

관용적 표현의 대응 관계에 기반한 영어-한국어 기계 번역

윤 성 희, 김 영 택
서울대학교 컴퓨터공학과

English-Korean Machine Translation based-on Bilingual Relation of Idiomatic Expressions

Yoon, Sung Hee Kim, Yung Taek
Department of Computer Engineering, Seoul National University

요 약

영어 문장을 한국어 문장으로 기계 번역하는 과정에는 분석 규칙이나 변환 규칙만으로는 해결하기 어려운 표현의 대응 관계들이 많이 나타난다. 본 논문은 영어-한국어 기계 번역에서 질적으로 향상된 한국어 문장을 얻기 위하여 두 언어 표현들 사이의 관용적 대응 관계에 기반하는 번역 방식을 논한다. 두 언어 표현들 사이의 다양한 직접 대응 관계를 제공하는 번역 사전을 이용하며, 입력 영어 문장으로부터 이와 같은 표현들을 인식하고 한국어 표현으로 직접적으로 대응시키는 번역 방식이다. 이러한 번역 방식은 기존의 변환 규칙 기반의 번역 방식보다 자연스러운 한국어 문장을 생성할 뿐만 아니라, 많은 구조적-의미적 모호성을 해결함으로써 시간적-공간적 처리 효율을 크게 높일 수 있다.

1. 서론

기계 번역은 컴퓨터를 이용하여 두 자연 언어 사이의 번역의 자동화를 실현하고자 하는 전산학적 연구이다. 특히, 기계 번역은 다른 자연 언어 처리 과정과는 달리 서로 다른 두 자연 언어가 동시에 관계되고 있으며, 두 대상 언어에 대한 어휘적, 구조적, 의미적 지식 등의 복합적인 지식이 컴퓨터에 의해 처리되어야 한다.

본 연구의 동기는 무엇보다도 기계 번역의 질적 향상이다. 질적으로 우수한 기계 번역은 입력 문장의 뜻을 정확하게 전달하는 자연스러운 출력 언어 문장을 생성해야 한다. 기계 번역이 갖는 가장 큰 문제는 두 자연 언어 사이에는 규칙으로 처리되기에는 매우 어려운 표현 및 어휘의 대응 관계가 있다는 것이다. 기존의 기계 번역 시스템들은 주로 다수의 구조 변환 규칙에 기반한 복잡한 변환 과정을 통하여 입력 문장 구조를 출력 문장 구조와 관련시키는 방법에 의존하여 왔다.

그러나, 이와 같은 변환 규칙에 기반한 기존의 기계 번역 방식은 영어와 한국어처럼 어휘적으로나 구조적으로 큰 차이를 갖는 언어 사이에서는 번역 결과에 있어서 질적 한계를 보여왔다. 계층적 구조나 어순에 크게 의존하고 있는 영어에 비해 한국어는 조사나 어미 등의 기능에 크게 의존하고 있다. 따라서, 영어 문장을 한국어 문장으로 번역하고자 할 때는 영어 어휘의 쓰임새나 문장의 다양한 패턴에 대한 분석으로부터 한국어에서의 적절한 조사, 어미, 어순 등을 결정하는 매우 불규칙적이며 어려운 작업이 요구된다. 뿐만 아니라, 계열이 크게 다른 영어와 한국어의 많은 어휘들은 번역되는 단어의 갯수에 있어서도 일대일 대응을 이루지 못하는 경우가 많으며, 일대다(예: ("snore" "코를 골다") ("tall" "키가 크다")), 다대일(예: ("take a walk" "산책하다") ("shake hands" "악수하다")), 다대다(예: ("do one's best" "최선을 다하다") ("have a headache" "머리가 아프다")) 등의 다양하고도 불규칙적인 대응 관계를 갖는 현상을 쉽게 볼 수 있다.

본 연구의 접근 방법은 입력되는 영어 문장을 한국어 문장으로 번역하고자 할 때, 입력 문장 자체의 구조적 분석보다는 출력 언어인 한국어 문장으로의 자연스러운 표현을 얻는, 즉, 관용적 대응 관계에 중점을 두는 것이다. 기존의 변환 기법의 번역 방식이 규칙에 의존한 반면, 이러한 번역 방식은 규칙으로 해결되기 어려운 실제의 언어 현상들에 기반하여 보다 자연스러운 번역을 가능하게 한다. 본 논문에서는 이와 같은 입출력 대응 표현들을 기계 번역을 위한 '관용적 표현(bilingual idiomatic expressions)', 또는 숙어적 표현이라고 정의한다.

본 논문에서의 기계 번역을 위한 관용적 대응 표현들이란 흔히 불리는 일반적 의미의 숙어보다는 범위가 훨씬 넓어서 영어에서의 전통적 은유 표현(예: "kick the bucket")들은 물론 반복적 구조 변환으로는 출력 언어로의 적절한 번역이 어려운 모든 표현(예: "rid ~ of ~", "prevent ~ from ~ing")의 대응 관계를 말한다. 영어-한국어 기계 번역의 경우를 볼 때, 이러한 확장된 범주의 관용적 대응 표현들은 영어에서의 일반적인 관용구는 물론 특히 한국어의 구조에는 존재하지 않아 번역을 위한 명백한 대응 구조를 얻기 어려운 영어의 부정사구, 분사구, 전치사들의 다양한 용법의 많은 부분에 대해 기본적인 어려움을 쉽게 해결하여 정확한 번역을 가능하게 한다[11]. 한국어 문장의 생성에서 문장의 많은 부분들이 관용적 표현의 직접 대응에 의해 조사, 어미, 선어말 어미 등이 매우 쉽게 결정되며, 어순도 쉽게 결정될 수 있다. 기존의 몇몇 연구들을 살펴볼 때, 자연 언어에서의 숙어적 표현들을 고려한 예가 있었으나, 이들 연구에서는 입력 언어의 은유적 표현에 국한되어 있다[1,7,8].

이와 같은 관용적 표현의 대응에 기반하여 번역 사전이 구성되며, 입력 문장의 분석 과정이나 번역 과정은 번역 사전이 제공하는 한국어 표현으로의 직접 대응 과정을 중심으로 이루어진다. 특히 입력 문장으로부터 우선적으로 관용적 표현들이 인식·번역된 뒤, 이들을 중심 단위로 문장의 전체적 구조를 완성한다. 즉, 일반적 경우처럼 입력 영어 단어 사이의 계층 구조의 파악보다는, 먼저 한국어로의 번역을 위한 대응 관용 표현들에 대해 그 요소들을 평면적으로 인식하는 것에 중점을 두고 있다.

기계 번역을 위한 이와 같은 접근 방식을 제안하는 이유를 번역의 질적 측면과 처리 효율의 측면에서 다음과 같이 말할 수 있다.

- (1) 규칙 기반의 기계 번역에서는 입력 문장의 분석에 성공하더라도 출력 언어 구조로의 변환을 위해 복잡한 변환 규칙과 변환 과정을 요구함에도 불구하고 만족스러운 번역 결과를 얻기가 매우 어렵다.
- (2) 많은 처리 시간과 공간을 소비해서 얻게 되는 분석 구조가 문장의 의미와는 무관한 경

우가 많다. 예를 들어, 영어에서의 전형적 은유 표현의 예가 될 수 있는 "kick the bucket"의 경우, 이 구절에 대한 구조적 분석 [VP "kick" [NP "the" "bucket"]]은 "물통을 차다" 대신 그 실제 의미인 "죽다" 라는 한국어 표현으로 옮겨 번역되기 위한 아무런 의미도 갖지 못한다. "take the bulls by the horns" 나 "take care of ~" 도 같은 유형의 표현들이다.

(3) 영어 표현의 문법적 정보가 번역될 적절한 한국어 표현의 문법적 구조와 크게 다른 경우가 많다. 예를 들어, "prevent ~ from ~ing"라는 영어의 표현에 대한 문법적 정보는 "~가 ~하지 못하게 하다"라는 한국어의 부정 표현을 얻기 위한 정보를 주지 못한다. 또한, "be suprised at ~"의 경우에는 영어의 구조로서는 수동태의 표현이지만, 한국어 표현으로는 "~에 놀라다"라는 자동사의 표현으로 번역되어야 한다.

(4) 영어에서 문법적으로 동일한 구조로 보이는 여러 표현들이 한국어에서는 각각 다양한 표현들로 바뀐다. 예를 들어, "allow ~ to~", "persuade ~ to~", "promise ~ to~", "enable ~ to~" 등은 영어에서는 동일한 구조를 갖지만 한국어로 번역될 때는 "~가 ~하는 것을 허락하다", "~를 ~하도록 설득하다", "~에게 ~할 것을 약속하다", "~가 ~할 수 있게 하다"등의 다양한 조사나 어미 등이 적절히 선택되어야 한다.

(5) 입력 문장의 분석 과정을 매우 비효율적이게 하는 주된 원인은 모호성의 많은 부분이 번역의 입장에서는 모호하지 않다. 예를 들어, "prevent ~ from ~ing"와 같은 동사의 사용 패턴도 [VP "prevent"[NP ~] [PP "from" ~ing]]와 [VP "prevent" [NP ~ [PP "from" ~ing]]]의 두 가지 구조로 분석될 수 있는 모호성을 가지고 있으며 구조 분석 후에도 그 조사나 어미의 적절한 선택이 매우 어려운 경우로서 관용적 대응 관계로 정의될 수 있다. 관용적 대응 관계로 정의된 후의 이 표현은 ["prevent" ~ "from" ~ing]로서 모호성 없는 유일한 평면 구조로 인식되고 "~가 ~ing 하지 못하게 하다"라는 한국어 표현으로 번역될 수 있다. 이와 같이, 관용적 표현의 직접 대응에 의한 모호성의 해결은 문장 분석은 물론 번역 과정의 효율을 크게 높여 준다.

본 논문은 관용적 표현의 대응 관계에 기반한 기계 번역 방식에 대한 두 가지 측면의 실험과 그 결과 자료를 제시하고 있다. 우선, 입력 문장의 분석 측면에서 관용적 표현을 우선적으로 추출하는 방법은 문법 규칙에만 의존하는 기존의 일반적 기법과 비교할 때 처리면에서 얼마나 효율적인가에 대한 실험과 그 결과를 제시한다. 또한, 이와 같은 관용적 표현의 대응에 기반하는 번역 방식에 의해 번역의 결과가 질적으로 향상될 수 있음에 대한 실험과 그 결과를 제시한다.

2. 영어-한국어 번역을 위한 관용적 대응 표현들

본 논문에서는 기계 번역을 위해 두 언어사이의 관용적 표현의 직접 대응 관계를 도입하고 있다. 기계 번역의 과정에서 반복적 구조 변환과 어휘 대치에 의해 적절한 번역구를 얻기 어려운 입력 언어의 표현들과 그 번역 표현의 모든 대응 관계가 곧 관용적 직접 대응 관계들로 정의된다. 영어-한국어 기계 번역을 위한 관용적 대응 표현들을 다음과 같이 다섯가지 범주로 분류할 수 있다.

- (1) 입력 언어에서의 전통적 은유 표현에 대한 한국어 의미들 (예) ("kick the bucket" "죽다") ("take the bull by the horns" "용감히 난국에 맞서다")
- (2) 복합 단어들 (예) ("popular election" "보통 선거") ("general hospital" "종합 병원")
- (3) 영어의 빈번한 사용 패턴으로서 한국어 표현으로의 기능어 제공에 유용한 표현들 (예) ("promise ~ to ~" "~에게 ~할 것을 약속하다") ("prevent ~ from ~ing" "~가 ~ing하지 못하게 하다")
- (4) 한국어 문장과 문법적 구조가 전혀 다른 영어의 표현들 (예) ("catch a cold" "감기에 걸리다") ("take a walk" "산책하다") ("be suprised at ~", "~에 놀라다")
- (5) 영어의 구어적 관용구들에 대한 한국어 표현 (예) ("Excuse me." "죄송합니다.") ("How do you do?" "처음 뵙겠습니다.")

관용적 직접 대응 관계에 의해 얻게 되는 한국어 표현은 영어 구조에 대한 구조 변환이나 어휘 대체 방식의 번역이 적절하지 못한 경우에 번역 문장의 골격을 제공하는 것이 그 기능이다. 그 밖의 번역 과정으로서 관용 표현이외의 각 단어들에 대해 가장 적절한 한국어 어휘를 선택하는 과정이 요구된다. 즉, 영어에서의 한 단어는 항상 특정의 한국어 단어로 대응되는 것이 아니라, 사용되는 경우에 따라 다수의 번역 어휘 중의 하나로 결정되어야 한다. 여기서, 사용되는 경우만 문장내에서 가장 깊은 관련을 갖는 관계어에 따라 함께 그 의미나 번역어가 결정되는 현상으로 해석될 수 있으며, 흔히 연어(collocation) 현상이라고 한다. 본 논문에서는 영어-한국어 번역에서 최적의 번역어 선택을 위해 다음과 같은 세 가지 연어 관계를 활용한다[3].

- (1) 타동사와 그 목적어
("reach" + "object" = "달성하다" + "목적")
("reach" + "branch" = "뺏다" + "가지")
- (2) 자동사와 그 주어
("river" + "run" = "강" + "흐르다")
("watch" + "run" = "시계" + "돌아가다")
- (3) 형용사와 수식하고 있는 명사
("good" + "time" = "즐거운" + "시간")
("good" + "egg" = "신선한" + "달걀")

번역 사전의 각 단어는 관용적 대응 표현 및 연어와 함께 디폴트 번역어를 갖는다. 디폴트 번역어는 문장의 연어쌍이 사전에 등록되지 않은 경우 그 어휘를 번역하기에 가장 일반적이라고 판단되는 출력 언어의 단어이다. 번역 사전의 모든 어휘는 그 디폴트 번역 어휘를 갖는다 (예: ("reach" = "도달하다") ("run" = "달리다") ("good" = "좋은")).

본 번역 방식은 번역을 위한 많은 지식을 관용적 표현의 직접 대응 관계를 제공하는 번역 사전으로부터 얻는다. 즉, 번역 사전이 주된 지식베이스이다. 따라서, 보다 많은 관용적 대응 표현들과 연어적 어휘의 쌍을 수집하는 것이 번역 결과의 질을 결정하는 중요한 작업이며, 이

러한 번역 사전의 구축 방법이 제시되어야 한다. 본 연구를 위한 번역 사전은 다음의 그림1에서 보이는 바와 같이 세가지 지식 추출 채널을 통해 구성된다. 첫째로, 영어-한국어 번역을 위한 관용적 대응 관계는 기존 출판된 영한 사전에 수록된 속어 표현들과 예문들을 이용하여 얻을 수 있다. 둘째로, 일반 영어 예문들을 통해서 영어 어휘들의 빈번한 전형적 사용 형태를 수집하며, 그 번역 표현과 함께 사전에 수록할 수 있다. 셋째로, 이와 같이 구축된 초기 사전을 이용하여 다수의 예문을 일차 번역한 뒤, 그 번역 결과로부터 의미적으로 틀리거나 어색한 표현들을 추출하고 새로운 정보를 사전에 수록한다. 이와 같은 사전 정보의 수집과 다수의 실제 예문을 이용한 실험을 통하여 관용적 표현의 직접 대응 정보를 갖는 번역 사전은 보다 안정된 상태로 수렴되어 갈 수 있다.

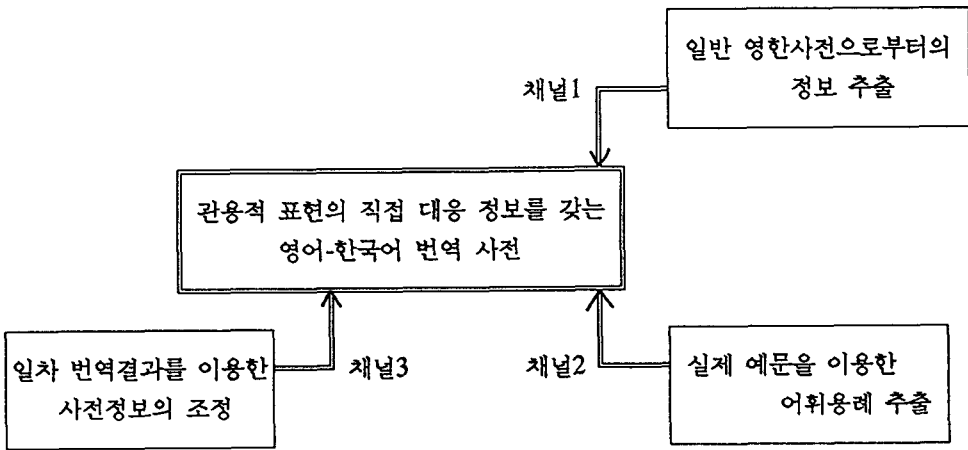


그림1. 영어-한국어 기계 번역을 위한 번역 사전을 위한 지식 채널

3. 관용적 표현의 직접 대응에 기반한 영어-한국어 기계 번역

관용적 표현의 대응 관계에 기반한 기계 번역 시스템은 처리 단계별로 (a) 입력 문장의 어휘 분석 (b) 입력 문장내의 관용적 표현들의 인식과 번역 표현의 직접 대응 (c) 문장의 전역 구조 완성(global structuring) (d) 연어 및 디플트에 의한 번역어 결정 (e) 한국어 문장 생성 과정으로 이루어진다. 번역 시스템을 위한 세 가지의 사전으로 영어 어휘 사전, 영어-한국어 번역 사전, 한국어 생성 사전이 요구된다. 그림2는 관용적 표현의 직접 대응 관계에 기반한 영어-한국어 기계 번역 시스템의 전체 구성을 보인다.

입력 문장은 관용적 대응 표현의 인식 과정에 이어서 전역 구조 완성 단계를 통하여 분석된다. 관용 표현의 인식 과정에서 인식되는 표현에 대해서는 대응되는 한국어 표현과 함께 단위 번역 구조를 만들고, 이 구조를 테이블(차트) 내에 하나의 완성된 부분 구조로 추가한다 [10]. 관용 표현의 인식을 위해서는 단어의 비교와 함께 표현의 변수적 요소를 얻기 위해 문장의 특정 지점에서 지역적으로 규칙이 적용될 수 있다. 전역 구조 완성 과정은 이미 인식되고 번역된 관용적 표현의 구조들을 중심으로 입력 문장의 전체 구조를 얻는 과정이다.

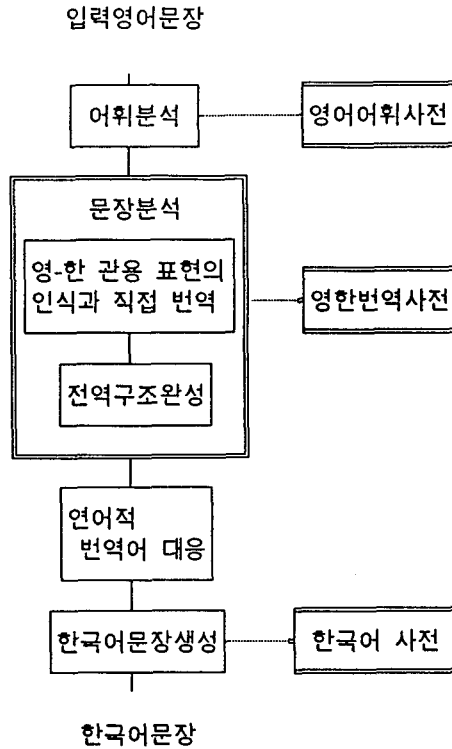


그림2. 관용적 표현의 대응 관계에 기반한 영어-한국어 기계 번역 시스템

4. 번역을 위한 관용적 대응 표현의 인식 과정

관용적 표현의 인식 과정은 자연스럽게 우수한 한국어 번역문을 얻기 위해 가장 중요한 처리 단계이다. 우선, 관용 표현의 인식 방법의 구현 측면에서 관용적 표현들을 형태적으로 두 가지로 분류한다.

(1) 단순 관용 표현(simple idiomatic expressions) : 고정된 단어들의 열(sequence of fixed words)로서 이 단어들 중 하나가 번역 사전의 키워드(keyword)이다. 문장내 단어들과의 비교(string match)에 의해 인식이 가능하다 (예: ("kick the bucket", 죽다), ("at the same time", 동시에)).

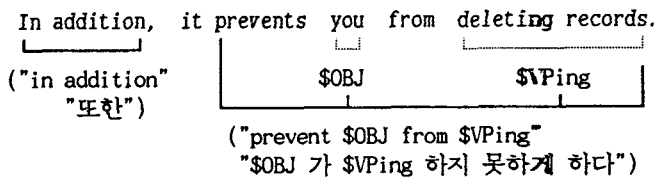
(2) 복합 관용 표현(complex idiomatic expressions) : 고정 단어로 시작되며, 고정 단어들과 함께 한개 또는 두개의 메타 기호(meta symbol)을 갖는 열이다 (예: "take care of \$POBJ" "prevent \$OBJ from \$VPing"). 본 논문에서 이와 같은 메타 기호(\$OBJ \$POBJ \$VPing \$VPinf 등)들은 관용 표현의 변수(variable)라고 불린다. 이러한 복합 형태의 관용 표현들을 인식하기 위해서는 차트에서 속어 변수를 대치할 부분 구조를 얻는 지역적 탐색 과정이 요구된다. 속어 인식 과정 중에 문장의 특정 지점을 시작점으로 하여 요구되는 변수를 대치할 구조를 찾는 지역적인 분석 과정을 '지역 파싱(local parsing)'이라고 한다. 예를 들어, 메타 기호 \$OBJ를 위해서는 목적격의 명사구(NP)에 대한 지역 파싱이 요구된다. 이와 비교하여, 이어지는 전역 구

조 완성 과정을 '전역 파싱(global parsing)'이라고 부를 수 있다. 중요한 점은 지역 파싱과 전역 파싱은 동일한 규칙 집합을 이용하며, 동일한 제어 방식을 갖는다는 점이다.

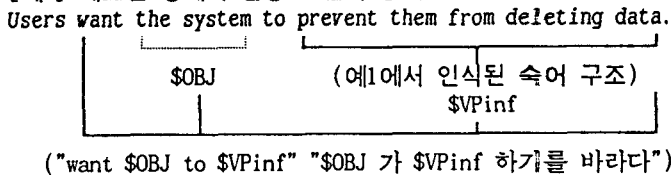
또한, 관용 표현들의 정확한 인식을 위해 사전의 각 표현들은 요소들의 어휘적-문법적 특성들 표현할 수 있어야 한다. 관용 표현의 요소들 중 문장에서 어휘적으로 변형되어 나타날 수 있는 단어들, 선택적으로 취할 수 있는 요소, 또한 삽입 수식이 가능한 경우 등이 그 표현의 명세로서 제공되어야 한다. 예를 들어, "\$POBJ를 돌보다"로 번역될 "take^o care of \$POBJ"에서 take^o는 입력 문장에서 단어 "take" 뿐만 아니라 "took" "taken" "taking" 등을 모두 의미하며, ("take^o (A+) care of \$POBJ", "\$POBJ를 (A^v+) 돌보다")는 "take^o care of \$POBJ" 뿐만 아니라 "take^o good care of \$POBJ" 등을 동시에 표현할 수 있다. 또한, 속어 "mind^o \$ONE'S (own) business"에서 단어 "own"은 입력 문장에서 선택적으로 나타날 수 있음을 말한다.

다음에서는, 몇개의 단순한 실제 문장들을 예로 관용적 표현의 인식에 의한 번역 표현의 직접 대응 과정을 보여준다. 자연 언어 문장의 특성으로서 부분 표현들은 내포(nested, embedded)될 수 있으므로 문장의 각 단어들은 역순으로 오른쪽 끝 단어부터 키워드가 되어 사전을 참조하고 정의된 각 관용 표현들에 대해 인식이 시도된다. 왼쪽 끝 단어부터 시도되는 경우엔 지역 파싱이 스택에 보류되었다가 last-in-first-out 순으로 완성되어야 하는데, 오른쪽 끝 단어부터 인식하는 것은 동일한 결과를 얻으면서 스택을 사용하지 않는 처리 상의 잇점을 갖는다. (그림에서 — 는 지역 파싱에 의해 얻어지는 구조를 표시한다.)

[예1] 단순 및 복합 관용 표현의 인식



[예2] 내포된 형태의 관용 표현의 인식



어떤 단어를 키워드로 하는 관용 표현들의 집합은 고정 단어 우선, 긴 속어 우선의 원칙으로 정렬되어 가장 먼저 인식되는 표현이 가장 바람직한 관용 표현으로 선택할 수 있다. 입력 문장에 나타난 관용적 표현들은 계층 구조를 갖지 않는 평면 구조로 인식되며, 이는 곧 구조적 모호성을 갖지 않음을 말한다. 예를 들어, 위 [예2]의 문장은 관용적 표현의 직접 대응 관계에 의한 경우 최종적으로 한개의 분석 구조를 갖지만, 일반적 관점에서는 네 가지 구조를 갖게 된다. 즉, 복수 분석 구조의 확률이나 그 수는 관용적 표현을 대응시킴으로써 크게 감소될 수 있다. 가장 큰 장점은 기계 번역의 과정에서 관용적 표현으로 인식되고 한국어 표현

으로 이미 대응된 부분에 대해서는 번역을 위한 복잡한 변환 과정이 필요치 않다는 점이다 [11]. 즉, 조사나 어미, 선어말 어미, 어순 등의 결정을 위한 많은 처리가 관용적 표현의 직접 대응에 의해 쉽게 해결될 수 있다.

번역을 위한 핵심 부분인 관용 표현들이 이미 부분적으로 인식된 번역된 문장은 테이블내에 이들 표현에 대한 각 구조를 가지고 있다. 이어지는 전역 구조 완성 과정은 부분적으로 관용 표현들이 인식되고 번역된 문장으로부터, 문법 규칙에 의해 문장 전체의 구조를 얻는 단계이다. 특히, 단어로부터의 분석이 아니라 이미 인식된 관용 표현의 번역 단위들이 중심이 되는 점이 중요하다. 입력 문장의 많은 부분이 관용적 표현으로 인식되었다면 전역 구조 완성 과정은 아주 적은 양의 처리만으로 완료될 수 있다.

계속하여 분석 완료된 구조에서 관용적으로 번역된 부분 이외의 단어들에 대해 언어적 관계에 의한 번역어 대응이 이루어진다. 마지막으로 한국어 문장 생성 과정에서는 지금까지의 처리로 얻어진 구조를 입력으로 하여 관용적 표현의 대응으로 주어진 한국어 표현과 언어 관계로 선택된 한국어 어휘들을 한국어 문법에 맞는 완성된 문장으로 생성한다. 앞의 예의 문장은 한국어 문장 “또한, 그것은 당신이 레코드를 삭제하지 못하게 한다.”로 번역될 수 있다.

5. 관용적 표현의 직접 대응에 의한 번역 방식의 장점

관용적 표현의 직접 대응에 기반한 기계 번역 방식은 번역의 결과와 처리 효율의 측면에서 다음과 같은 장점을 갖는다[11].

(1) 일반적으로 관용적 표현에 대한 계층적 구조 분석과 이에 대한 반복적인 구조 변환이나 단순한 어휘 대치는 전혀 무관하거나 부적절한 의미의 번역문을 낳게 된다. 영어의 관용적 표현들에 대해 구절 구조 규칙으로 얻어지는 계층 구조는 출력 언어인 한국어로의 번역에 있어서 무의미하며, 이러한 경우에 관용적 표현들 단순화된 유일의 평면적 구조로 인식하고 직접 번역 표현을 대응시키는 것이 효율적이다. 그림3에서 그 예를 볼 수 있다.

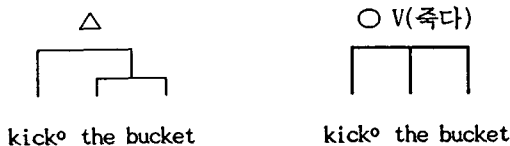


그림3. 관용적 대응 표현에 기초한 평면 구조와 직접 번역

(2) 문법 규칙만으로는 해결되지 못하는 문장의 많은 모호성(ambiguity)들이 관용적 표현의 직접 대응에 의해 해결될 수 있다. 예를 들어, (“take care of \$POBJ”, “\$POBJ를 돌보다”)는 모호한 두 개의 구조 대신 관용적 표현에 대한 평면 구조로 인식된다. 관용적 대응에 의한 모호성의 제거는 전역 구조 완성 과정의 공간적 시간적 효율을 향상시킴으로써, 번역의 효율 또한 크게 높일 수 있다. 그림4에서 그 예를 볼 수 있다. 전치사구나 부정사구를 비롯한 영어에서의 구조적 모호성의 원인이 되는 많은 표현들이 관용적 표현의 직접 대응 방식에 의해 처리될 수 있다.

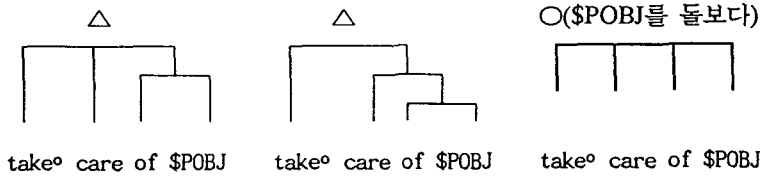


그림4. 관용적 표현의 평면적 인식에 의한 구조적 모호성의 제거

- (3) 영어-한국어 관용적 표현의 직접 대응 관계는 한국어 표현에 있어서 다양하고 불규칙적인 조사, 어미, 선어말 어미 등이 번역 사전으로부터 직접적으로 제공되며, 동시에 어순도 결정된다.
- (4) 입력 영어 문장의 분석에 필요한 규칙의 일부가 번역 사전의 관용적 표현으로 표현될 수 있으므로, 분석 규칙이 상대적으로 단순할 수 있다.
- (5) 번역 사전을 통하여 원하는 표현과 그 번역구를 쉽게 관련지을 수 있고, 추가, 수정이 쉬우므로 번역 시스템이 보다 융통성을 갖는다.

6. 실험 결과

본 논문에서는 영어-한국어 사이의 어휘적 구조적 차이를 쉽게 해결하여 질적으로 우수한 한국어 번역 문장을 얻기 위하여 두 언어의 관용적 표현의 직접 대응 관계에 기반하는 기계 번역 방식을 제안하였다. 이와 같은 번역 방식은 분석 규칙이나 변환 규칙보다는 실질적인 언어 정보의 수집에 중점을 둔 통계적이고 예문 중심적인 접근 방식을 취하고 있다. 이에 대한 실험 도메인으로서 실제 영문 컴퓨터 지침서의 3800여 문장에 대해 입력 문장 분석 과정의 효율과 기계 번역에의 질적 효용에 대해 실험하였다. 그 첫번째 실험으로서 번역을 위한 입력 문장 분석 과정의 효율에 대한 실험에서 관용적 표현의 평면적 인식과 직접 대응에 기초한 문장 분석 과정은 문법 규칙에만 의존하는 한 단계의 분석 과정에 비해 처리 시간은 평균적으로 0.68배로 감소되었으며, 모호한 분석 결과의 평균 갯수에 있어서도 0.60배로 감소되었다. 이 실험 결과는 전체 문장에 대한 평균적 평가로서 관용적 표현의 평면적 인식에 의한 잇점과 그 오버헤드를 모두 포함하고 있다. 또한, 관용적 표현의 직접 대응에 의한 번역의 결과는 90%에 가까운 번역 성공률을 보임으로써 실제 기계 번역에 적용되었을 때 번역의 결과를 크게 향상시킬 수 있음을 알 수 있다.

7. 결론

영어와 한국어는 구조적으로나 어휘적으로 차이가 매우 큰 언어이다. 따라서, 영어와 한

국어 사이의 기계 번역에 있어서 어색하지 않고 적절한 의미를 전달하는 번역문을 얻기 위해서 두 언어 사이의 관용적 표현의 직접 대응 관계를 도입하였다. 관용 표현의 직접 대응 방식에 의한 번역 과정에서 번역의 큰 단위인 관용적 표현들은 무의미한 계층 구조를 갖지 않는 유일한 평면 구조로 인식됨으로써 많은 모호성을 제거하여 입력 문장 분석 및 번역 과정의 시간적 공간적 효율을 크게 높일 수 있다. 또한, 기존의 번역에서의 같은 변환 규칙에 의한 반복적 구조 변환이나 부적절한 어휘 대치없이 질적으로 향상된 우수한 번역 문장을 얻을 수 있다.

실험을 통하여 본 논문에서 제안한 관용적 표현의 대응 관계에 기반한 기계 번역 방식은 기존의 일반적 번역 방식에 비해 시간적 공간적으로 매우 효율적이며, 많은 입력 문장들이 매우 자연스럽게 정확한 한국어 문장으로 번역될 수 있음을 보여주었으며, 기계 번역 시스템의 실용화에 크게 기여할 것으로 기대된다.

8. 참고문헌

- [1] Abell'e, A. and Schabes, Y., Parsing Idioms in Lexicalized TAGs, pp.1-9, Proceedings of the Fourth Conference of the European Chapter of the Association for Computational Linguistics, 1989.
- [2] Jensen, K., PEG 1986: A Broad-Coverage Computational syntax of English. Technical Report, IBM T.J. Watson Research Center, 1986.
- [3] Lee, H.S. and Kim, Y.T., The Collocation Structure of the Transfer Dictionary for Machine Translation in English-to-Korean. pp.24-28, Proceedings of ICCPCOL'92, 1992.
- [4] Linden, E. and Kraaij, W., Ambiguity Resolution and the Retrieval of Idioms: Two Approachs, pp.245-248, Vol.2, Proceedings of COLING-90, 1990.
- [5] Longman, Longman Dictionary of English Idioms: The up-to-date dictionary of idioms, 1987.
- [6] Santos, D., Lexical Gaps and Idioms in Machine Translation, pp.330-335, Vol.2, Proceedings of COLING-90, 1990.
- [7] Schenk, A., Idioms in the Rosetta Machine Translation, pp.319-332, Proceedings of COLING'86, 1986.
- [8] Stock, O., Getting Idioms into Lexicon based Parser's Head, Proceedings of the Annual Meeting of the Association for Computational Linguistics, pp.52-58. Stanford California, 1987.
- [9] Tomita, M., An Efficient Augmented Context-Free Parsing Algorithm, Computational Linguistics, 13:1-2, pp.31-46. 1987.
- [10] Wiren, M., Interactive Incremental Chart Parsing, Proceedings of the European Chapter of the Association for Computational Linguistics, Manchester, England, 1989.
- [11] Yoon, S. H., "Idiomatical and Collocational Approach to English-Korean Machine Translation," Proceedings of 2nd PRICAI, 1992.