

영한 기계번역기 트래니Trannie 96

성열원, 박치원, 정희선

인문정보연구회, The Institute of Human Information Research

1. 트래니Trannie 96의 작동 환경

10월 9일 영한 기계번역기 트래니Trannie 96가 발표되었다.

트래니Trannie 96은 Windows 95 환경에서 작동한다. 최소 설치 사양은 HDD 25MB, RAM 4M이며 원활한 작동을 위해서는 RAM 16M이상을 권장한다. 그리고 별도의 작업 공간은 필요하지 않다. 최적의 작동을 위해서는 빠른 속도의 CPU와 HDD가 필요하지만, 작동의 기본 환경으로는 어떤 IBM 호환기종이라도 관계없다.

2. 트래니Trannie 96의 특징 및 구조

2.1. 트래니Trannie 96의 특징

트래니Trannie 96 번역 엔진의 현저한 특징은 빠르고 효과적인 번역을 위해서 엔진을 통합적으로 구성해 효율성을 극대화했다는 점이다.

기존의 번역기에서는 이론과 엔진, 그리고 사전이 잘 맞아 들어 가게 통합적으로 구성되지 않는 경우가 많다. 이럴 경우에 번역기의 설계와 코딩을 담당하는 사람들이 각각 다르고 의사 전달이 원활히 이루어지지 않아 이상적으로 최적화된 설계가 이루어지기 어렵다. 또한 구현에 있어서도 설계의 의도가 제대로 반영될 수 없다.

트래니Trannie 96의 경우, 번역기를 구성할 때, 언어학 이론에 바탕한 각 부분의 엔진이 다른 엔진의 작동과 환경을 고려하여 유기적으로 설계하였기 때문에, 통합되었을 때 상호 데이터의 형식이나 전달 요소들이 어긋나지 않는다. 따라서 그 효율성이 극대화되는 것은 당연하다.

이러한 유기적 최적화는 '명확하고 세밀한 처리 단계의 구분', '언어 원리에 맞는 부분 엔진과 규칙의 구성', '사전, 규칙, 처리 및 각 부분 모듈 엔진들의 통합적인 구성' 등의 모습으로 나타난다.

2.2. 트래니Trannie 96의 구성

2.2.1. 전처리기(Pre-processor)

전처리기는 크게 문서 전처리기와 문장 전처리기로 나뉜다. 문서 전처리기는 문서 전체의 형식을 컴퓨터 내부에서 처리하기 쉬운 형식으로 표준화하는 것으로, 인용부호 등과 같이 한 문장을 넘어서는 영역에 걸쳐 나타나는 기호를 정보화하는 한편, 선, 그림, 도표 등 처리 대상이 아닌 정보들을 제거할 수 있도록 만들어 주는 엔진을 말한다.

한편 문장 전처리기에서는 여러 기호들과 약호들을 처리하여 정보화하고, 번역이 불가능한 형식의 문장을 가려내 프로그램의 오동작을 방지하는 작업을 수행한다.

2.2.2. 태거(Tagger)

트래니Trannie 96에서는 태깅 과정을 하나의 독립된 부분으로 설정하고 있다. 이 부분에서는 낱말들의 정보를 사전에서 뽑는 것 이외에도, 속어들의 정보화, 품사 애매성의 수치화 및 우선순위 결정 등, 직접적으로 다음 단계인 문장 단위 구문 분석에 필요한 형식으로 어절들을 정보화하는 작업을 수행한다.

2.2.3. 파서(Parser)

문장의 구조를 분석해 낸다. 속도와 정확성을 동시에 고려하여 설계되었고, 또한 변환과 생성에 관한 나무 구조를 생성해 내기 위해 최적화되었다. 일반적인 전산학 이론에서 찾아보기 힘든 새로운 기법들이 많이 사용되었다.

2.2.4. 변환기(Convertor)

일반적인 변환기와 달리 먼저 의미구조로 변환한 후, 다시 국어 나무 구조를 바꾸어내는 작업을 행한다. 일반적으로 파서에서 내보내는 나무 구조는 문장의 통사적 구조, 다시 말해, 주어와 목적어, 수식어와 피수식어 등의 정보만을 포함하고 있는 것에 비해, 의미구조는 행위, 행위자, 행위의 영향을 받는 것, 행위에 수반되는 다른 물질이나 행위와 같이 문장의 구조를 의미적인 관계로 표현한다. 이를 통해 몇 가지 의미처리 작업을 행할 수 있으며, 한결 자연스러운 국어 문장 구조를 만들어 낸다.

2.2.5. 생성기(Generator)

국어의 실사와 허사 정보 및 구조 정보를 조합하여 국어 문장을 생성해 낸다. 국어의 나무구조가 생성되면, 그 말단부에서 생성될 각 낱말의 원형과 그 쓰임에 따른 여러 허사 정보들이 나오게 된다. 예를 들어 I love the girl. 과 같은 문장에서 love는 타동사, 현재, 서술형, 직설법이라는 정보가 나오는데, 이를 [{"사랑하-"} : [현재], [서술], [직설]] 이라는 정보로 바꾸는 것이다. 생성기에서는 이 정보를 받아들여 '사랑한다' 와 같이 실제 문장에서 사용되는 꼴로 바꾼다. 트래니Trannie 96의 생성기는 대용량의 국어 사전과 정리된 처리 알고리즘을 통해 깨끗하게 국어 어절을 조합해 내며 경어체와 평어체의 선택 사양을 제공한다.

2.2.6. 사전(Dictionaries)

트래니Trannie 96에서는 여러 종류의 처리 사전(Dictionaries for Engines)을 사용한다. 먼저 기본적인 낱말 정보를 얻기 위해, 영어 낱말 사전을 사용한다. 이 외에도 영어 낱말의 통사, 의미적 정보를 담고 있는 특수 정보 사전을 여러 종류 사용하며, 각 낱말의 대역어를 담고 있는 대역 사전도 물론 사용한다. 또한 복합어, 속어를 처리하기 위한 사전도 사용하며, 전처리 등을 위해 약어 사전도 사용한다. 중의성 해소를 위한 의미 사전으로는 논항 사전이 사용되며, 몇 가지 세부적인 처리를 위한 부분 정보 사전이 엔진의 각 부분에서 사용된다. 또 국어 생성을 위한 국어 형태 사전이 사용된다.

2.3. 통합 구성

일반적인 번역기의 구성과 마찬가지로, 트래니Trannie 96도 위의 과정을 순차적으로

진행하여 번역을 행한다. 번역의 속도를 높이기 위해, 되도록 순환적인 구조를 없애려고 노력하였으며, 처리가 실패한 것으로 판단되는 경우, 최소한의 순환으로 다시 새로운 시도를 가능하게 구성하였다. 또한 각 부분의 엔진은 통합적인 설계 위에서 구성되어 엔진 사이에 빠진 부분이나, 겹친 부분이 없도록 하였다.

3. 기대 효과

트래니Trannie 96은 위에서 설명한 바와 같이 언어학에 바탕을 두고 여러 새로운 아이디어를 반영하였다. 이에 따라 신속성과 정확성의 목표를 동시에 이룰 수 있게 되었다.

<<ABSTRACT>>

English-Korean Machine Translator "Trannie 96"

The aim of this presentation is to show the structures and characteristics of English-Korean Machine Translator "Trannie 96".

"Trannie 96" consists of five main engines and various types of dictionaries. With respect to the engines, the English sentences filtered by Pre-processor are tagged and parsed. After the conversion from English sentence structure to Korean one, "Trannie 96" constructs Korean sentence. As for dictionaries, each engine has more than one optimized dictionaries.

The algorithms employed by this machine is based on Linguistic theories, which make it possible for us to produce speedy and accurate translation.