

번역도우미 FK-Tranassis 의 FKT-DynaDic 사전

최민옥*, 서래원
배재대학교 컴퓨터공학과
*e-mail : mochoi@mail.pcu.ac.kr

FKT-DynaDic of Translation Assistance FK-Tranassis

Min-Ok Choi *, Laiwon Seo
Dept. of Computer Engineering, paichai University
*e-mail : mochoi@mail.pcu.ac.kr

요 약

본 논문은 다국어 번역기술의 일환으로 Nlip 의 불어 형태소 분석기를 기반으로 한 불-한 번역 도우미 FK-Tranassis 의 구성에 대하여 설명하고 불-한 번역 도우미 FK-Tranassis 에서 제공되는 FKT_DynaDic 사전의 논리적 구조에 대하여 설명하여 FKT_DynaDic 사전에서 번역가의 성향에 맞는 대역어를 제공해주는 방법에 대하여 설명하고자 한다.

1. 서론

정보 창출 및 소비를 위한 정보 기기 보급의 가속화와 더불어 인터넷과 같은 정보 기반시설에 대한 지속적인 투자가 이루어지고 있으며 그 결과 오늘날의 사용자들은 국경과 지리적인 한계를 극복하고 다양한 정보들을 신속히 획득할 수 있게 되었다. 그러나 그와 같은 풍성한 물적 조건에도 불구하고, 유통되고 있는 정보들이 표현된 언어의 이질성으로 인해 사용자들은 원하는 정보에 대한 검색에서 뿐만 아니라 획득한 정보의 이용에서도 여전히 많은 어려움을 겪고 있는 것이 현실이다[5].

1994 년 4 월 한국고속철도 건설공단과 프랑스 알스톰사가 고속철을 TGV 로 계약함으로써 고속철을 유치하게 되었고, 프랑스어로 되어있는 TGV 에 관한 광대한 기술 및 운영 시스템 자료에 대한 처리가 요구되었다. 뿐만 아니라, 나날이 빈번해지는 한불 두 국가간의 교류에 보다 효율적인 정보 검색과 활용을 위해서는 물리적 기반들만큼 한불 번역 기술 개발에 대한 많은 관심과 투자가 시급하다.

영한 혹은 일한에 관한 자동번역시스템이 꾸준히 연구되어 왔으며, 최근에는 중국의 발달과 함께 중한 자동번역에 관한 많은 연구가 국책 및 기업 연구소에서 이루어지고 있으나 한불에 관한 기초기술연구는 상업성의 미비로 인해 활발히 이루어지고 있지 못한

형편이다.

일부에서 자동번역시스템을 이용하여 방대한 양의 자료들을 번역한다고 하지만 기계번역에 대한 한계성으로 완벽한 번역이 이루어지지 못하고, 자동번역 시스템으로 번역을 한 후에도, 수작업을 통하여 결과를 검토하여 수정해야 하는 실정이다. 자동 번역은 실질적인 면에서 실현성이 없다는 결론에서, 컴퓨터가 번역가를 지원할 수 있는 시스템에 대한 필요성이 제기되었고, 시스템 CAT(Computer Aided Translation)이 등장하게 되었다.

CAT 는 전문가를 위해 대량의 문서를 번역하는데 도움을 주는 기능을 제공한다[6].

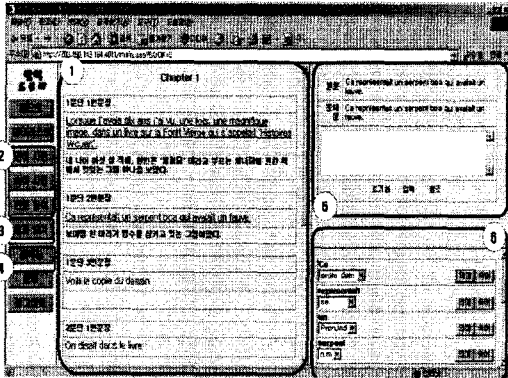
같은 맥락에서 이루어지고 있는 다국어 번역 기술 일환으로, 개발되고 있는 불-한 번역 도우미 프로토타입인 FK-Tranassis 프로토타입의 구성을 본 논문에서 소개하고 FK-Tranassis 프로토타입에서 제공해주는 FKT-DynaDic 사전의 논리적 구조에 대하여 논하고자 한다.

2. 번역 도우미 FK-Tranassis 구성

사용자에게 불-한 번역 도우미 FK-Tranassis 프로토타입은 기존의 자동기계번역기의 관점이 아닌 다른 관점에서의 번역방법을 제시한다. 자동으로 번역을 해주는 것이 아니라 전문가가 번역을 하기 위해서 필요

한 부수적인 기능들을 제공해주는 전문가를 위한 인터페이스이다.

이 FK-Tranassis 의 기능은 첫째 모든 정보를 인터넷 브라우저에서 편리한 인터페이스로 웹페이지를 이용하여 실시간으로 번역 작업을 할 수 있도록 하고, 둘째 번역 작업을 하는 동안 작업의 내용을 수정하고 삭제 할 수 있게 하며 맞춤 출력 기능을 제공하여 전문가가 원하는 형식으로 번역 내용을 출력 할 수 있게 한다. 셋째 문서내의 검색기능과 책갈피 기능과 같은 보조 기능을 제공하여 번역을 하는데 있어서 편리함을 제공한다. 넷째 불한 전자 사전과 별도로 자신에 맞게 구성할 수는 사전 기능이 제공되고 또한 전문 용어에 대한 사전을 제공하여 오역 없이 번역을 할 수 있게 하며, 다섯째 통계기반 방법과 어휘 빈도에 대한 시소러스 방법을 사용하여 기존 번역 사례를 참조문으로 제시해 불필요한 중복 번역을 피해 번역속도를 향상시킨다. 마지막으로 현재 사이트에 접속한 다른 번역 전문가에게 실시간으로 도움을 요청하는 서비스를 제공한다. 이와 같이 불-한 번역 도우미 FK-Tranassis 는 전문가에게 번역을 위한 보조 기능을 제공해 다량의 문서처리에 도움을 주는 웹상에서 서비스되는 인터페이스이다.



<그림 1 FK-Tranassis 프로토타입의 본문부분 >

위의 <그림 1>은 불-한 번역도우미 FK-Tranassis 프로토타입의 본문부분으로 ①원문부분과 ②사전부분 ③결과보기부분 ④검색창 부분 ⑤번역입력 및 참조부분 ⑥단어의 의미 숙어부분으로 구성된다.

원문부분은 문장이 번역단위로 각각의 문장은 링크 되어 있고 각 문장에는 정보를 담은 정보 탭을 두어 문단과 문장의 위치를 한눈에 파악이 가능하다. 또한, 다음 문단으로 넘어갈 때 번역한 앞 문단에 번역된 문장을 1/3 정도 함께 제시해줌으로써 번역의 단절감을 막을 수 있어 자연스러운 번역을 할 수 있도록 한다. 사전부분은 첫째 동철의어를 구분하기 위해서 정렬을 위한 단계를 따로 두어 정렬 표제를 설정하고, 둘째 표제어와 예문의 교차 검색 기능을 제공하여 등재

단계의 문제를 해결한다. 셋째 하나의 표제어에 파생되는 어휘를 내포하는 구조를 각각 독립된 표제어로 제시하고 마지막으로, 불규칙 굴절 변화형을 표제어로 등재하고 기본형 정보 및 굴절 정보를 제공하거나, Nlip 불한 형태소 분석기를 이용하여 기본형과 굴절형 간의 관계를 정의해 준다[7].

결과보기부분은 지금까지 해온 번역문의 결과를 보여주는 기능이다. 결과보기 유형에는 원문과 번역문을 함께 보여주는 형태와 번역문만을 보여주는 형태, 그리고 원문만 보여주는 형태 등 세 가지 유형을 제공한다. 원문과 함께 번역문을 보여주는 형태는 원문과 번역문을 한 문장씩 보여주어 직역과 의역에 대한 원문과 번역문을 비교 분석함으로써 편집이 용이하고 번역문만 보여주는 형태는 전체적인 번역결과가 원본의 내용에 맞게 원활하게 번역되었는지 확인할 수 있고, 원문만 보여주는 형태는 원본에 대한 내용이 대략 어떤 내용으로 번역되어있는지 확인할 수 있어 전문가가 필요한 상황에 따라 결과를 여러 유형으로 볼 수 있게 한다.

검색창 버튼은 원본과 번역된 결과문에 대하여 완전한 문장을 찾고자 할 때나 한 문장에 단어에 대하여 찾고자 할 경우 사용한다.

번역입력 및 참조부분은 전문가가 원문을 보고 번역을 입력하는 부분으로 원문과 원문을 분해한 분해문 그리고 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 참조문으로 제시한다.

3. FKT-DynaDic 사전

번역도우미 FK-Tranassis 프로토타입의 FKT-DynaDic 사전은 Translation Dic 사전과 Concept Dic 사전으로 구성된다. Concept Dic 사전은 개념어 사전이고 Translation Dic 사전은 대응어 사전이다. FKT-DynaDic 사전의 특징은 번역가의 번역 성향에 맞게 대역어를 제공해주고, 또한 대역어가 될 수 있는 확률이 높은 최적의 대역어를 우선순위로 제공해준다는 것이다. 사전의 논리적 구조를 설명하면서 FKT-DynaDic 사전의 특징에 대하여 알아 보도록 한다.

4. FKT-DynaDic 사전의 논리적 구조

FKT-DynaDic 사전의 사용에 있어서 기처리 번역 문장 일 경우와 미처리 번역 문장일 경우 분해문 DIC DATABASE 와 FKT-DynaDic 사전의 논리적 구조는 변한다. 분해문 DIC DATABASE 는 기처리 문장을 구성하는 기본 단어들의 정보를 수록하고 있는

DATABASE 이다. 직접적인 정보구성으로는 기본단어, 대역어, 기본 단어의 순서, 기본단어의 소속, 기본단어의 사전정보 위치가 제공된다. Concept Dic 사전은 번역을 위한 일반적이고 기초적인 기본 사전으로 특정 문장에 따른 전문용어 사전이나 개인 사전을 의미하는 것이 아니라 보편적이고 기본이 되는 어휘 사전으로 구성된다. Concept Dic 사전에서 제공되는 정보는 표제어, 품사, 개념어, count 가 있다. Concept Dic 사전 구성의 특징은 유일형태의 표제로 구성된 사전이 아니라 하나의 대응어를 추출하기 위해서는 표제어, 품사 개념이 조합되어야만 유일한 개념을 적용한 표제어를 선정한다. Concept Dic 사전에서 선정된 하나의 표제어가 다수의 개념을 나타낼 수 있는 경우 번역가의 성향에 따라 사용되는 개념은 달라지게 된다. Count 가 시스템을 사용하는 번역가의 기존 개념 설정 성향을 보관하는 목적으로 해당 개념의 사용 빈도를 Count 로 표시한다.

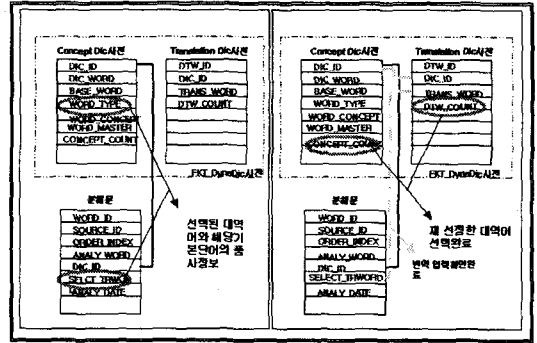
Translation Dic 사전은 번역가가 표제어에 맞는 문장을 번역을 위해 선택되는 대역어들로 구성된다. Translation Dic 사전은 최적의 대응어를 제시해준다. Translation Dic 사전에서 제공되는 정보는 대역어, 대역어의 소속, 대역어의 개념정보 그리고 count 가 있다. 번역가가 한 표제어에서 한 대응어에 대해서만 번역하는데 많이 사용하였을 경우 이는 다음에 같은 표제어가 나올 때 그 표제어에 대한 대응어가 될 수 있는 확률이 높다. 한 표제어에 대하여 최적의 대응어를 우선순위로 제공하기 위해서는 대역어의 사용 빈도를 Count 로 표시한다.

FKT-DynaDic 사전과 분해문 DIC DATABASE 의 논리적 구조가 어떠한 변화가 있는지 기처리 번역 문장일 경우와 미처리 번역 문장일 경우를 살펴 보도록 한다.

기처리 번역 문장일 경우 기본 단어들로 수록되어 있는 분해문 DIC DATABASE 의 정보구성 요소 중 번역에 사용된 대역어와 그 대역어의 사전 정보 위치 즉, Concept Dic 사전에서 제공해주는 품사의 정보를 알 수 있다. 이는 기 처리 된 문장의 대역어 정보가 무엇이며 또한 번역가가 문장 번역시 그만의 고유한 번역 스타일을 알 수 있는 정보가 된다. 다른 대역어를 재 선정하여 문장을 편집하고자 할 경우 Concept Dic 사건의 사용된 빈도가 높은 개념어 즉 count 가 높은 개념어에 대한 Translation Dic 사전에서 가장 많이 제시된 대역어 즉 count 가 높은 대역어를 제시해준다. 대역어를 선택하여 완료하면 Concept Dic 사건의 count 와 Translation Dic 사건의 count 의 논리적 구조가 변하므로 번역가의 성향에 따라 대역어 선택의 확률이 높아진다. 또한 번역가가 번역을 완료하였을 경우 분해문 DIC DATABASE 의 정보구성 요소중 선택된 대역어 저장부분과 그 대역어의 사전정보부분이 Update 되면서 논리적 구조가 변한다.

<그림 2>은 기처리 번역 문장일 경우 분해문 DIC

DATABASE 와 Concept Dic 사전과의 관계를 나타내는 다이어그램이다.

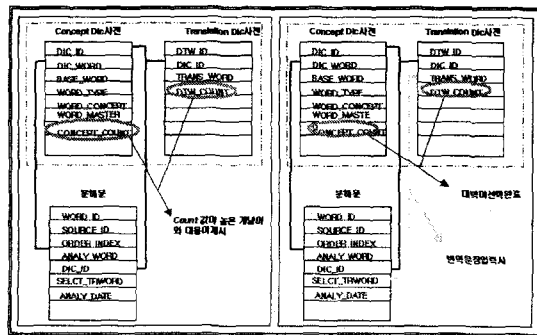


<그림 2 기처리 문장의 분해문 DIC DATABASE 와 FKT-DynaDic 사전과의 Relationship>

<그림 2>의 좌측 부분은 번역된 대역어와 대역어와 관련된 해당 원 문장의 기본단어의 품사 정보를 나타내는 그림이고, 우측 부분은 대역어를 재 선정 시 분해문 DIC DATABASE 와 FKT-DynaDic 사전과의 Relationship 을 나타낸 그림이다.

미처리 번역 문장일 경우는 nlip 형태소분석기가 문장을 분해하여 분해문 DIC DATABASE 에 어순에 맞게 기본 단어를 저장한다. 분해문 DIC DATABASE 의 기본단어를 Concept Dic 사전에서 검색하여 사용빈도가 높은 즉 count 가 높은 개념어에 대한 Translation Dic 사전에서 가장 많이 제시된 대역어 즉 count 가 높은 대역어를 먼저 제시해준다. 이는 제공되는 대역어가 번역가의 성향에 고려하여 먼저 제시해 준 것이다. 대역어를 선택하여 완료하면 Concept Dic 사건의 count 와 Translation Dic 사건의 count 의 논리적 구조가 변한다. 또한 번역가가 번역을 완료하였을 경우 분해문 DIC DATABASE 의 정보구성 요소 중 선택된 대역어 저장부분과 그 대역어의 사전정보부분이 Update 되면서 논리적 구조가 변한다.

<그림 3>은 미처리 번역 문장일 경우 분해문 DIC DATABASE 와 Concept Dic 사전과의 관계를 나타내는 다이어그램이다.



<그림 3 미처리 문장의 분해문 DIC DATABASE 와 FKT-DynaDic 사전과의 Relationship>

<그림 3>의 좌측 부분은 번역할 개념어에 대한 대역어를 제공해주는 그림이고, 우측 부분은 대역어를 대역어 선택 완료와 번역문장 입력완료시 DIC DATABASE 와 FKT-DynaDic 사전과의 Relationship 을 나타낸 그림이다.

기처리 번역 문장 일 경우 분해문 DIC DATABASE 와 FKT-DynaDic 사전의 각각의 논리적 구조와 연관성, 그리고 미처리 번역 문장일 경우 분해문 DIC DATABASE 와 FKT-DynaDic 사전의 각각의 논리적 구조와 연관성에 대하여 알아 보았고 이들 통하여 번역가에게 번역 스타일에 맞는 대역어를 제공해 주는 과정에 대하여 알아보았다.

5. 결론

본 논문에서 기존의 자동번역에서의 이론적 접근 방법이나 규칙들을 사용하여 번역기를 구현한 것이 아니라 번역가와 컴퓨터가 서로 상호 보완하여 빠르고 정확한 번역을 위하여 도움을 줄 수 있는 불-한 번역 도우미 FK-Tranassis 의 구성과 이 시스템의 핵심 모듈인 FKT-DynaDic 사전의 논리적 구조를 살펴 보았다.

불-한 번역 도우미 FK-Tranassis 프로토타입은 불어를 한국어로 번역하는 대량의 문서를 처리함에 있어서 전문가에게 원문을 분해와 기존 번역 사례 참조 최적 단어 추천 기능을 제공하는 사전을 제공해주어 빠르고 효율적인 번역 작업을 도와주기 위한 도구이다. 불-한 번역 도우미 FK-Tranassis 프로토타입에서 최적 단어 추천 기능을 제공하는 사전은 FKT-DynaDic 사전이다. FKT-DynaDic 사전은 Translation Dic 사전과 Concept Dic 사전으로 구성된다. Concept Dic 사전은 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 대응어를 추출하기 위해서는 표제어, 품사, 개념이 조합된 사전이다. Translation Dic 사전은 표제어 번역가가 문장에 맞는 번역을 위해 선택되는 대역어들로 구성된 사전이다. FKT-DynaDic 사전의 논리적 구조를 설명하면서 FKT-DynaDic 사전이 번역가의 성향에 맞게 대응어를 제시해주는 방법과 한 표제어에 대하여 대역어를 제시해주는 확률적인 방법에 대하여 설명하였다.

참고문헌

- [1] 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.
- [2] Gerald D., "Translation Memory: Concept, products, impact and prospects", Major project report South Bank

University School of Electrical, Electronic and Information Engineering, 1995.

- [3] Berrendonner, A. "Grammaire pour un analyseur: aspects morphologiques". Lyon: document de travail du groupe SYDO, 1984.
- [4] Seo, L. Contribution au traitement automatique du verbe francais: lexicque, morphologie et syntaxe. these en doctorat. Lyon: univresite-lumiere, 1995.
- [5] 서래원, 박세영, 유성준 "다국어 정보처리를 위한 불어 전처리에 관한 연구" 정보처리학회 Vo.17, No.4 1999.
- [6] 박주형, 이창우, 강명주 "자동번역과 CAT 의 현황과 전망" ㈜클릭큐 부설 기술 연구소
- [7] 윤애선 "표준화된 전자사전 개발을 위한 정보구조 및 사용자 환경 설계와 그 응용", 한국프랑스학회논문집, Vol.33, No.0, 2001
- [8] <http://www-rali.iro.umontreal.ca>
- [9] <http://kibs.kaist.ac.kr/experien/etdms72.htm>