

한영 기계번역을 위한 타동사 변환사전 구성

옥철영
전자계산학과

〈요 약〉

변환방식의 기계번역은 변환사전에서 제공하는 정보의 종류와 그의 정밀성에 따라서 변환과정의 복잡도와 번역의 질이 결정된다. 사람에 의한 번역은 양국어 사전에서 제공하는 구절 중심의 번역정보를 이용함으로써, 그 번역의 결과는 정확하고 자연스럽다.

본 논문에서는 양국어 사전에서 제공하는 구절 중심의 여러가지 번역정보들을, 한영기계번역시스템이 이용할 수 있는 형태의 동사 변환사전을 제안하였다.

제안된 변환사전에서는 첫째로, 구절 중심의 번역에서 동사의 역어가 선택되는 기준을 제공하여, 변환과정에서 추가적인 의미해석없이도 역어를 효과적으로 선택할 수 있도록 하였다. 둘째로 동사의 역어가 취하는 구체적인 구문구조를 제공하여, 여러 단계의 구조변환의 복잡도를 줄이면서도 두 언어간의 표현방식의 차이점을 해결할 수 있게 하였다.

A Construction of Transitive Verb Transfer Dictionary for Korean to English Machine Translation

Cheol-Young Ok
Dept. of Computer Science

〈Abstract〉

In a transfer machine translation, transfer dictionary decides the complexity of transfer phase and the quality of translation according to the types and precision of information supplied in the dictionary. Using the phrasal level translated information within human readable dictionary, human being translates a source sentence to the corresponded target sentence

correctly and naturally.

This paper represents a verb transfer dictionary in which the various information is constructed so the machine readable format that the Korean-to-English machine translation system can utilize them.

The dictionary first provides the criteria by which an appropriate target verb is selected in phrase-to-phrase translation without additional semantic analysis in transfer phase. Second, the dictionary provides the concrete sentence structure of a target verb to resolve the expressive gaps between two languages and reduce the complexity of the various structure transfer in word-to-word translation.

1. 서 론

기계번역은 컴퓨터를 이용하여, 하나의 언어로 표현된 문장을 원문장이 지닌 의미와 같은 의미를 전달하는 다른 언어의 문장으로 번역함을 목적으로 한다. 기계번역은 번역방식이 채택하는 전략에 따라서 변환방식과 피봇방식으로 나누어진다 [18]. 이러한 기계번역의 전략을 결정하기 위해서는 번역될 언어간의 문법적인 유사성 정도와 언어 사용의 관습이나 문화적인 유사성 정도 등이 고려되어야 한다.

변환방식의 기계번역은 분석, 변환, 그리고 생성의 세과정으로 구성된다. 분석과정은 원문의 구문과 의미를 분석하여 추상적인 내부표현방법으로 바꾸고, 변환과정은 원문에 대한 목적문장으로의 어휘와 구조변환을 수행한다. 마지막 생성과정은 목적언어의 문법에 맞는 자연스러운 번역문을 생성한다 [14].

변환방식의 기계번역은 변환과정에서 두 언어간의 언어 용법상의 차이를 처리해 줄 수 있기 때문에, 언어체계가 상이한 두 언어간(예, 일본어와 영어, 한국어와 영어)의 기계번역에서 주로 채택되었다 [1, 8, 12, 13]. 이러한 기계번역시스템들은 단어 중심

의 번역으로 변환과정에서의 역어선택시 발생하는 의미적 모호성(semantic ambiguity)을 해결하기 위해서, 변환사전에서 역어선택을 위한 의미제약조건 및 구문구조정보와 기타 필요한 변환규칙등을 제공한다 [7, 16]. 그러나 이러한 단어 중심의 번역방법에서는 역어선택을 위한 정확한 의미해석이 필요하며, 설사 정확한 의미해석이 이루어지더라도, 단어 중심의 변환방식은 기본적인 구조변환 뿐만 아니라 보다 자연스러운 번역문을 생성하기 위한 추가적인 구조변환 및 조정작업이 필요하다 [11].

그러나, 단어 중심의 변환방법이나 규칙에 의한 기계번역과는 달리, 사람이 어떤 문장을 번역할 때는 일반적으로 양국어 사전을 참조하여 문장의 의미를 전달하는 핵심 구절(phrase)을 중심으로 그 의미를 가장 정확히 전달할 수 있는 표현으로 번역을 시도한다. 이때 참조되는 양국어 사전에는 입력 단어가 지닌 의미와 대등한 의미의 역어뿐만 아니라, 구절 내에서의 문맥에 가장 적합한 역어와 해당 구절에 대한 전형적인 번역 용례와 양국어간의 언어 용법상의 차이를 나타내는 여러가지 선언적인 번역정보를 제공한다. 따라서 사전에 기술된 용례와 같거나 비슷한 구절에 대해서는 단어변환이나 변환규칙을 적용하지 않고서도 직접 번

역문을 생성할 수 있다. 이러한 양국어 사전이 컴퓨터가 인식할 수 있는 형태의 사전(MRD: Machine Readable Dictionary)으로 체계적으로 구성되어 있어 기계번역시스템에 이용될 수 있다면, 기계번역의 결과는 정확하고 번역의 질은 향상될 것이다. 이러한 구절 변환을 위한 사전에 대한 연구로는 [9, 10, 15, 17]에서 있었다.

본 논문은 [15]에서 제안한 사전을 근간으로 하여, 문장의 머리어(head)인 동사를 위주로 구절 변환을 위한 한영 타동사 변환사전을 구성함을 목표로 한다. 이를 위하여 [4]의 한영사전을 기초로 하여, 구절 내에서의 문맥에 가장 적합한 역어를 선택할 때 고려하여야 하는 요인들을 분석하여 이들 역어를 결정할 수 있는 기준을 제시하고, 한국어와 영어의 표현방법의 차이등을 고려하여 이러한 차이점들이 구절 변환시 효과적으로 반영될 수 있는 한영 타동사 변환사전을 제안한다.

제안된 동사변환사전은 [15]에서와는 달리, 입력문의 표층구조에서 나타나는 조사의 표층격 정보만을 이용하여 구절의 역어를 결정함으로써 기계번역과정시의 과중한 의미해석을 줄일 수 있다. 또한, 두 언어간에 구조변환이 필요한 부분을 하나의 변환단위로 보아 구조변환이 어떻게 일어나는지를 직접 사전에서 제공하기 때문에, 단어 중심의 변환과정에서의 구조변환 및 조정작업을 생략될 수 있으므로 전체 변환과정이 대폭 축소될 수 있다.

2. 구절변환에서 역어선택

(2.1) “고기를 굽다”

: roast a meat

(2.2) “생선을 굽다”

: grill a fish

(2.3) “식빵을 굽다”

: toast a bread

다른 예로, 한국어 타동사 “낳다”의 경우 “동물이 새끼나 알을 몸 밖으로 내놓다”의 뜻으로, 일반적으로 ‘breed’의 역어로 번

문장을 구성하는 각 단어는 적어도 하나의 의미를 가지며, 같은 의미를 지닌 다른 언어로 번역이 가능하다. 그러나, 번역될 단어가 동형의어어(homograph)이거나 다의어(polysemy)이면, 먼저 문장내에서 어떤 의미로 사용되었는지를 해석하고 그 의미에 적합한 역어를 결정하여야 한다. 이때 동사의 의미는 동사에 하위범주화(subcategorization)되는 구성성분과의 의미적 호응관계에 의해 결정된다. 이러한 의미적 호응관계를 이용하여, 변환사전에서 해당 의미를 지닌 역어로 번역하기 위한 의미제약조건을 제시함으로써 동사의 적합한 역어를 선택할 수 있다. 이러한 의미제약조건은 명사의 의미속성에 따라서 계층구조의 개념구조나 의미범주(semantic category)로 설정되어 의미표(semantic marker)로 주어진다 [7, 16].

그러나, 한영기계번역의 많은 경우에 이러한 의미제약조건은 동사의 의미를 결정하기 위해 필요하지만, 적합한 역어를 선택하기 위해서는 충분하지 않다. 예를 들면, 한국어 타동사 ‘굽다’는 4가지 다른 의미를 가지며, 각 개별 의미는 <음식>, <사진>, <숯>, <도자기>등의 목적어 명사와의 호응관계에 의해 구분되며, 변환사전에서 적당한 의미표와 같은 의미의 역어를 제공함으로써 기본적인 번역문을 생성할 수 있다. 그러나 ‘굽다’의 의미중에서 <음식>의 의미표를 가지는 “요리하다”의 뜻으로 사용된 경우 ‘cook’으로 번역될 수 있으나, 다음 예문에서와 같이 각기 다른 역어로 번역되는 것이 더욱 자연스럽다.

역될 수 있으나 주어인 개별 동물에 따라서 각기 다른 역어를 가진다.

- (2.4) “소가 새끼(송아지)를 낳다” : a cow claves a calf
 (2.5) “말이 새끼(망아지)를 낳다” : a horse foals a colt
 (2.6) “개가 새끼(강아지)를 낳다” : a dog drops a pup
 (2.7) “돼지가 새끼를 낳다” : a pig farrows her young

물론 위의 경우 개념구조나 의미범주를 더욱 세분하여 각 목적어가 각기 다른 의미표를 가져 적합한 역어를 선택하게 할 수 있으나, 그렇게 되면 의미범주가 광범위해지며 개별 역어에 따른 의미표가 만들어질 가능성이 있다.

의미표에 의한 역어선택정보를 제공하는

- (2.8) “손바닥으로 때리다” (strike with the palm of the hand) : ‘slap’
 (2.9) “몽둥이로 때리다” (strike with a stick) : ‘drub’
 (2.10) “회초리로 때리다” (strike with a whip) : ‘flog’

이런 경우 단어 중심의 변환을 한다면, 입력문의 부사구를 그에 대응되는 전치사구로 변환할 수 있다. 그러나 조사 ‘로(으로)’는 조사와 결합되는 명사와 동사의 의미에 따라서 약 14가지의 의미격으로 사용될 수 있으므로, 전치사구의 적합한 전치사를 결정하기 위해서는 명사의 의미속성과

경우와는 달리, 한영기계번역에서 동사를 수식하는 부사구가 적합한 역어를 결정하는 정보를 제공해 주기도 한다. 한국어 동사 “때리다”는 ‘strike’의 역어로 번역될 수 있으나, 사용되는 도구에 따라서 각기 다른 역어로 번역하는 것이 문맥의 의미를 보다 명확히 전달할 수 있다.

동사의 의미를 결정해야 한다 [7].

이상의 예에서는 세분화된 의미표에 의한 역어결정방법이나 전치사구로의 변환등으로 단어 중심의 변환방법에 의해서도 적절한 번역문을 생성할 수 있지만, 아래 예문들인 경우에는 단어 중심의 변환으로는 입력문의 의미를 전달하는 번역문을 생성할 수 없다.

- (2.11) “몸을 꼬다” : ‘writhe’
 (2.12) “코를 골다” : ‘snore’
 (2.13) “표지를 갈다” : ‘recover’
 (2.14) “뒤축을 갈다” : ‘reheel’
 (2.15) “정치에 손을 대다” : ‘meddle’ in politics
 (2.16) “술을 입에 대다” : ‘drink’ wine

이와같은 형태의 번역은 양국 언어에서 개별 어휘가 전달하고자 하는 의미범위의 차이와 하나의 의미를 전달하는 어휘수의 차이때문으로, 각 언어에서 어휘가 생성되고 쓰여지는 양태에 따라 다르다. 이러한 구절은 같은 의미를 지닌 역어선택의 입장에서 보면 동사와 함께 사용되는 성분들이 고정되어(frozen) 다른 변형을 허용하지 않는 숙어로, 해당 의미를 전달하는 구절(phrase)을

중심으로 역어선택이 시도되어야만 정상적으로 번역될 수 있다. 이와같은 구절 중심의 변환은 영한기계번역에서도 고려되는 사항으로, (9,17)에서는 이러한 구절을 관용어구(idiom)으로 보았다.

상기 (2.11)~(2.16)의 예문에서 뿐만 아니라 나머지 다른 예문에서도 개별 동사의 역어가 선택되는 요인들을 숙어로 보아, 변환과정시에 이들 요인들을 하나의 변환

단위로 취급할 수 있다면 번역의 결과는 정확하고 자연스러울 것이다. 본 논문에서는 [4]의 한영사전을 기초로 하여 개별 동사의 역어가 결정되는 요인들을 분석하여, 변환과정의 역어선택시 고려되어야 할 변환 단위의 기준을 제시하고자 한다.

3. 구절변환에서 구조변환

언어체계가 다른 두 언어간의 기계번역에서는 문장을 구성하는 성분의 어순 차이를 해결하기 위해, 기본적인 어휘변환 후에 구조변환을 수행한다. 한국어 문장은 일반적으로 “주어 + 목적어 + 동사”의 어순을 가지나, 격조사에 의해 구성성분의 문법적 관계가 결정되기 때문에 동사를 제외한 구성성분간에는 비교적 자리옮김이 자연스럽다. 반면에 영어는 “주어 + 동사 + 목적어”의 어순으로 구성성분간의 어순은 엄격히 지켜진다 [2, 3].

- (3.1) “도로에 돌을 깔다” : cover a road with stone
- (3.2) “상자바닥에 종이를 깔다” : cover the bottom of a box with paper
- (3.3) “뺨 다리에 붕대를 감다” : swathe a broken limb in bandage
- (3.4) “생선에 야채를 곁들이다” : garnish fish with a green vegetable
- (3.5) “문에 자물쇠를 걸다” : fasten the door with a lock
- (3.6) “실로 노끈을 꼬다” : twist threads into a string
- (3.7) “나무로 숯을 굽다” : burn wood into charcoal

이러한 언어간의 표현의 차이를 처리하기 위해서 일영기계번역시스템에서는, 개별 동사가 의미적으로 요구하는 부사구에 대해 표층격과 심층격으로, 그리고 해당 동사의 역어에서 요구되는 전치사구에 대해 표층격 등으로 세분하여 동사변환사전에 수록하여 변환과정에서 이러한 의미격의 차이가 반영될 수 있게 하였다 [11, 16]. 그러나, 부사구의 의미역은 조사와 결합하는 명사의 의

[1]에서의 한영기계번역시스템(KETS)은 한국어 문장의 분석구조를, 변환 및 생성규칙의 사전을 이용하여 그에 대응되는 영어의 심층구조로 변환한다. 또한 [8]의 한영기계번역시스템(KoEng)에서는 의미적 소성(feature)에 한영 양국어간의 문장구조에 대한 어순을 지정함으로써, 단일화(unification)가 이루어 지는 동안에 구조변환을 행한다. 그러나, 이러한 두 시스템에서는 한국어와 영어와의 기본적인 문장 구조변환은 처리할 수 있으나, 양국 언어에서 의미를 표현하는 방식에 차이가 있는 문장들에 대한 구조변환은 쉽게 처리할 수 없다.

양국 언어에서 의미를 표현하는 방식의 차이는 주로 언어 사용의 용법이나 관습 차이에 의한 것으로, 다음 예문들에서와 같이 한국어 동사가 요구하는 부사구의 의미격이 역어 동사에 의해 요구되어지는 의미격과 다르기 때문이다. 이러한 차이들이 기계번역에서 제대로 반영되지 않으면 한국식 영어가 되어 버린다.

미속성과 동사의 의미를 알아야만 결정됨으로, 입력문의 의미해석부에 과중한 부담을 주게 된다.

위와 같은 표현 차이외에도, 다음 예에서와 같이 한국어 문장에서 목적어를 수식하는 소유격이 영어 문장에서 목적어로, 목적어는 영어 동사의 숙어 형태로 표현되는 경우도 있다.

- (3. 8) "(그녀의) 소매를 당기다" : pull her by the sleeve
 (3. 9) "(손님의) 멍살을 잡다" : seize a guest by the lapels of his coat
 (3.10) "(소년의) 용기를 기리다" : praise boy for his courage
 (3.11) "(철수의) 일을 거들다" : help Cheol-Su to do his work

위의 예문에서와 같이 한국어 문장에서는 목적어를 수식하는 소유격은 일반적으로 생략되어 문장에서 나타나지 않는다. 이런 경우, 문장내의 구절변환만으로는 정확한 번역을 수행할 수 없으나, 변환사전에서 이러한 생략 가능한 정보를 제공하여 차후 문단(paragraph) 단위의 번역에 반영될 수 있도록 하여야 한다.

- (3.12) "홍진에 걸리다" : catch the measles
 "감기에 걸리다" : catch cold
 (3.13) "안개가 걷히다" : the mist clear off
 "목소리가 끊기다" : his voice breaks
 (3.14) "머리를 깎이다" : have one's hair cut
 "의자의 천을 깔리다" : have a chair upholstered

이러한 형태의 구조변환은 [1]에서는 변환규칙으로 처리하였고, [11]에서는 기본적인 구조변환의 전후에 추가적인 구조변환을 두어 처리하였다. 그러나, [1]에서는 변환규칙들을 개별 어휘에 대해 일괄적으로 적용하지 못하며, [11]에서는 추가적인 구조변환으로 전체 변환과정이 복잡해지는 문제점등이 있다.

결국 이러한 문제점들은 변환사전이 규칙이나 단어 중심의 변환정보만을 제공할 뿐 변환된 동사구의 구조변환정보를 직접 제공하지 못하기 때문이다. 이러한 문제점들을 해결하기 위한 방법으로 [6,7]에서는 Hornby 패턴을 기반으로 한국어 동사에 대응되는 영어 동사의 구문구조를 제공하였다. 그러나, Hornby 패턴의 구문구조도 개별 어휘에 대해 일괄적으로 적용되지 못하기 때문에 충분하지 못하다.

본 논문에서는 동사의 역어를 결정할 때 고려되어야 할 동사구를 하나의 변환단위로

이러한 경우외에도 한국어의 문장이 여러 다른 형태의 영어 문장으로 변환된다. 예문(3.12)는 한국어에서의 자동사가 영어에서의 타동사로, (3.13)은 피동형이 능동형으로 변환됨을 보여 주며, (3.14)는 사역의 의미를 나타내는 표현방식의 차이를 보여 주고 있다.

처리했던 것과 같은 관점에서, 개별 동사 어휘에 대해 양국 언어간에 구조변환이 필요한 부분들을 하나의 변환될 구절로 보아, 이들 구성성분들이 어떻게 변환되는지를 직접 변환사전에 수록함으로써 상기의 문제들을 해결하고자 한다.

4. 변환사전의 구조

사람이 번역을 할 때 이용하는 양국어 사전에는 표제어 단어와 대등한 의미의 역어와 해당 역어로 번역되는 전형적인 번역 용례 뿐만 아니라, 표제어 단어와 함께 자주 사용되는 구절에 대해 그 구절의 문맥에 가장 적합한 다른 역어와 양국 언어간의 언어용법 차이등의 번역에 필요한 여러 정보들이 수록되어 있다. 이러한 정보들은 양국 언어의 전문가들에 의해 수집 정리된 정보들로서, 기계번역에서도 이용될 수 있다면 정확하고 자연스러운 양질의 번역 결과를 얻을

수 있을 것이다.

본 논문은 한영기계번역시스템이 동사구를 중심으로 번역을 시도할 때, 동사구의 문맥을 가장 정확하게 전달할 수 있는 역어선택의 기준과 구조변환정보를 함께 제공하는 동사변환사전을 구성함을 목적으로 한다. 이러

한 목적을 위해 한영사전 {4}에서 제공하는 타동사 약 1300개의 개별 동사의 번역정보들을 중심으로 역어가 결정되는 요인과 양국 언어간에 의미를 전달하는 표현방식의 차이점들을 수집 분석하여 다음과 같은 기본 구조의 한영변환사전을 제안한다.

<p>K_Verb <Sent_Type1> Factor11_of_Selection = {(K_Noun,E_Noun),...} VP_English Factor1M_of_Selection = {(K_Noun,E_Noun),...} VP_English <Sent_Type2> Factor21_of_Selection = {(K_Noun,E_Noun),...} VP_English Factor2M_of_Selection = {(K_Noun,E_Noun),...} VP_English <Sent_TypeN></p>

위의 사전구조에서 K_Verb는 표제어 동사로, 동형이의어는 일반사전에서 수록하는 방법에 따라 별도의 표제어를 가지도록 하였다. 그러나, 동형이의어와 다의어는 구절 변환의 입장에서 특별히 구분되어질 필요는 없어 하나의 표제어에 다른 항목으로 수록되어도 무방하나, 단어의 의미적인 입장에서 기본적인(default) 역어를 제공하기 위해서는 분리해 두는 것이 타당할 것 같다.

표제어의 동사가 다의어인 경우, 일반적으로 각 의미를 표현하기 위해 통사적으로 조금씩 다른 문장구조를 구성한다. 이것은 동사에 하위범주화되는 보어(complement)들의 문법기능과 의미역할의 차이에 의한 것으로, 입력문장의 구문분석시 뿐만 아니라 동사의 의미를 결정하기 위해서도 이용될 수 있다. {7,15}에서는 동사의 의미적 사용에 따라서 문장을 분류하여 동사의 역어선택시에 이용하였다. 그러나, 동사의 의미적 사용에 따른 분류는 동형의 조사가 사용되어 통사적으로 같은 구문구조를 지닌

문장이라 할지라도 조사의 의미역할에 따라 다른 문장으로 분류됨으로, 문장구조를 결정하기 위해 조사의 의미역을 결정해야 하는 추가적인 부담이 있다.

구절변환에서 단어나 조사의 의미해석을 하지 않고 입력 문장의 표층구조만으로 역어선택과 구조변환을 수행하기 위해서, 변환사전에는 가능한 모든 용례에 대해 역어선택과 구조변환의 요인들을 수록하여 번역에 필요한 정보를 제공한다. 따라서 본 논문에서는 통사적인 관점에서 하위범주화되어지는 성분들에 의해 문장을 분류한 {5}의 방법을 채택하였다. 또한, 다의어 동사는 일반사전에서와 달리 의미별로 항목을 분류하지 않고, 같은 구문구조를 지닌 의미들은 모두 같은 항목으로 분류하고 이들의 구문구조를 <Sent_Type>에 표시하였다.

Factor_of_Selection은 역어선택시 고려되어야 할 문장성분의 지시표로서, 동사에 하위범주화되는 보어(SBJ,OBJ,ADV), 참여(MNR), 소유격(POS), 반대어(NEC),

관용어구(IDM), 그리고 기본적인 역어(DFT)로 나누었다. 여기서 하위범주화되는 보어중 ADV는 부사격 조사의 의미역할로 나누지 않고, 단순히 표층구조에서 나타나는 형태로 ADV(‘에’), ADV(‘에게’), ... 등으로 표시하였다. 첨어(MNR)는 동사에 하위범주화되지는 않지만 특별한 부사와 사용되었을 때, 보다 적합한 역어를 제공하기 위한 지시표이다. 소유격(POS)는 예문 (3.8)~(3.11)에서와 같이 한국어의 소유격이 구조변환을 필요로 한 경우를 구분하기 위해서 사용한다. 반대어(NEG)는 해당 동사가 문장에서 부정으로 사용되었을 때, 특별히 다른 역어가 제공되는 경우에 사용된다. 관용어구(IDM)은 [5]에서 분류한 구문구조외에 특별한 형태의 구문구조를 필요로 하는 경우와 속담등의 역어선택 및 구조변환정보를 제공하기 위해 사용한다.

{{(K_Noun, E_Noun), ...}}는 역어선택 요인들과 함께 결합되어 VP_English로 번역되는 명사들의 리스트(collocation list)

로, 가능한 모든 명사들을 수록한다. 여기서, K_Noun은 한국어 명사이고 E_Noun은 K_Noun에 대응되는 영어 명사로 1 : 1 대응관계를 가진다. 만약 이러한 리스트의 요소가 무한히 많고 이들이 특정 의미속성을 가진다면, DFT에 그들의 의미속성을 기술하여 중의성을 피하도록 하였다. 이렇게 VP_English로 번역되어지는 모든 명사들의 리스트를 역어가 결정되는 요인과 함께 기술함으로써, 동형의 조사가 다른 의미적으로 사용된 경우라도 정확한 번역정보를 제공할 수 있다.

VP_English는 해당 동사가 {{(K_Noun), ...}}의 명사와 함께 사용되었을 때 영어 동사구의 표층구조로, 역어선택시 고려되는 성분들이 어떻게 구조변환이 되어야 하는지에 대한 정보를 직접 제공한다.

다음은 2,3절의 예문들에 대해 구절변환을 위한 변환사전에서의 구체적인 예를 보이고 있다

굽다 (타)

1 <KVP11>

- ADV(‘로’) = {{(석쇠,?)}} grill [OBJ]
- OBJ = {{(고기, a meat)}} roast [OBJ]
- OBJ = {{(생선, a fish)}} broil [OBJ]
- OBJ = {{(식빵, a loaf of bread), (김, a laver), (떡, a rice cake)}} toast [OBJ]

2 <KVP10>

- OBJ = {{(벽돌, bricks), (도자기, china)}} burn [OBJ]
- OBJ = {{(사진,?)}} print
- MNR = {지나치게} overprint

낳다 (타)

1 <KVP10>

- OBJ = {{(자식, a child), (아이, a baby), (새끼, young), (알, eggs)}}
 - SBJ = {{(소, a cow)}} [SBJ] calve [OBJ: a calf]
 - SBJ = {{(새, a bird)}} [SBJ] lay [OBJ]
 - DFT = {동물} [SBJ] breed [OBJ]

때리다 (타)

1 <KVP11>

ADV('로') = {(손바닥, ?)} slap [OBJ]

ADV('로') = {(뭉둥이, ?)} drub [OBJ]

폴다 (타)

1 <KVP10>

OBJ = {(코, ?)} snore

대다 (타)

1 <KVP9>

OBJ = {(청진기, a stethoscope), (컵, a glass), (손, *POS hand)}

ADV('에') = {(가슴, *POS chest), (입, *POS lips)} put [OBJ] to [ADV]

ADV('에') = {(귀, the ear)} put [OBJ] behind
(ADV)

ADV('에') = {(어깨, the shoulder)} touch [POS] on [ADV]

ADV('에') = {(입, ?)}

OBJ = {(술, wine)} drink [OBJ]

OBJ = {(과일, a fruit), (과자, a cake)} eat [OBJ]

IDM = {손대다}

ADV('에') = {(사업, a business)} start [ADV]

ADV('에') = {(정치, politics)} meddle in [ADV]

깔다 (타)

1 <KVP9>

OBJ = {(종이, a paper), (돌, stone)} cover [ADV] with [OBJ]

ADV('에') = {(바닥, the bottom), (도로, a road), (마루, the floor)}

OBJ = {(이부자리, a bed)} make [OBJ]

OBJ = {(깔개, a carpet), (융단, a rug), (돛자리, a mat)} spread [OBJ]

위의 예에서 collocation list의 E_Noun 이 ?으로 표시된 것은, 명사에 해당 역어를 가지지 않는 경우이다.

5. 결 론

사람이 번역을 행할 때 이용하는 양국어 사전이 체계적으로 구성되어 있어 기계번역 시스템이 이용할 수 있다면, 번역의 결과는 정확하고 자연스러울 것이다.

본 논문은 사람이 이용하는 사전을 MRD 형태의 사전을 구성하기 위한 기초적인 연구로서, 첫째로 한 문장에서 동사의 역어를 선택할 때 고려되어야 할 기준을 제시하며, 둘째로 한영 언어간에 의미를 표현하는 구조적인 차이를 직접 제공할 수 있는 동사 변환사전을 제안하였다. 제안된 사전을 이용한 기계번역시스템은, 역어선택시 의미정보를 사용하지 않으며 각 동사별로 역어선택 및 구조변환이 필요한 성분들이 하나의 변환 단위로 취급할 수 있기 때문에, 단어

중심의 변환방법에 비해 변환과정이 단순해지고 번역 결과는 정확하고 자연스러운 잇집이 있다.

그러나 제안된 변환사전을 이용할 한영기계번역시스템은 사전에서 제공하는 역어선택 정보들을 빨리 찾을 수 있는 알고리즘과, 역어선택의 요인이 복합적으로 나타나는 경우에 우선적으로 적용할 알고리즘의 개발이 시급하다.

제안된 변환사전은 상용의 한영사전에서 제공하는 기본적인 번역용례를 중심으로 구성하였기 때문에 용례와 같은 형식의 문장의 번역은 비교적 정확하고 자연스러울 수 있으나, 사전에서 제공되지 않은 문장들에 대해서는 특별히 번역정보를 제시하지 못한다. 이를 위해서는 더 많은 용례를 수집하여 계속적인 사전의 추가 및 변경작업이 필요하며, 이런 작업을 손쉽게 할 수 있는 사전개발 환경에 대한 연구가 필요하다.

또한 제안된 사전에서는 의미정보를 거의 사용하지 않으므로, 사전의 번역정보들이 중복되고 사전에서 생략된 정보를 기본적으로 처리하지 못하는 문제점이 있다. 따라서 이를 위해서는 이미 연구되어진 단어 중심의 변환사전에서의 의미정보들을 접목하는 연구가 병행되어야 하며, 보다 완전한 구절 변환을 위해서 명사, 형용사, 그리고 조사에 대한 변환사전의 연구도 추가되어야 할 것이다.

참 고 문 헌

1. 김영택, 한영 변환기 및 영어 생성기 모형 개발, 88 국책 연구 개발사업과제 보고서, "한영,영한 기계번역 시스템 개발(I)", 과학기술처, 1989
2. 남기심, 고영근, 표준 국어문법론, 탑출판사, 6판, 1989
3. 박병수, "기계 번역에서 본 한국어의 특징", 정보과학회지, Vol.7, No.6, P.31-39, 1989
4. 시사영어사, 뉴월드 한영 대사전, (주) 시사영어사, 서울, 1987
5. 신효필, "HPSG를 기초로 한 한국어 동사의 하위범주화", 언어학 연구, 제 7호, 서울대학교 언어학과, 1990
6. 신효필 외2인, "한-영 동사의 하위범주화와 대응에 관한 연구", 제 2회 기계번역 Workshop 발표논문집, 시스템공학센터, 한국과학기술원, P.217-236, 1989
7. 이상조 외3, "한영 기계번역을 위한 사전 구성에 관한 연구", 제 2회 기계번역 Workshop 발표논문집, 시스템공학센터, 한국과학기술원, P.171-184, 1989
8. 정희성 외5, "한영 기계번역 시스템 (KoEng)의 설계", 제 2회 기계번역 Workshop 발표 논문집, 시스템공학센터, 한국과학기술원, P.87-96, 1989
9. Kim, Y.T., "A Construction of Transfer Dictionary for Machine Translation", Proceedings of The Korea-US Bilateral Workshop on Computers, Artificial Intelligence and Cognitive Science, Seoul, Korea, P.1-6, 1991
10. Lee, H.S., and Kim, Y.T. "A Construction of Semantic Markers in the Transfer Diction-

- ary from Relative Examples", Proceedings of The Korea-US Bilateral Workshop on Computers, Artificial Intelligence and Cognitive Science, Seoul, Korea, P.296-300, 1991
11. Nagao, M and Tsujii, J.I., "The Transfer Phase of the Mu Machine Translation System", Proceedings of 11th International Conference on Computational Linguistics (COLING 86), P.97-103, 1986
 12. Nagao, M., Tsujii, J.I., and Nakamura, J.I., "Machine Translation from Japanese into English", Proceedings of IEEE, Vol.74, No.7, P.993-1012, 1986
 13. Nagao, M., Tsujii, J.I., and Nakamura, J.I., "The Japanese Government Project for Machine Translation", Computational Linguistics, Vol.11, No.2-3, P.91-110, 1985
 14. Nirenburg, S., Machine Translation: Theoretical and methodological issues, Cambridge Univ. Press, 1987
 15. Ok, C.Y., "The Construction of Verb Transfer Dictionary for Korean to English Machine Translation", Proceedings of The Korea-US Bilateral Workshop on Computers, Artificial Intelligence and Cognitive Science, Seoul, Korea, P.287-291, 1991
 16. Sakamoto, Y., Ishikawa, T., and Satoh, M., "Concept and Structure of Semantic Markers for Machine Translation", COLING 86, P.13-19, 1986
 17. Santos, D., "Lexical gaps and idioms in machine translation", Proceedings of 13th International Conference on Computational Linguistics (COLING 90), P.330-335, 1990
 18. Slocum, J., "A Survey of Machine Translation: Its History, Current Status, and Future Prospects", Computational Linguistics, Vol.11, No.1, P.1-17, 1985