

HS-eLearner: 영어 학습과 영한번역 시스템 평가를 위한 도구

김성동, 박성훈
한성대학교 컴퓨터공학과
e-mail : sdkim@hansung.ac.kr, garion01@hansung.ac.kr

HS-eLearner: A Tool for English Learning and Evaluation of English-Korean Machine Translation System

Sung-Dong Kim, Sung-Hoon Park
Dept. of Computer Engineering, Hansung University

요 약

본 논문에서는 영어 학습을 도와주고 영한 기계번역 시스템의 평가를 통해 기계번역 시스템의 성능 개선을 위한 도구인 HS-eLearner 를 설계, 구현하였다. HS-eLearner 는 영어 문장의 번역 뿐만 아니라 입력 문장의 구조, 문장에 사용된 단어의 의미를 제공하여 사용자의 효과적인 영어 학습을 보조하는 기능을 가진다. 또한 사용자가 번역된 문장을 평가하여 사용자에게 의한 객관적인 번역 시스템에 대한 평가를 기대할 수 있으며 평가 결과를 개발자에게 제공함으로써 번역 시스템의 성능 개선에 사용될 수 있다. 즉 사용자와 개발자간의 커뮤니케이션을 제공함으로써 사용자의 요구를 수용할 수 있는 시스템으로의 개선을 용이하게 한다.

1. 서론

현재 여러 가지 영한 기계번역 시스템이 상용화 되어 판매되고 있다 [1]. 상용 시스템은 영어 문장의 대화식 번역, 문서의 자동 번역, 인터넷 웹사이트 번역 등의 기능을 제공하고 있다. 또한 Microsoft 사의 Word, Excel, PowerPoint 등의 문서와 Outlook Express 상의 e-mail 도 번역하는 기능을 제공함으로써 사용자에게 여러 가지 편리를 제공하고 있다. 그러나 번역률이 저조하여 널리 사용되지 못하고 있는 상황이다. 그러나 비교적 짧고 그 구조가 분명한 문장들에 대해서는 좋은 번역을 생성한다. 이를 활용한 초등학교, 중학교, 고등학교 학생들의 영어 공부를 보조할 수 있는 도구가 있다면 학생들이 영어를 보다 재미있고 효과적으로 공부할 수 있을 것이다.

영한 기계번역 시스템을 개발하는 과정에서 많은 문장을 수집하여 번역 테스트를 수행한다. 테스트를 위한 문장은 주로 개발자 측에서 수집하며 번역 테스트를 통해 번역 시스템의 성능을 향상시킨다. 개발자

측에 의한 테스트 문장의 수집과 시스템 개선은 다양한 분야, 다양한 형식의 문장에 대한 테스트를 제한하기 때문에 번역기의 실제적인 성능 향상을 기대하기 어렵다고 할 수 있다. 많은 사용자에게 의한 번역 시스템의 평가를 통해 번역 시스템의 부족한 부분을 보다 객관적으로 판단할 수 있으며 사용자의 커뮤니케이션을 통해 사용자의 번역 테스트 과정에서 사용한 다양한 분야, 다양한 형식의 테스트 문장을 확보할 수 있게 된다. 이를 통해 사용자의 요구를 만족하는 방향으로 번역 시스템을 개선할 수 있게 된다.

본 논문에서는 위에서 제시한 두 가지 목적을 달성하기 위하여 HS-eLearner(HanSung-English-Learner)라는 영어 학습 및 영한 번역 시스템 평가 도구를 설계, 구현하였다. 영한 번역 시스템은 (주)엘앤티(Language & Technology)에서 제공받았으며 사용자 인터페이스(graphical user interface)는 MFC(Microsoft Foundation Class)를 이용하여 HS-eLearner 를 구현하였다.

2 장에서는 HS-eLearner 의 구조와 각 부분의 기능을 설명하고 3 장에서 HS-eLearner 를 이용의 기대 효과를

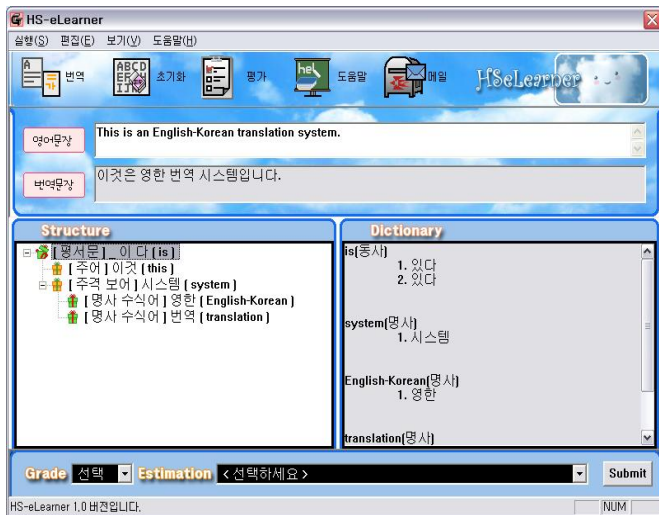
분석하였으며 4 장에서 본 논문을 결론짓는다.

2. HS-eLearner 의 구조와 기능

본 절에서는 HS-eLearner 의 구조와 각 부분의 기능을 상세히 설명한다.

2-1. HS-eLearner 의 구조

HS-eLearner 는 영어 문장을 번역하여 문장 구조와 문장에 사용된 단어의 뜻과 함께 번역문을 제공함으로써 사용자의 영어 학습을 보조할 뿐만 아니라 사용자가 번역 결과를 평가하고 그 결과를 번역 시스템 개발자에게 쉽게 전달할 수 있는 도구를 제공한다. 그림 1 은 HS-eLearner 의 모습을 보여준다.



(그림 1) HS-eLearner 의 모습.

2-2. HS-eLearner 의 기능

HS-eLearner 는 학습을 보조하는 기능과 번역 시스템을 평가하는 두 가지 기능을 가진다. 본 절에서는 그림 1 을 이용하여 HS-eLearner 의 각 부분의 기능을 설명한다.

HS-eLearner 상단에는 여러 가지 버튼을 배치하여 쉽게 사용자가 HS-eLearner 를 사용하도록 하였다. **번역**, **초기화**, **평가**, **도움말**, **메일** 버튼. **번역** 버튼을 클릭하면 입력한 영어 문장을 번역하여 번역 결과를 보여주며 동시에 왼쪽 아래쪽 창에 입력 영어 문장의 구문 구조(syntactic structure)를 보여주고, 그리고 오른쪽 아래 창에는 문장에 사용된 단어가 영한 번역 사전에서 가지는 모든 의미를 나열하여 준다. **초기화** 버튼은 새로운 문장을 번역하려고 할 때 현재 창에 나타나 있는 모든 내용을 지우는 기능을 한다. 물론 초기화를 하지 않고 현재 영어 문장을 간단하게 편집하여 다시 번역을 시도해도 된다. **평가** 버튼은 HS-eLearner 아래쪽 부분에 평가를 위한 리스트 박스가 나타나도록 한다. **메일** 버튼은 저장된 평가 결과를 e-mail 을 사용하여 번역 시스템 개발자에게 보내는 기능을 가진다.

영어 문장 박스에는 번역하고자 하는 영어 문장을 입력한다. **번역 문장** 박스에는 번역된 한국어 문장이 출력된다. 왼쪽의 **'Structure'** 창에는 영어 문장의 구문 구조가 트리 형식으로 보여지는데 처음에는 '평서문', '의문문' 등의 하나의 노드(node)가 나타나며 '+'나 해당 노드를 클릭하면 그 아래의 하위 노드가 펼쳐지면서 나타난다. 사용자가 문장의 트리 구조를 직접 다룰 수 있게 함으로써 학습 효과를 향상시킬 수 있을 것으로 기대된다. 오른쪽의 **'Dictionary'** 창에는 영어 입력 문장에 사용된 모든 단어에 대해 각 단어가 사전에서 가지는 모든 의미를 제시해준다. 문장의 번역에 사용된 의미 뿐만 아니라 다른 의미도 알려줌으로써 사용자 영어 단어를 학습 하는데 도움을 줄 수 있을 것이다. 아래 부분에서는 번역 결과에 대해서 A, B, C, D 등급으로 평가를 할 수 있게 하였으며 A 등급이 아닌 경우 잘못된 번역에 대해 평가를 할 수 있도록 하였다. 여기서는 잘못된 번역으로 다음 6 가지를 제시하였다.

- 잘못된문장 구조
- 대역어 선정의 오류
- 전치사 번역의 어색함
- 한국어 어순이 잘못됨
- 부분적으로 해석이 빠진 부분이 있음
- 전체적으로 이해할 수 없음

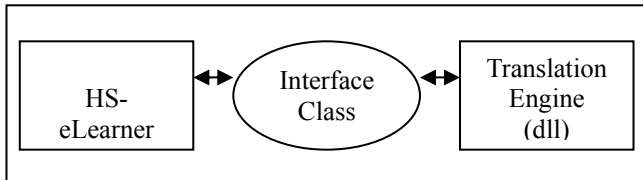
사용자가 평가한 결과에 따라 영어 문장과 번역 문장을 따로 일정한 곳에 저장하도록 하였다. 즉 6 가지 원인별로 입력 문장과 출력 문장을 서로 다른 파일에 저장하도록 하였다. 이 기능은 영한 번역 시스템의 오류를 수정에 사용할 수 있는 데이터를 원인별로 분류하고 구축할 수 있도록 한다. 구축된 데이터는 후에 영한 번역 시스템 개발자에게 피드백 됨으로써 번역 시스템의 개선에 소중한 자료로 활용될 수 있을 것이다. 또한 A 등급에 해당하는 문장도 저장함으로써 현재 번역 시스템이 좋은 성능을 보이는 문장도 수집할 수 있게 하였다. 이 데이터는 시스템의 개선이 이루어지는 과정에서 시스템의 검증(validation)에 필요한 문장으로 활용될 수 있을 것이다.

2-3 번역 엔진과의 인터페이스

HS-eLearner 의 핵심이 되는 번역 엔진은 (주) 엘앤티 (Language & Technology)에서 개발한 것으로 자연스러운 번역을 위한 속어 번역 방식[2, 3]을 택하고 있으며 긴 문장의 효율적인 번역을 위해 문장 분할 방식[4, 5]을 적용하여 개발 되었다. 영어 문장의 구문 분석의 표현을 위해 predicate-argument 구조[6]를 이용하므로 HS-eLearner 는 이 구조를 사용자에게 보여준다.

영한 번역 시스템은 dll(dynamic linking library) 형식으로 제공되며 HS-eLearner 는 번역 시스템에 영어 문장을 제공하고 번역된 한국어 문장 뿐만 아니라 영어 문장의 구문 구조를 나타내는 트리 구조, 영어 단어의 사전 내용 등을 제공받는다. 또한 코드 변환 등 번역 엔진 내부에서 사용하는 함수도 HS-eLearner 에서 사

용할 수 있어야 한다. HS-eLearner 는 번역 엔진과의 인터페이스를 위해서 하나의 인터페이스 클래스(interface class)를 이용한다. 인터페이스 클래스는 번역 엔진의 초기화와 번역 엔진 내부 자료 접근과 함수 호출을 위한 기능을 한다. 그림 2 는 HS-eLearner 와 번역 엔진과의 인터페이스를 보여준다.



(그림 2) HS-eLearner 와 번역 엔진간의 인터페이스

응용 프로그램과 번역 엔진 사이에 인터페이스 클래스를 둬으로써 응용 프로그램과 번역 엔진간의 독립성을 보장한다. 즉 응용 프로그램의 요구를 만족시키기 위해 번역 엔진을 수정하는 대신 인터페이스 클래스의 기능을 추가하거나 수정하면 된다. 이는 번역 엔진을 다른 응용에 적용하는 것을 용이하게 해줄 수 있다.

3. HS-eLearner 의 기대 효과

HS-eLearner 는 학습 도구로서의 역할과 영한 번역 시스템 평가를 위한 도구로서의 역할을 가지는 두 가지 목적으로 개발되었다. 이 절에서는 HS-eLearner 를 활용함으로써 기대할 수 있는 효과를 설명한다.

첫째로 HS-eLearner 는 학습 도구로서의 역할을 수행한다. 학습 도구란 관련된 특정 주제를 학습하는데 도움을 주는 도구를 지칭한다. 그것은 일반적인 학습 방법을 개선하거나 효과를 증진하는 것이 그 역할의 중심이 된다. HS-eLearner 의 영어 학습 도구로서의 역할은 단순히 문장의 의미만을 출력하는 것이 아니라, 문장 구조, 어법, 단어 의미를 출력하여 사용자에게 언어적 배경 지식을 학습하는 데에 도움을 주며 이는 이러한 데이터를 얻기 위해 여러 문법책, 어휘 사전을 통해 이러한 지식들을 얻는 기존의 학습 방법보다 커다란 사용자 편의를 제공한다. 영어 학습 도구는 사용자가 영어를 학습하는 데에 최대한의 편의와 정보를 제공하여 사용자의 학습 능력을 고취 시키고 학습 효율을 늘리는 데에 그 궁극적인 역할과 목표가 있다.

둘째로 HS-eLearner 는 영한 번역 시스템의 개발자로 하여금 현재 시스템에서 개선될 부분을 확인할 수 있도록 하는 데에 그 역할이 있다. 단순히 영문 번역을 요구하는 사용자의 만족도를 측정하는 것이 아니라, 영한 번역 시스템의 현재 수준을 시스템 개발자가 아닌 사용자에게 의한 객관적 평가가 이루어지게 된다. 또한 현재 시스템이 높은 만족도를 주는 부분은 어느 부분이며, 상대적으로 만족도가 적은 부분은 어느 부분인지를 측정, 분석할 수 있는 도구의 역할을 한다. 또한 만족도가 낮으며, 개선이 필요한 부분에 대해서

는 입력과 출력 데이터, 연관되어 있는 번역 시스템의 알고리즘과 논리(logic)에 관련된 로그(log)를 작성하여 개발자가 번역 시스템의 개선에 쉽게 참고 할 수 있는 데이터를 생성 할 수 있는 역할도 하게 된다. 궁극적으로 HS-eLearner 는 개발자가 번역 시스템의 알고리즘과 코드의 개선에 도움을 주어 번역 시스템의 성능 향상에 필요한 도구의 역할을 할 것으로 기대된다.

4. 결론

본 연구에서 개발한 HS-eLearner 는 사용자의 학습 효과를 향상시켜 주고 동시에 사용자와 영한 번역 시스템 개발자간의 커뮤니케이션을 용이하게 해준다. 기존의 번역 시스템은 개발자가 일방적으로 제공할 수밖에 없는데 HS-eLearner 는 이러한 단점을 보완함으로써 사용자의 요구를 빨리 수용할 수 있다. 일반적으로 기계번역 시스템은 개발이 완료되면 계속 사용될 수 있는 것이 아니며 계속적인 보완이 필요하다. 이러한 시스템의 개선, 보완 과정이 사용자와의 원활한 커뮤니케이션에 바탕을 둔다면 사용자를 만족시킬 수 있는 시스템으로 보다 빨리 개선될 수 있을 것이다. 이는 개발자에게는 더 좋은 성능을 가지는 시스템 개발을 할 수 있도록 해주며, 시스템의 개선점이 어떤 부분에 있는지를 축적된 데이터를 통해 알 수 있으므로 개발의 편의성 또한 제공해준다. 사용자는 시스템의 지속적이고 효과적인 개선을 통해서 학습의 편의를 제공받는 동시에 더욱 효율적이고 더욱 개선된 학습 내용을 얻게 되는 것을 기대할 수 있다.

결론적인 가장 큰 기대효과는 처음부터 완전한 것을 만들어 낼 수 없는 영한 기계번역 시스템을 개발하고자 할 때, 사용자와의 커뮤니케이션을 극대화하고 그것을 데이터화 함으로써 얻어지는 영한 번역 시스템의 점진적 발전 가능성이다. 이는 단순히 개발과정의 반복이 아닌 사용과 평가의 반복을 통해 더욱 사용자의 요구에 부응하는 영한 번역 시스템을 개발할 수 있다는 가능성을 창출한다는 것과, 개발자로서 스스로 알지 못한 부분에 대해 개선점을 더욱 쉽게 찾을 수 있다는 점이 가장 크게 기대되는 부분이다.

참고문헌

- [1] Yu Seop Kim, Sung-Dong Kim, Seong Bae Park, Jong Woo Lee, Jeong Ho Chang, Kyu Baek Hwang, Jang Min O, Yung Taek Kim, "Machine Translation Systems: E-K, K-E, J-K, K-J," In proceedings of Envisioning Machine Translation in the Information Future: 4th Conference of the Association for Machine Translation in the Americas, AMTA 2000, Cuernavaca, Mexico, October 10-14, 2000.
- [2] Sung Hee Yoon, "Efficient Parser to Find Bilingual Idiomatic Expressions for English-Korean Machine Translation," In Proceedings of the 1994 ICCPOL, pp. 455-460, 1994
- [3] Ho Suk Lee, "Automatic Construction of Transfer Dictionary based on the Corpus for English-Korean Machine Translation," Ph.D. thesis, Seoul National

University, In Korean, 1993.

- [4] Sung-Dong Kim, Byoung-Tak Zhang, Yung Taek Kim, "Reducing Parsing Complexity by Intra-Sentence Segmentation based on Maximum Entropy Model," In Proceedings of Empirical Method in Natural Language Processing, pp. 164-171, 2000.
- [5] Sung-Dong Kim, Byoung-Tak Zhang, Yung Taek Kim, "Learning-based Intrasentence Segmentation for Efficient Translation of Long Sentences," Machine Translation, Volume 16, Number 3, pp. 151-174, 2001.
- [6] James R Hurford, "The Neural Basis of Predicate-Argument Structure," In Behavioral and Brain Sciences, Volume 23, Number 6, 2003.