역을 한 후에도, 수작업을 통하여 결과를 재검토하고 수정해야 하는 실정이다. 현재의 기술로는 자동 번역이 실질적인 면에서 실현성이 없다는 결론에 따라, 컴퓨터가 번역가를 지원할 수 있는 시스템에 대한 필요성이 제기 되었고, 이에 따라 시스템 CAT(Computer Aided Translation)이 등장하게 되었다.

1.2 CAT 등장

CAT은 전문가에게 번역에 필요한 부수적인 기능들을 제공하여 번역 작업의 능률을 높이고, 빠르고 정확한 번역을 할 수 있게 도와주는 시스템이다.

1980년대 PC의 보급과 함께 과도한 기대로 시작되었던 자동 번역의 희망은 자연언어처리 기술의 한계에 부딪쳐 전문 번역가들의 회의를 불러일으켰다. 이에 번역가들은 완전한 자동번역의 요구보다 번역의 편리성과 정확성을 제공할 수 있는 전문 용어 데이터 처리 및 처리 결과물의 관리를 해결해 줄 수 있는 S/W와 H/W적 대안을 요구하게 되었다. 이러한 요구에 따라 1990년대 초 문서처리와 데이터 관리가 접목된 번역 워크스테이션의 개발이 이루어졌다. 그간 통계기반 방법과 코퍼스 기반 방법이 주로 연구되어 왔으며 음성인식, 음성 합성, 규칙기반 또는 코퍼스 기반 번역모듈 등의 기술이 접목되면서 대화 음성 번역에 대한 연구도 시작되었다²⁾. 현재 개발되어 있는 응용프로그램들은 Atril(Deja Vu X), IBM(Translation Manager), SDL(SDLX2003) 그리고 Trados(Translator's Workbench) 등이 개발되어 있다. Atril 사의 Deja Vu X는 EBMT(example-based machine translation) 기술과 fuzzy matching이 주요기술이고, project managers에 의하여 고품질의 번역을 제공하고 있다³⁾. SDL사의 SDLX2003은 기번역과 미번역에 의하여 새로운 TM(translation memory)을 생성하고, 프레임마커로부터 html파일과 같은 다양한 Project Wizard를 제공하며, 데이터베이스관리기술로 개념지향기술을 제공하고 있다⁴⁾. Trados는 1984년에 Jochen Hummel 과 Iko Knyphausen에 의하여 개발된 것으로, 첫 번째 버전은 DOS를 관리하기 위한 memory-resident multilingual 기술이고, 두 번째 버전은 Windows를 위한 linguistic fuzzy matching기술이다. 그리고 1998년에는 S-Tagger technology를 개발하였고, 2000년에는 HTML/XML editor인 TagEditor, T-Window for PowerPoint 등과 연계하여 포맷 형태를 제공해주는 기술을 개발하였다⁵⁾.

2) 박주형, 이창우, 강명주, 『자동번역과 CAT의 현황과 전망』, 정보과학회지 Vo.19, No.10, 2001

3) <http://www.atril.com>

4) <http://www.sdlintl.com>

5) <http://www.localizationworks.com/DRTOM/Trados/TRADOS.html>

CAT시스템들의 공통적인 요소들은 정렬도구, 데이터 관리도구, 용어 관리도구, 문서 처리도구이다. 첫째 정렬도구는 기 번역된 전자 문서를 TM데이터로 만들 수 있게 해주는 도구로 대부분 GUI 형태이다. 정렬도구는 세그멘테이션 규칙을 통해 구문단 위 원문과 대역문을 데이터화하며, 구두점을 이용한 문장 경계 구분과 RTF(Rich Text Format)의 파일 형식 정보를 이용하여 편집자의 의도에 맞게 데이터를 추출한다⁶⁾. 둘째, 데이터 관리 도구는 텍스트, html, 프로그램소스와 같이 다양한 파일 형태를 포함하는 프로젝트 관리 도구이다. 용어 데이터와 품질관리 및 일관성 유지를 위해서 History데이터를 관리해주고, 프로젝트별 데이터를 효과적으로 검색, 적용할 수 있는 데이터 관리 프로그램이다. 셋째, 용어 관리도구는 사용자가 번역 프로젝트에 필요한 용어를 정의하고 번역 작업 환경에서 참조할 수 있게 하는 프로그램이다. 최근에는 대규모 프로젝트를 빠르고 일관성 있게 처리하기 위해 대부분 퍼즈 매칭 알고리즘을 사용하고 있으며, 클라이언트/서버 환경에 적합한 네트워킹을 제한된 범위에서 지원하고 있다. 넷째, 문서 처리도구는 워드프로세서와 같이 문서 처리하는 기능을 가지고 있어 번역과정을 용이하게 만들어주는 기능을 가지고 있다. 또한 문서 처리도구는 맞춤법 및 철자교정기능을 가지고 있으며 번역 메모리 시스템과의 연동을 최우선으로 지원한다. 문서처리 도구의 파일 형식을 분석하는 필터(Format Filter)는 독립적인 문서 처리를 할 경우 다른 타상 프로그램이나 문서 처리파일을 편집하여 다양한 형태로 import/export를 할 수 있도록 지원한다⁷⁾.

현재 국내의 CAT시스템은 기반기술 및 번역 연동 기술 확보에 어려움을 겪고 있다. 몇몇 연구소와 기업들이 공동으로 동양 4개국 언어에 대한 TM기반의 통합 CAT 시스템 개발을 하고 있는 실정이다. 이러한 시대적 상황에 따라 불한 번역 기술 개발의 일환으로, 불한 번역 도우미 FK-Tranassis⁸⁾를 설계 구현하게 되었으며, 이 시스템의 핵심 요소기술에 해당되는 FKT-DynaDic⁹⁾을 개발하게 되었다. 따라서 본 논문에서는 전반부에 불한 번역 도우미 FK-Tranassis의 개괄적인 시스템 구성을 소개하고, FK-Tranassis의 핵심 구성요소인 FKT-DynaDic사전의 논리적 구조와 기능에 대하여 논하고자 한다.

<http://www.trados.com>

- 6) Achim B., *Workflow using linguistic technology at the Translation Service of the European Commission*, EAMT Workshop, Geneva, pp. 7-18, April 1998.
- 7) Gerald D., *Translation Memory: Concept, products, impact and prospects*, Major project report South Bank University School of Electrical, Electronic and Information Engineering, 1995.
- 8) French Korea Translation Assistance의 약자
- 9) French Korea Translation Dynamic dictionary의 약자

II. 불한 번역 도우미 FK-Tranassis 구성

'번역 도우미'란 명칭은 기존의 번역 관련 도구들과의 차별성을 도모하기 위한 용어이다. FK-Tranassis의 특성이 번역을 보조해 주는 역할인데 착안하여, 93' 대전 엑스포에서 처음 사용되기 시작한 "행사 안내를 맡거나 남에게 봉사하는 요원"이란 의미를 가진 '도우미'라는 단어를 사용하였다. 불한 번역 도우미 FK-Tranassis는 자동 번역기의 기술적 보완을 위한 시스템이라기보다는 다른 관점에서 번역 방법을 제안하는 CAT시스템의 하나이다. 즉, 번역을 해주는 것이 아니라 번역가가 번역을 하는 과정에서 필요한 부수적인 기능들을 제공하여 작업 능률을 높여서 빠르고 정확하게 번역을 할 수 있게 도와준다.

불한 번역 도우미 FK-Tranassis는 다음과 같은 요구 사항들을 만족해야한다. 원문을 분해하여 축약¹⁰⁾되기 전의 원형으로 전환해주는 원문 분해기능¹¹⁾과 기존 번역 사례를 검색하여 참조할 수 있도록 제공하는 참조 구문 기능 그리고, 최적 단어 추천을 제공하는 사전 기능이 제공되어야한다. 그리고 기본적으로 번역작업을 하는데 있어서 입출력을 용이하게 하기 위해 기본 기능들이 제공되는 최적의 웹서비스가 제공되어야한다. 이러한 요구조건을 충족하고자 불한 번역 도우미 FK-Tranassis를 개발하게 되었다.

이 FK-Tranassis의 주요 기능으로는 첫째, 모든 정보를 웹 브라우저라는 편리한 인터페이스로 제공한다는 것이다. 웹페이지 기반으로 구현되었기에 네트워크 상에서 장소에 구애됨이 없이 실시간 번역 작업을 할 수 있다. 둘째, 번역 작업을 하는 동안 모든 작업의 내용을 수정하고 삭제 할 수 있으며, 맞춤 출력 기능을 제공하여 전문가가 원하는 형식으로 번역 내용을 출력 할 수 있게 한다. 셋째, 문서내의 검색기능

10) 불어는 형태소 결합 특성을 가지고 있다. 불어에서 나타난 형태소 결합현상은 다음과 같은 4가지로 분류된다.

- 연결 형태 (concatenation form) : 정관사 /le/와 관계사 /quel/이 결합되어 복합관계사 /lequel/이 됨과 같이, 단순히 두 단어가 연결되어 구성된 형태.

예) /lequel/, /auquel/, /duquelles/,

- 축약 형태 (elision form) : 두 단어 사이의 모음이 탈락되면서 apostrophe가 추가되어 결합되는 형태.

예) /s'il/, /d'une/, /c'est/, /s'aime/, /l'on/,

- 결합형태 (hyphenated form) : 자음이나 모음의 탈락이 없이 단순히 hyphen으로 연결되는 형태.

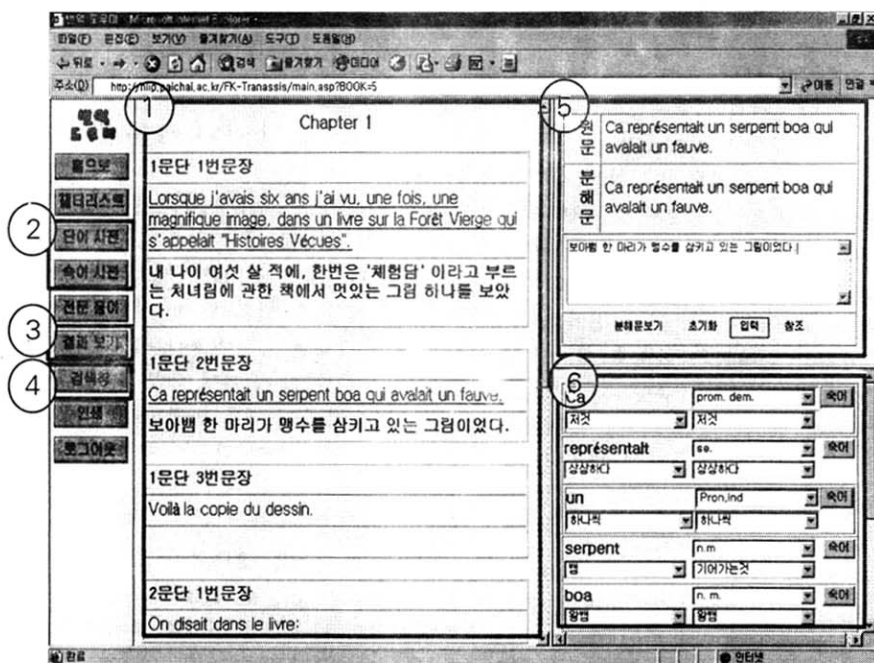
예) /l -dedans/, /ci-contre/, /aime-je/,

- 아말감 형태 (amalgamation form) : 앞에서 본 축약관사의 예에서와 같이 두 단어가 아말감 현상에 의해 결합된 후 새로운 형태의 어절로 변환된 형태.

예) /au/, /du/, /aux/, /des/, /aucun/,

11) 서래원, 박세영, 유성준, 『다국어 정보처리를 위한 불어 전처리에 관한 연구』, 정보처리학회 Vo.17, No.4 1999.

과 책갈피 기능과 같은 보조 기능을 제공하여, 번역을 하는데 있어서 다양한 편리성을 제공한다. 넷째, 불한 전자 사전과 별도로 번역자 개개인에게 맞게 구성할 수 있는 개별사전 추가 기능이 제공되며 또한, 전문 용어 사전을 제공하여 전문 분야의 번역을 편리하게 할 수 있게 한다. 다섯째, 통계기반 방법과 어휘 빈도에 대한 시소러스 방법을 사용하여 기존 번역 사례를 참조문으로 제시해 불필요한 중복 번역을 줄여줌으로써 번역속도와 질을 향상시킨다. 마지막으로 사이트에 접속한 다른 번역 전문가에게 실시간으로 도움을 요청할 수 있는 기능도 제공한다. 이와 같이 불한 번역도우미 FK-Tranassis는 전문가에게 번역을 위한 보조 기능을 제공해 다량의 문서처리에 도움을 주는 웹상에서 서비스되는 인터페이스이다.



<그림 1. FK-Tranassis의 본문부분>

위의 <그림 1>은 불한 번역도우미 FK-Tranassis의 본문부분으로 ①원문부분과 ②사전부분 ③결과보기부분 ④검색창 부분 ⑤번역입력 및 참조부분 ⑥단어의 의미속어부분으로 구성된다.

원문부분은 번역 단위가 문장이고, 문단과 문장의 위치를 한눈에 파악할 수 있다

록 위치 정보를 수록한다. 또한 이미 번역된 문장을 1/3정도 함께 제시해줌으로써 번역의 단절을 막을 수 있어 자연스러운 번역을 할 수 있도록 한다.

사전부분은 동철이의어를 구분하기 위해서 정렬을 위한 단계를 따로 두어 정렬 표제를 설정하며 하나의 표제어에 파생되는 어휘를 내포하는 구조를 각각 독립된 표제어로 제시한다. 또한, 불규칙 굴절 변화형을 표제어로 등재하고 기본형 정보 및 굴절 정보를 제공하거나, Nlip 불어 형태소 분석기를 이용하여 기본형과 굴절형 간의 관계를 정의해 준다.

결과보기부분은 지금까지 해온 번역문의 결과를 보여주는 기능으로, 결과보기 유형에는 원문만 보여주는 형태와 번역문만을 보여주는 형태 그리고 원문과 번역문을 함께 보여주는 형태 등 세 가지 유형을 제공한다. 원문만 보여주는 형태는 원문에 대한 전체적인 내용을 파악하게 함으로써 번역 진행 방향을 가늠하게 하고, 번역문만 보여주는 형태는 전체적인 번역결과가 원본의 내용에 맞게 원활하게 번역되었는지 확인할 수 있게 한다. 그리고 원문과 함께 번역문을 보여주는 형태는 원문과 번역문을 한 문장씩 보여주어, 원문과 번역문의 비교 분석을 가능하게 하고, 최적의 번역 결과를 얻기 위한 편리한 편집기능을 제공한다.

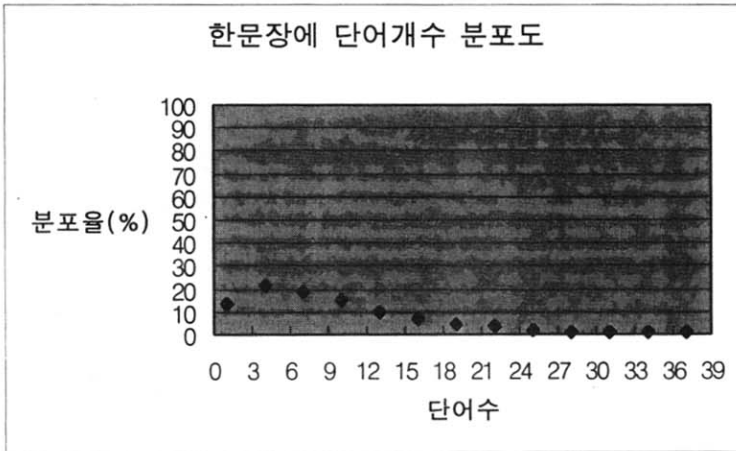
검색창 버튼은 원본과 번역된 결과물에 대하여 완전한 문장을 찾고자 할 때, 즉 전문검색을 할 때나 혹은 한 문장의 일정 단어를 찾고자 할 경우 사용한다.

번역입력 및 참조부분은 전문가가 원문을 보고 번역을 입력하는 부분으로, 원문과 원문을 분해한 분해문과 이전에 번역된 문장을 참조할 수 있게 한다. 분해문은 원문에 제시된 단어를 원형으로 제시한 것으로 불어의 다양한 동사변형과 축약에서 야기되는 중의성을 다소 줄일 수 있다. 또한, 분해문은 각각의 단어에 대해 대응되는 번역문장의 단어들이 출현하는 빈도를 통계적 지식으로 활용하여 번역하는 방법인 통계 기반 방법을 이용해 이전에 번역되어온 번역문과 비교하여 같은 문장을 참조문¹²⁾으로 제시한다.

불한 번역 도우미 FK-Tranassis의 예문¹³⁾으로 사용된 생텍쥐페리의 『Le petit prince』의 문장에 대한 단어의 분포를 살펴보면, 한 문장을 구성하는 단어의 수가 15개 미만인 경우가 79.10%정도이고 15개 이상인 경우는 20.90% 정도였다. 아래 <그림 2>은 『Le petit prince』의 27 개의 chapter 가운데 표본 13개의 chapter를 문장단위로 나누어 단어를 추출한 후 단어의 출현빈도를 분포도로 나타낸 것으로 위 사실을 잘 나타내고 있다.

12) 참조문이란 기번역 문장 중 번역하고자 하는 문장과 유사한 문장을 의미한다.

13) 본 시스템의 검증을 위해 생텍쥐페리의 『Le petit prince』를 번역 대상으로 실험하였다. 본 실험에서 사용된 『Le petit prince』는 총 1570개의 문장으로 구성되었고, 총 12072개의 단어가 사용되었다. 그 중 유일한 표제어의 수는 2356개이다.



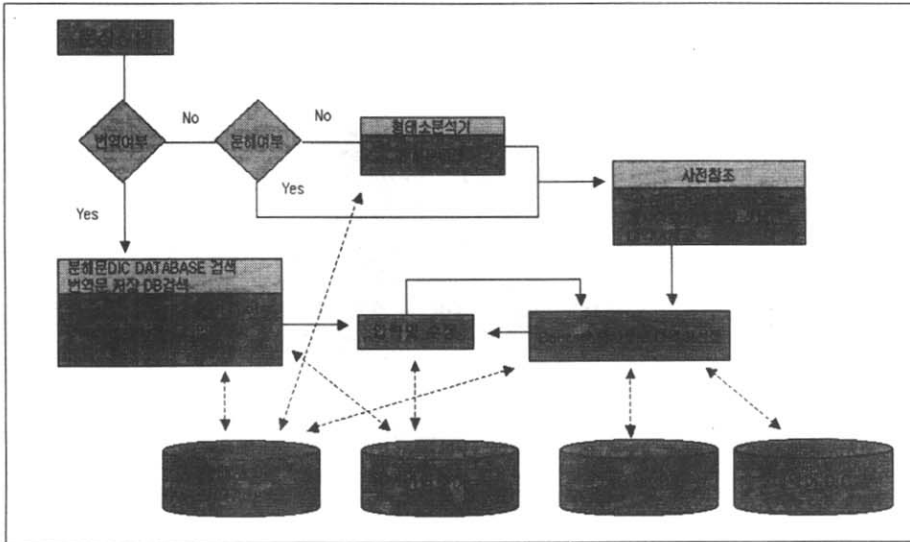
<그림 2. FK-Tranassis 예문의 단어 개수 분포도>

일반적으로, 한 문장을 구성하는 단어의 수가 4개에서 10개 사이인 경우가 가장 많다. 실험에서 단어 수가 9개인 예문을 번역한다고 가정할 때, 이 예문을 분해한 분해문의 기본 단어가 10개로 되었다면, 이 분해문의 기본 단어들과 기존에 이미 번역된 문장의 분해된 기본 단어들과 비교해, 그 유사도에 따라 우선순위를 두어 기번역 문장을 참조문으로 먼저 제시해준다. 유사도가 낮은 경우 번역하고자 하는 분해문의 기본 단어 수 절반에 해당하는 임의의 기본 단어 5개까지를 비교하여 유사도가 제일 높은 기번역 문장을 참조문으로 제시함으로써 번역 속도와 정확률을 증대시켰다.

Ⅲ. FK-Tranassis의 번역 흐름도

FK-Tranassis에서 한 문장을 번역하는데 있어 선택된 문장에 따라 기처리 문장인 경우와 미처리 문장인 경우가 있다. 이 선택된 문장을 번역하는데 있어서 FK-Tranassis의 FKT-DynaDic사전과 분해문 DIC DATABASE 그리고 번역문 DB가 각각 어떠한 의미의 연관관계를 가지고 번역 하게되는지 FK-Tranassis의 문장 번역 흐름도를 통하여 알아보려고 한다.

<그림 3>은 FK-Tranassis의 문장 번역흐름도로 기처리 문장과 미처리 문장 번역 시 제공되는 FKT-DynaDic사전과 분해문 DIC DATABASE 그리고 번역문 DB 각각의 관계를 나타낸 그림이다.



〈그림 3. 번역도우미 FK-Tranassis의 문장 번역 흐름도〉

FK-Tranassis에서, 번역가가 번역을 위해 선택한 문장이 미처리 문장인 경우와 기처리 문장인 경우 각각 서로 다른 번역 흐름 절차를 나타낸다. FK-Tranassis에서 번역을 위해 선택한 문장이 미처리 문장인 경우 번역되는 흐름 절차를 살펴보면 첫째, 번역도우미 FK-Tranassis 본문 부분(그림 1)의 원문모듈에서 한 문장을 선택한다. 둘째, 선택한 문장이 미처리 문장으로 분해문이 존재하지 않는 경우, Nlip 형태소 분석기¹⁴⁾로 분해문을 생성하여 분해문 DIC DATABASE에 저장한다. 셋째, 분해문 DIC DATABASE에서 분해된 분해 단어를 찾아 각 분해 단어마다 FKT-DynaDic사전의 Concept Dic사전에서 개념어의 가중치가 높은 개념어와 품사정보를 제시해주고 이 개념어에 대한 Translation Dic사전에서 대역어에 대한 가중치가 높은 대역어를 자동으로 선택하여 먼저 제시해준다. 넷째, 제시해준 개념어, 품사정보, 대역어를 번역에 맞게 선정하여 번역문을 입력하면 번역문DB에 번역문이 저장된다. 만약 번역문을 수정할 경우에는 개념어, 품사정보, 대역어를 번역에 맞게 재 선정할 수 있다.

FK-Tranassis에서 선택한 문장이 기처리 문장인 경우 번역되는 흐름 절차를 살펴보면 첫째, 분해문 DIC DATABASE에서 번역시 사용된 대역어와 그 대역어의 사전정보 그리고 분해문을 검색하고 번역문DB에서 기록된 번역문을 검색하여, 각

14) Seo, L. A Study on the Regularization of the French Sentence for Multilingual Information Processing. Proceedings of The 9'th KISS Fall Conference Region Chung-cheong

서래원, *Analyseur de la Phrase Simple - A.P.S.*, 한국불어불문학회 1997년 하계발표회, 1997

분해 단어마다 개념어, 품사정보 그리고 대역어를 제시해주고 번역문을 제시해준다. 둘째, 기처리 문장을 수정하게 되면 번역가가 개념어를 재 선정하게 되므로 개념어의 가중치가 바뀌게 되고 또한, 재 선정된 개념어에 의해 대역어의 가중치도 바뀌게 된다. 셋째, 재 선정된 개념어, 품사정보, 대역어를 기반으로 새로운 번역문을 입력하면 번역문DB에 번역문이 저장된다. 만약 새로이 번역된 문장을 재 수정할 경우에는 개념어, 품사정보, 대역어를 재 선정 할 수 있다.

IV. FKT-DynaDic사전

일반적으로 전자사전이란 용어는 '기계 가독형 사전'과 자연언어 자동처리에 쓰일 수 있도록 표제어의 상세하고 일관된 언어적 정보를 제공하는 '온라인 어휘부'란 용어를 모두 포함한다¹⁵⁾. 번역도우미에서 사용되는 전자사전은 이와 같이 정의된 일반적인 전자사전과 동일한 의미의 사전이라고 할 수 있다.

불한 번역도우미 FK-Tranassis의 FKT-DynaDic사전은 Translation Dic사전과 Concept Dic사전으로 구성되어있다. Concept Dic사전은 개념어 사전으로써, 하나의 표제어에 파생된 어휘를 모두 표제어로 제시하고 용언을 제외한 불규칙 굴절 변화형을 표제어로 등재하여 기본형 정보 및 굴절 정보를 제공한다. 또한 용언에 있어서는 용언의 기본형만을 사전에 등재하고, 굴절된 단어는 Nlip 불어 형태소 분석기를 이용하여 기본형과 굴절형 간의 관계를 정의해 준다. Translation Dic사전은 대역어 사전이다. FKT-DynaDic사전의 특징은 번역가의 번역 성향에 맞게 대역어를 제공해주고, 또한 대역어가 될 수 있는 확률이 높은 최적의 대역어를 우선순위로 제공해준다.

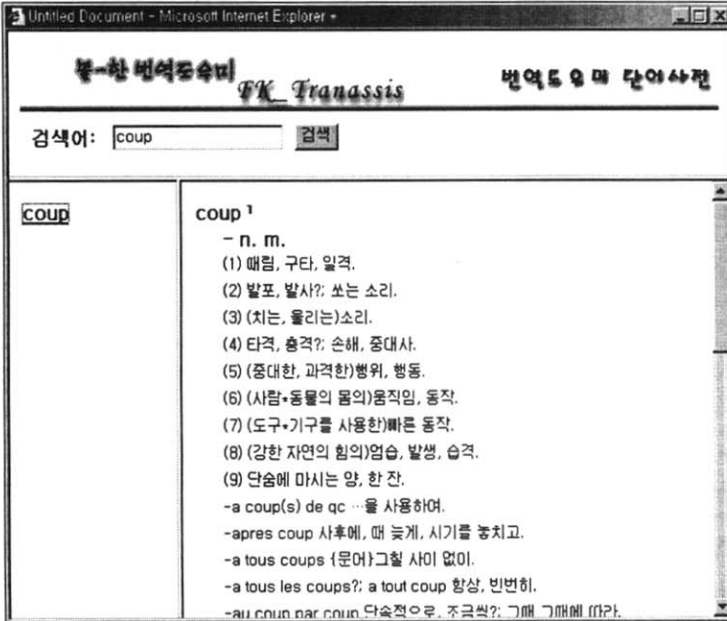
아래 <표 1>은 불한 번역 도우미 FK-Tranassis의 FKT-DynaDic사전의 기본 문서 구조를 규정하는데 사용한 어휘 기술 항목의 분류와 특성을 보여준다.

어휘 기술 항목	특 성
표제어	불어사전에 등재된 단어
기본형	표제어로 선정된 단어의 기본형
구분자	표제어 내부의 동철이의어와 같은 하위 표제어구분
품사	표제어의 품사
개념어	품사에 따른 개념어
대역어	대역어 리스트
상세의미	품사에 따른 상세 의미들

<표 1. 기본사전구성의 어휘기술항목과 특성>

15) 윤애선, 『표준화된 전자사전 개발을 위한 정보구조 및 사용자 환경 설계와 그 응용』, 한국프랑스학회논문집, Vol.33, No.0, 2001

〈그림 4〉는 불한 번역도우미 FK-Tranassis의 FKT-DynaDic사전이다. FKT-DynaDic의 Concept Dic사전에서 표제어에 대하여 동철이의어를 구분하고 품사에 따라 개념어를 추출하였으며, 추출된 개념어에 따라 Translation Dic사전에서 대역어를 추출하여 구현된 모습이다.



〈그림 4. 불한 번역도우미 FK-Tranassis의 FKT-DynaDic사전〉

구현된 불한 번역도우미 FK-Tranassis의 FKT-DynaDic사전은 번역시 한 표제어에 대하여 많이 사용한 대역어를 우선순위로 표시한다. 즉, 사용 빈도가 높은 대역어 가중치에 의해ダイナミック한 불한 번역도우미 FK-Tranassis의 사전이 구축된다.

숙어사전은 전문가가 해당 단어에 대하여 숙어만을 찾고자 할 경우 사용되는 사전이다. 〈표 2〉는 숙어사전의 기본 구성을 규정하는데 사용한 어휘 기술항목과 특징을 보여준다.

어휘 기술 항목	특 성
표제어	불어사전에 등재된 단어
구분자	표제어 내부의 동철이의어와 같은 하위 표제어구분
품사	표제어의 품사
숙어	품사에 따른 숙어
숙어의미	숙어에 대한 의미

〈표 2. 숙어사전의 어휘적 기술항목과 특징〉

숙어사전은 표제어의 품사에 따라 숙어를 제공한다. 이 사전은 웹서비스용 숙어사전으로 Concept Dic사전 혹은 Translation Dic사전과 마찬가지로 관리자에 의해서 표제어의 등재 여부와 품사 정보들을 수정하거나 삭제를 할 수 있게 한다. 또한 전문가가 자주 쓰는 단어들이나 신조어들을 따로 등록하여 사용할 수 있는 사용자 숙어 사전이 제공된다. 관습과 문화의 차이 또는 특정 학문의 전문적인 내용일 경우에는 일반 사전만으로는 번역에 한계가 있다. 이런 경우 전문용어 사전을 이용하면 보다 정확한 번역을 할 수 있다. 단어사전이나 숙어사전과 마찬가지로 전문용어 사전도 웹서비스용 사전으로 표제어의 품사에 따라 어휘 정보를 제공한다.

V. FKT-DynaDic사전의 논리적 구조

FK-Tranassis의 실행에 있어서 번역하고자하는 문장이 기처리 번역 문장인가 혹은 미처리 번역 문장인가에 따라 분해문 DIC DATABASE와 FKT-DynaDic사전의 논리적 적용 방법이 변한다. 분해문 DIC DATABASE는 기처리 문장을 구성하는 기본 단어들의 정보를 수록하고 있는 DATABASE이다. 직접적인 정보구성으로 기본단어, 대역어, 기본단어의 순서, 기본단어의 소속, 기본단어의 사전정보 위치가 제공된다.

Concept Dic사전은 번역을 위한 일반적이고 기초적인 기본 사전으로 특정 문장에 따른 전문용어 사전이나 개인 사전을 의미하는 것이 아니라 보편적이고 기본이 되는 어휘 사전으로 구성된다. Concept Dic사전에서 제공되는 정보는 표제어, 품사, 개념어 그리고 개념어의 사용 빈도에 따른 개념어의 가중치를 나타내는 정보가 있다. Concept Dic사전은 유일형태의 표제로 구성된 사전이 아니므로, 하나의 대역어를 추출하기 위해서는 표제어, 품사, 개념어 조합되어야만 유일한 개념을 적용한 표제어를 선정하게 된다. Concept Dic사전에서 선정된 하나의 표제어가 다수의 개념을 나타낼 수 있는 경우 번역가의 성향에 따라 사용되는 개념은 달라지게 된다. 개념어의 사용빈도에 의하여 개념어의 가중치를 부여하는 것은 시스템을 사용하는 번역가가 사용해 왔던 기존 개념 설정 성향을 차기 번역에 재 사용하고자 함이다.

Translation Dic사전은 번역을 위한 대역어들로 구성된 대역어 사전이다.

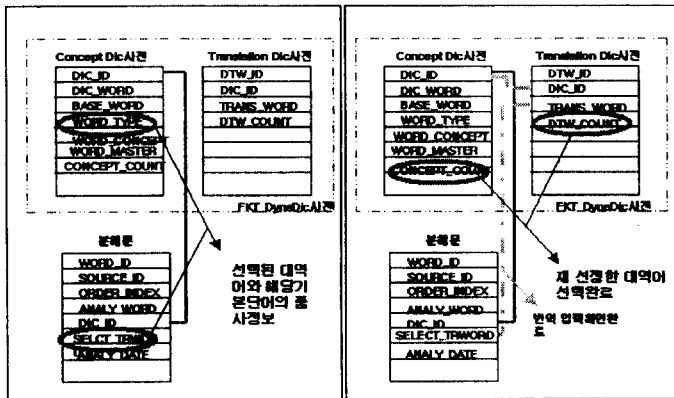
Translation Dic사전은 최적의 대역어를 제시해준다. Translation Dic사전에서 제공되는 정보는 대역어, 대역어의 소속, 대역어의 개념정보 그리고 대역어의 사용 빈도에 따른 대역어의 가중치 등이다.

번역가가 한 표제어에 대해 특정 대역어만을 주로 사용하여 번역하였을 경우 이는 다음에 같은 표제어가 나올 때 그 표제어에 대한 대역어가 될 수 있는 확률이 매우 높다. 특정 개념어에 대한 대역어의 사용빈도에 따른 대역어의 가중치 정보는, 한 표제어에 대하여 최적의 대역어를 우선순위로 제공하기 위한 중요한 정보로 간주된다.

FKT-DynaDic사전과 분해문 DIC DATABASE의 논리적 구조에 어떠한 변화가 있는지 기처리 번역 문장 일 경우와 미처리 번역 문장일 경우를 살펴보도록 한다.

기처리 번역 문장일 경우 기본 단어들로 수록되어있는 분해문 DIC DATABASE의 정보구성 요소 중에는, 번역에 사용된 대역어와 그 대역어의 사전적 정보 즉, Concept Dic사전에서 제공해주는 품사의 정보가 있다. 이는 기처리된 문장에서 대역어의 사전정보가 무엇인지 알 수 있으며 또한, 문장 번역시 번역가 자신만의 고유한 번역 성향을 알 수 있게 하는 정보가 된다. 다른 대역어를 재선정하여 번역문장을 수정할 경우에는 선택된 대역어에 따라 개념어의 재선정이 요구될 수 있으며, 개념어의 재선정시 개념어의 가중치가 바뀌게 된다. 또한 재선정된 개념어에 의해 대역어의 가중치가 바뀌기도 한다. 재선정된 대역어에 의하여 Concept Dic사전과 Translation Dic사전의 데이터 구조가 변하게 되므로 번역가의 성향에 따른 대역어 자동 선택의 정확률이 높아진다. 또한 번역가가 번역을 완료했을 경우 분해문 DIC DATABASE의 정보구성 요소 중 선택된 대역어 저장부분과 그 대역어의 사전적 정보부분이 Update되면서 논리적 구조가 변한다.

〈그림 5〉은 기처리 번역 문장일 경우 분해문 DIC DATABASE와 Concept Dic사전과의 관계를 나타내는 다이어그램이다.

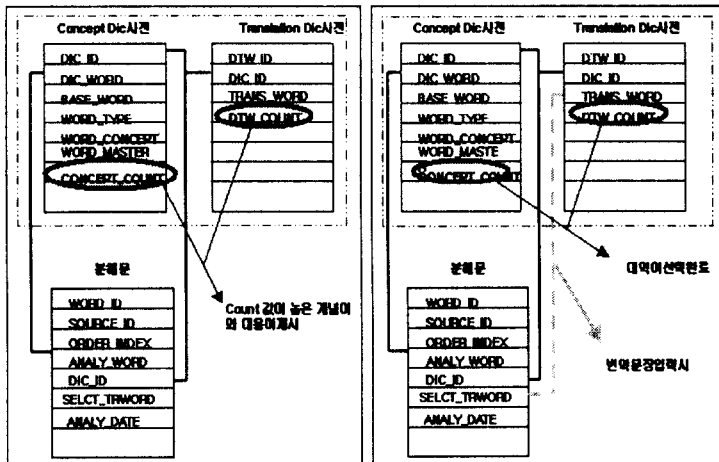


〈그림 5. 기처리 문장의 분해문 DIC DATABASE와 FKT-DynaDic사전과의 Relationship〉

<그림 5>의 좌측 부분은 번역된 대역어와 관련된 해당 원 문장의 기본단어의 품사 정보를 나타내는 그림이고 우측 부분은 대역어를 재선정시 분해문 DIC DATABASE와 FKT-DynaDic사전과의 Relationship을 나타낸 그림이다.

미처리 번역 문장일 경우는 Nlip 불어 형태소 분석기가 문장을 분해하여 분해문 DIC DATABASE에 어순에 맞게 기본단어를 저장한다. 분해문 DIC DATABASE의 기본단어를 Concept Dic사전에서 검색하여 사용빈도가 높은 개념어를 선택하고, 이 선택된 개념어에 해당하는, Translation Dic사전에서 가장 사용빈도가 높은 대역어를 먼저 제시해준다. 이는 제공되는 대역어가 번역가의 성향을 고려하여 먼저 제시되었음을 의미한다. 번역가가 선택한 개념어에 대해 대역어를 선택하여 완료하면, 선택된 개념어와 선택된 대역어에 따라 Concept Dic사전과 Translation Dic사전의 데이터 구조가 변한다. 이는 번역가의 번역성향에 따른 대역어를 제공해주는 정확률을 높이는 정보가 된다.

<그림 6>은 미처리 번역 문장일 경우 분해문 DIC DATABASE와 Concept Dic사전과의 관계를 나타내는 다이어그램이다.



<그림 6. 미처리 문장의 분해문 DIC DATABASE와 FKT-DynaDic사전과의 Relationship>

<그림 6>의 좌측 부분은 번역할 개념어에 대한 대역어를 제공해 주는 그림이고, 우측 부분은 대역어 선택 완료 및 번역문장 입력 완료시 분해문 DIC DATABASE와 FKT-DynaDic사전과의 Relationship을 나타낸 그림이다.

기처리 번역 문장일 경우 분해문 DIC DATABASE와 FKT-DynaDic사전 각각의

논리적 구조와 연관성, 그리고 미처리 번역 문장일 경우 분해문 DIC DATABASE 와 FKT-DynaDic사전 각각의 논리적 구조와 연관성에 대하여 알아보았고, 이를 통하여 번역가에게 번역 성향에 맞는 대역어를 제공해 주는 과정에 대하여 알아보았다.

VI. 결 론

본 논문은 번역 작업 능력을 높이고 정확한 번역을 도와주는 CAT시스템의 등장과 정의에 대하여 설명하였고, 이 같은 다국어 번역기술의 일환으로 Nlip 붙어 형태소 분석기를 기반으로 한, 불한 번역 도우미 FK-Tranassis의 핵심 부분인 FKT-DynaDic의 구조와 기능에 대하여 설명하였다.

FK-Tranassis은 기존의 자동번역에서의 이론적 접근 방법이나 규칙들을 사용하여 번역기를 구현한 것이 아니라 번역가와 컴퓨터가 서로 상호 보완하여 빠르고 정확한 번역을 위하여 도움을 줄 수 있는 불한 번역 도우미이다. 또한, FK-Tranassis은 불어를 한국어로 번역하는 대량의 문서를 처리함에 있어서 전문가에게 원문 분해기능과 기존 번역 사례 참조 기능, 최적 단어 추천 기능을 제공해주어 빠르고 효율적인 번역 작업을 도와주기 위한 도구이다. 불한 번역 도우미 FK-Tranassis에서 최적 단어 추천 기능을 제공하는 핵심 요소기술은 FKT-DynaDic이다. FKT-DynaDic은 Translation Dic사전과 Concept Dic사전으로 구성되어 있다. Concept Dic사전은 번역을 위한 일반적이고 기초적인 기본 사전으로 특정 문장에 따른 전문용어 사전이나 개인 사전을 의미하는 것이 아니라 보편적이고 기본이 되는 어휘 사전이고, 유일형태의 표제로 구성된 사전이 아니라 하나의 대역어를 추출하기 위해서 표제어, 품사, 개념이 조합된 사전이다. Translation Dic사전은 문장에 맞는 번역을 위해 선택되는 대역어들로 구성된 사전이다. 본 논문에서는 FKT-DynaDic의 논리적 구조와 기능을 설명하면서 FKT-DynaDic이 번역가 개인의 번역성향에 맞게 대역어를 제시해주는 방법과 한 표제어에 대하여 대역어를 제시해주는 확률적인 방법에 대하여 설명하였다.

참고문헌

- Achim B., *Workflow using linguistic technology at the Translation Service of the European Commission*, EAMT Workshop, Geneva, pp. 7-18, April 1998.
- Berrendonner A., *Grammaire pour un analyseur: aspects morphologiques*. Lyon: document de travail du groupe SYDO, 1984.
- Gerald D., *Translation Memory: Concept, products, impact and prospects*, Major project report South Bank University School of Electrical.

- Electronic and Information Engineering, 1995.
- 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.
- Seo L., *A Study on the Regularization of the French Sentence for Multilingual Information Processing*, Proceedings of The 9'th KISS Fall Conference Region Chung-cheong
- Seo L., *Contribution au traitement automatique du verbe francais: lexicque, morphologie et syntaxe*, these en doctorat. Lyon: univresite-lumiere, 1995.
- 박주형, 이창우, 강명주, 『자동번역과 CAT의 현황과 전망』, 정보과학회지 Vo.19, No.10, 2001
- 서래원, *Analyseur de la Phrase Simple - A.P.S.*, 한국불어불문학회 1997년하계발표회, 1997.
- 서래원, 박세영, 유성준, 『다국어 정보처리를 위한 불어 전처리에 관한 연구』, 정보처리학회 Vo.17, No.4 1999.
- 윤애선, 『표준화된 전자사전 개발을 위한 정보구조 및 사용자 환경 설계와 그 응용』, 한국프랑스학회논문집, Vol.33, No.0, 2001
- <http://www-rali.iro.umontreal.ca>
- <http://www.localizationworks.com/DRTOM/Trados/TRADOS.html>
- <http://www.trados.com>
- <http://www.atril.com>
- <http://www.sdlintl.com>

〈Résumé〉

Dictionnaire FKT-DynaDic pour un système d'Aide de la traduction Franco-Coréen nommé FK-Tranassis

SEO Lai-Won / CHOI Min-Ok

Dans ce présent, nous présentons le dictionnaire FKT-DynaDic pour un système d'aide de la traduction automatique Franco-Coréen nommé FK-Tranassis. FKT-DynaDic est un dynamique dictionnaire du système FK-Tranassis basé d'un analyseur morpho-syntaxique de NLIP. Ce

dictionnaire se compose de deux sous-dictionnaires : Translation Dic et Concept Dic. Le dictionnaire Concept Dic est une base de données de conceptions des mots à traduire. Le dictionnaire Translation Dic a des informations sur des mots traduits et des informations de la tendance du traducteur sur ce mot traduit. En présentant des relations de ces deux dictionnaires, nous montrons l'importance du dictionnaire FKT-DynaDic dans le système d'Aide de la Traduction Franco-Coréen nommé FK-Tranassis.

주제어(Mots-clés) : 자연언어처리(traitement automatique des langues naturelles), 불한번역도우미(aide de la traduction), 동적사전(dictionnaire dynamique), 형태소분석기(analyseur de la phrase), 지식베이스 (la base de donnée de la connaissances)