

언어과학연구 24 (2003)

한영기계번역을 위한 언어구의 처리 방법*

이 동 혁 (고려대학교)

Lee, Dong-Hyeok. 2003. A Way of Processing Collocational Phrases for the Korean-English Machine Translation. *The Journal of Linguistic Science* 24, 189-212. The primary purpose of this paper is to review some basic concepts and problems related to collocational phrases, and to introduce a way of processing collocational phrases for the Korean-English machine translation based on Feature Computational Grammar. In reviewing the previous definitions and characters dealing with collocation, we would find that collocation can be defined properly as the combination of words that have a certain mutual expectancy, and a collocational phrase can be defined as a kind of fixed expression. In order to apply to the Korean-English machine translation, collocation should be divided into a functional collocation and semantic collocation according to the role of collocation. Therefore we can disambiguate forms, grammatical categories, structures, and meanings with collocational information. In addition, we can reinterpret a phrasal construction into a lexical expression. (Korea University)

Key words 관용구(idiom), 연어(collocation), 언어구(collocational phrase), 연어관계(collocation), 일반구(free expression), 자질연산문법(Feature Computational Grammar), 중의성 해결(disambiguation), 한영기계번역(Korean-English machine translation)

* 이 글은 언어과학회 2003년 동계 전국 학술 발표대회에서 발표한 글을 읽고 다듬은 것이다.

1. 도입

이 글은 효율적인 한영기계번역을 위하여 언어관계(collocation)를 설정하고, 그러한 관계에 있는 표현들의 처리 방안을 제시하는 것이 목적이다. 일반적인 이론언어학 입장에서 보면, 이 글의 연구 대상이 두 가지 점에서 낯설게 느껴질 것이다. 하나는 ‘기계번역’이요, 다른 하나는 ‘언어관계’ 때문이다. 둘 다 극히 최근에야 국어학자들 사이에서 본격적으로 논의가 되기 시작한 대상이다.

전산학자들(또는 전산언어학자들) 중심으로만 진행되어 오던 기계번역(machine translation)의 연구가 국어의 본질에 대한 이해 부족으로 한계에 봉착하게 되자 국어학 연구의 도움을 절실하게 요구하게 되었고, 반면에 국어학자들은 자신이 머리 속으로만 추상적으로 이해하고 체계화하였던 이론 문법을 실제 기계에 모의하여 확인할 수 있다는 점에서 기계번역의 발전은 전산학자와 국어학자 사이에서 상보적인 것이었다.

언어관계는 본래 구조주의적 패러다임 속 결합관계(syntagmatic relation)의 한 분류로서 독일의 어휘장(lexical field) 이론과 영국의 Firth를 비롯한, 유럽 중심의 언어학에서 발달되어 오던 연구 대상이었다. 따라서 Chomsky에서 태동한 생성문법의 이론이 우위적 바탕을 이루고 있는 현재의 통사론 논의에서는 언어관계가 선택제약(selectional restriction)으로는 설명하지 못하는 우연한 관계일 뿐이라고 단순하게 언급되고 지나갈 뿐이었다. 그러나 사전편찬이나 자연언어처리를 연구하는 국어정보학의 절대적인 영향으로 현재는 언어관계가 국어학계에서 설명하고 해결해야 할 대상이 되었다.

기계번역은 컴퓨터를 매개로 하여 번역대상어(source language)를 분석하여 뜻에 맞는 번역목표어(target language)로 생성하는 일련의 절차를 말한다. 한영기계번역에서는 번역대상어가 한국어이고, 번역목표어는 영어가 된다. 한영기계번역에서 언어관계 표현의 처리가 중요한 이유는 무엇보다 번역의 질을 향상시킬 수 있기 때문이다.¹⁾ 기계번역의 과정이 크게 ‘분석 과정’과 ‘생성 과정’으로 나

1) 현재 기계번역의 연구는, 실제 한국어 자료를 엄밀하게 분석하는 것보다 기계번역의 거시적인 방법론을 중심으로 논의가 이루어지고 있다. 예를 들어, 예제 기반 기계번역, 코퍼스 기반 기계번역, 존재론(ontology) 기반 기계번역 등이 현재 논의되고 있는 기계번역의 거시적인 방법론이다. 그러나 한국어의 실제 예를 중심으로 정밀하게 기계번역의 전반적인 처리 방법을 보여주는 미시적인 연구가 미흡하다는 게 우리의 판단이다.

된다면²⁾, 우리는 이 중에서 ‘분석 과정’에 논의의 무게를 실을 것이다. 따라서 분석 대상이 한국어가 되고, 바로 이 한국어의 언어관계 처리 방안이 중심이 될 것이다.

이 글의 논의는 다음과 같이 진행된다. 2장에서는 언어구를 한정하여 이 글의 논의 대상을 분명히 할 것이다. 3장에서는 한영기계번역 시스템의 처리 방법을 자질연산문법을 중심으로 살펴보고 4장의 언어구의 처리 방법 제시에 도움이 되도록 할 것이다. 4장에서는 3장의 자질연산문법을 토대로 하여 언어구를 처리하는 방법을 구체적으로 제시할 것이다. 마지막으로 5장에서는 글의 내용을 요약하고, 해결하지 못한 문제점을 밝힐 것이다.

2. 언어구의 일반론

2.1 언어구의 특성

언어구가 구(phrase)의 일종인 이상, 항상 다음의 구조를 보인다.

$$(1) [X+Y]_{\text{PHRASE}}$$

언어구가 독립적인 논의 대상이 될 수 있는 것은 X와 Y의 관계가 특별하기 때문이다. 우리는 이 관계를 언어관계라 부르고자 한다. X와 Y의 관계는 일단, 결합관계라는 이름으로 우리에게 너무나 친숙하다. 결국 언어관계란 결합관계의 한 유형일 것이다.³⁾

이제 언어관계에 대한 일반적인 정의를 살펴보자. 언어관계는 ‘특정한 어휘 사이에서 나타나는 긴밀한 의존관계’라고 보통 정의하고 있다.⁴⁾ 정의의 핵심은 X와 Y가 특정한 어휘여야 하고, X와 Y의 관계가 긴밀하고 의존적이어야 한다는 것이다. 즉, 결합의 동기가 특정 어휘여야 한다는 조건이 가장 중요하다고 하겠다. X

2) 이 글의 모태가 되는 기계번역 시스템의 구조도는 3장에서 구체적으로 도식화하였다.

3) 논자에 따라서는 계열관계(paradigmatic relation)와 쌍을 이루는 결합관계를 언어관계라고 이룰하기도 한다.

4) 기존의 언어관계 정의에 대해서는 이동혁(1998)을 참조하기 바란다.

와 Y의 긴밀한 관계는 이러한 결합의 동기에 따른 결과적 현상이기 때문이다.

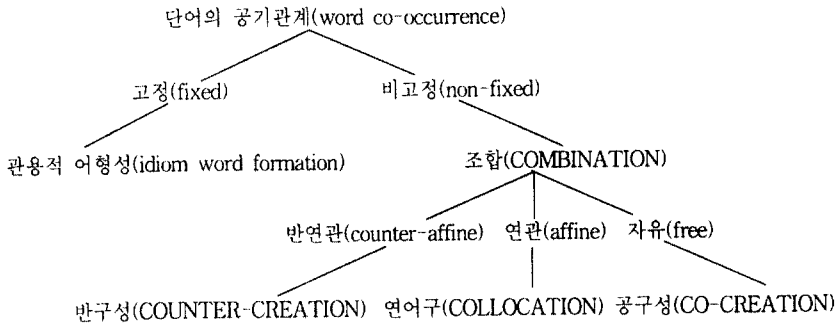
X와 Y는 어휘 내적으로 음운론적 특징과 통사론적 특징, 그리고 의미론적 특징 등을 수반하고 있다. 이러한 특징은 X와 Y의 사전(dictionary)에 기록되어 있다. X와 Y의 특징은 어휘항목으로서 각각의 정체성(identity)을 나타내는 것이기도 하지만 다른 어휘항목과의 관계성을 규정하는 것이기도 하다. 그러나 Lipka(1992:166)에서도 설명했듯이 언어관계는 이러한 특징들에 중립적이다. 왜냐하면 X와 Y의 관계가 언어관계인 것은 자질 차원의 결합관계가 아니라 특정한 어휘 차원의 결합관계이기 때문이다. 아래의 예를 통해 자세히 살펴보기로 한다.

- (2) 가. 밥을 먹다
나. 새빨간 거짓말

(2가)의 Y 자리에 있는 ‘먹다’는 [食]의 의미를 갖는 것으로 X가 [+음식물]이 기만 하면 X와 Y가 결합하는 데 문제가 없다. 자질 차원이나 부류 차원의 문제다. 한편, (2나)의 ‘거짓말’이 X에 대해 [+색깔]로 결합을 제한한다면 ‘새파란 거짓말’도 가능해야 한다. 그러나 그렇지 않다. 적어도 ‘빨간 거짓말’의 표현마저 불가능한 것은 X와 Y의 결합관계가 어휘 차원의 문제임을 증명하는 것이다. 따라서 X와 Y는 특정한 어휘의 긴밀한 의존관계이면서 어휘내항(lexical entry)의 문맥적 특성으로 설명할 수 없기 때문에 우연한 관계라고도 할 수 있다. 우리는 (2가)의 구 유형을 일반구(또는 자유구)라 하고, (2나)의 구 유형을 언어구라 할 것이다.

언어구는 특정 어휘 차원의 결합이기 때문에 X와 Y의 결합이 고정적일 수밖에 없다. (1)의 구 구성을 보이면서 X와 Y의 결합이 고정적인 구는 언어구 말고도 관용구(idiom)가 있다. 관용구는 잘 알려져 있는 대로, X와 Y의 의미 합으로는 전체 구의 의미를 도출해 낼 수 없는 구성이다. 예컨대, [落榜]의 의미를 지닌 ‘미역국을 먹다’는 ‘미역국’과 ‘먹다’의 개별적인 의미를 공식적으로 합쳐서는 도저히 구 전체 [落榜]의 의미를 도출해 낼 수가 없다. 이 때문에 관용구를 통째로 어휘부의 사전에 등재하는 것이 보통이다. 그러나 관용구를 제외한 일반구와 언어구는 구의 구성 과정이 공식적으로 설명이 된다.

지금까지 설명한 구 구성의 방식에서 언어구의 위치를 Hausmann(1985:399)를 따라 도식화하면 아래와 같다.



요컨대, 2.1절에서 논의한 바에 따라 언어구의 특성을 간략히 정리하면, 다음과 같다.

- 일반구: [+자유 결합, +투명 결합]
- 언어구: [+고정 결합, +투명 결합]
- 관용구: [+고정 결합, -투명 결합]

2.2 언어관계의 유형

언어관계의 유형은 X와 Y의 정체성을 확인함으로써 우선 분류할 수 있다. 우리는 2.1절에서 언어관계가 X와 Y는 특정한 어휘의 긴밀한 의존관계이면서 어휘 내향의 문맥적 특성으로 설명할 수 없기 때문에 우연한 관계라고 규정한 바가 있다. 이에 의하면 X나 Y 중 어느 하나가 어휘가 아니라 특정한 문법 형태소가 요구되되, 이 문법 형태소의 결합을 문법적으로나 의미적으로 설명할 수가 없다면 언어관계로 포함시킬 수 있다고 본다. 이를 Benson et al.(1986)과 金鎮海(2000), 임흥빈(2002)를 따라 문법적 언어관계(grammatical collocation)라고 부르기로 하겠다.⁵⁾

5) X나 Y 중 어느 하나가 어휘가 아니기 때문에 엄밀한 의미에서는 언어관계로 보기가 어렵기도 하다. 그러나 X와 Y가 합쳐져 독립된 문법 기능을 발휘할 수 있다는 점에서 X와 Y의 관계성을 규정할 필요가 있고, X와 Y의 통합 구성을 범주화할 필요가 있다. 현재로서는 이러한 구성을 엄밀한 의미의 언어관계에 준하는 준언어관계(semi-collocation)의 일부로 보고, 이를 설명하거나 응용하는 것이 최선이라고 본다.

(3) 가. 철수는 지금 가야 한다.

나. 철수가 미국에 가는 바람에 우리는 아무것도 하지 못했다.

(3)과 같은 문법적 언어관계를 보이는 언어구들은 한국어의 문법 형태소가 다 채우지 못하는 문법 기능들을 보충해서 완성하기에 상당히 중요하다. 따라서 언어관계는 X와 Y 중 문법 형태소의 포함 여부에 따라 문법적 언어관계와 어휘적 언어관계(lexical collocation)로 일차 분류할 수 있다.

(4) 언어관계의 유형(1)

가. 문법적 언어관계: '-어야 하다', '바람에'

나. 어휘적 언어관계

어휘적 언어관계는 X와 Y의 논항 관계, 수식 관계와 같은 문법적 관계에 따라 하위 분류할 수도 있고, X와 Y의 하위 범주적 특징에 따라 다시 하위 분류할 수도 있다. 이에 따라, '체언+용언' 구성, '수식어+체언' 구성, '부사어+용언' 구성 등으로 하위 분류할 수 있는 것이다. 보통은 어휘적 언어관계를 이와 같이 분류하나, 우리는 한영기계번역을 위한 언어구의 처리 방법을 논하기 때문에 그 목적에 부합하도록 언어관계를 재분류하는 것이 좋겠다.

앞서 (3)의 문법적 언어구는 한국어의 문법 기능을 보충하기 위하여 구성된 표현이었다. 한편, 어휘적 언어관계 중에서도 이와 유사한 유형이 있다. 대표적인 것이 이른바 경동사(light verb) 구성이다. 경동사의 앞에는 반드시 서술성 명사가 와야 하는 것은 경동사의 가장 중요한 분포적 특징이다. 문장의 의미역에 따른 논항구조를 서술성 명사가 결정한다면, 경동사는 순수히 통사적 필요성에 의해 특정 어휘로 표현된 것이다. 그리고 문장에서 경동사는 서술성 명사와 통합하여 서술어의 문법 기능을 담당하는 것이기에 문법적 언어관계와 닮아 있다고 보겠다. 우리는 이와 같은 언어관계를 기능적 언어관계로 부르겠다.

X나 Y 중 하나가 자체의 개념 구조에 공기 가능한 어휘를 하위명세화한 구성이 어휘적 언어관계의 한 갈래가 될 수 있다.⁶⁾ 가령,

(5) 가. '똥을 누다', '춤을 추다', '배가 고프다'

6) 金鑽海(2000)에서는 '전제적 언어'라고 부르고 있는 것이다.

나. '육을 먹다', '마음을 먹다', '나이를 먹다'

(5가)의 '누다', '추다', '고프다'와 (5나)의 '먹다'는 자신의 개념 구조에 '똥', '춤', '배'와 '육', '마음', '나이'와 같은 공기 가능한 어휘를 하위명세화하고 있다. 위에서 (5가)와 (5나)로 분류한 것은, (5가)의 '누다', '추다', '고프다'는 의미가 중의적이지 않는데 비해 (5나)의 '먹다'는 어휘 내적으로 의미적 중의성을 보이고 있는 어휘이기 때문이었다. 우리는 이러한 언어관계를 의미적 언어관계라고 부르겠다.

이로써 우리는 기존의 문법적 언어관계와 어휘적 언어관계를 언어구의 구성 목적에 따라 아래와 같이 재분류할 수 있게 되었다.

(6) 언어관계의 유형(2)

- 가. 기능적 언어관계: '-어야 하다', '결정을 내리다' 등
- 나. 의미적 언어관계: '똥을 누다', '육을 먹다' 등

3. 한영기계번역의 처리 방법론: 자질연산문법을 중심으로

3.1 자질연산문법의 개관

본고에서 채택하고 있는 한영기계번역을 위한 문법 모델은 자질연산문법 (Feature Computational Grammar, 줄여서 FCG)이다.⁷⁾ FCG의 목적은 분해기 문법이라는 정보처리를 위한 언어 분석 이론이다. FCG는 언어 자료체를 중심으로 하여 연구를 진행하며, 이 자료체들에서 귀납되는 규칙과 반복적으로 등장하는 원리를 인간의 언어 직관, 즉 언어 능력의 반영으로 간주한다(김원경 2000:31). 이와 같은 성격의 FCG는 다음과 같은 체계를 가진다.

(7) FCG의 체계

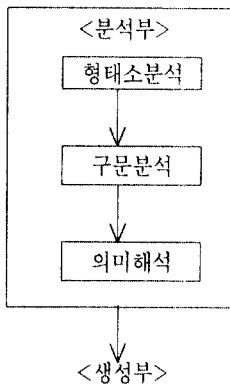
어휘부(포화된 자질 집합) ↕ 연산부(연산규칙)	← 연산 원리
시스템 구성	시스템 이론

7) FCG에 대한 자세한 설명은 김원경(2000), 유혜원(2002), 고창수(2002)를 참고하기 바란다.

어휘부는 어휘항목과 어휘항목의 형태적 정보, 통사적 정보, 의미적 정보, 시스템 정보가 포화된 자질의 집합으로 구성되어 사전은 이룬다.⁸⁾ 사전은 형태소 분석, 구문 분석의 각 분석 과정과 생성의 과정에서 참조하는 정보를 자질과 어휘항목으로 저장하고 있는 셈이고, 각 분석 과정에서 처리한 정보들을 사전에 반영하기 때문에 사전은 폐쇄적인 집합체가 아니며, 자질의 첨가와 삭제가 가능한 개방적인 집합체이므로 유연한 성격을 지닌다. 그리고 하나의 어휘항목에 대하여 대역어가 여러 개인 다의어 구성이라면, 사전 내부에 다의적인 중의성을 해결하는 연산 절차가 포함되어 있고, 하나의 어휘항목이 독립적인 기능을 갖지 못하고 다른 단어와 복합 구성을 이루어서 독립적인 기능을 갖게 되면, 이에 대한 처리도 사전 내부에서 하게 된다.

연산부에는 형태소 분석과 구문 분석과 의미해석을 담당하는 분석부와 단어를 형성하고 구문을 생성하는 생성부가 포함되어 있으며, 각 분석 과정의 연산 결과로 도출된 정보는 FCG에서 별도의 부문으로 구축한 분석결과표에 저장된다.

(8) 연산부의 조직



8) 김원경(2000)에서는 FCG에서 상정한 자질의 특성을 아래와 같이 설명하였다.

- 자질: 특정한 어휘항목의 외연을 표시하기 위한 내항
 - ㄱ. 어휘항목 α 는 항상 자질을 가지며, 이러한 자질 집합은 $\{f_1, f_2, f_3, \dots\}$ 로 표시된다.
 - ㄴ. 어휘항목 α 와 구별되는 어휘항목 β 는 적어도 하나의 자질에서 α 와 구별된다.
 - ㄷ. 어휘부 내 어휘항목의 자질체계는 개방적이다.
 - ㄹ. 어휘항목의 자질은 $\{\pm\}$ 로 표시되지 않는다.
 - ㅁ. 어휘항목들은 어휘 그물로 연결되어 있다.
 - ㅂ. 연결된 어휘항목들은 특정 자질을 공유한다.

정보 테이블은 분석의 결과를 저장하는 공간이고, 이 공간에 저장된 정보는 다음 분석에 참조가 될 수 있으며, 사전에서 대역어를 선택하는 근거가 된다. FCG에서는 이 정보 테이블을 명제 테이블(propositional table)과 양태 테이블(modal table)로 나눈다. 명제 테이블에 표시되는 정보는 문장성분의 정체다. 문장의 성분들이 문장에서 기능할 수 있는 최대의 후보는 테이블에 기능을 표시하여 칸으로 나누어져 있다. 그리고 각 칸에는 핵어가 되는 단어가 올 수 있으며, 핵어를 꾸미는 수식어는 핵어의 아래에 수식어열로 저장된다. 한편, 양태 테이블에는 Fillmore(1968)에 따라 부정, 시제, 법, 상 등의 정보가 표시된다. 양태 테이블도 각 기능마다 칸이 나뉘어져 있으나 명제 테이블과 달리 그 칸에는 문자열이 써지는 것이 아니라 기능이 문장에서 실현되고 있는지의 여부만 체크한다.

(9) 정보 테이블

가. 명제 테이블⁹⁾

	S	V	DO	P
Head	L1			
Mod1	L2			
Mod2	L3			
.....			

나. 양태 테이블

	NEG	TENSE			ST			
		Past	Present	Future	Decl	Imp	Que	Excl
	√	√			√			

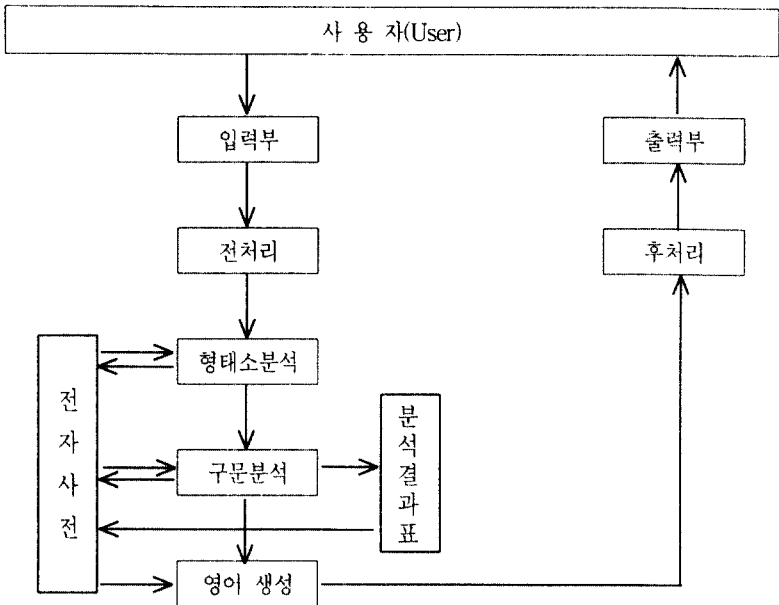
3.2 한영기계번역 시스템 구조도

본고의 한영기계번역 시스템은 다음과 같은 구조를 보이고 화살표에 따른 처

9) X⁰-이론에 근거한 수형도 표상 방법과는 달리 이 명제 테이블의 표상 방법은, 수식과 피수식 관계를 제외하면 위계적이지 않고 선형적이다. 그리고 목표어인 영어의 어순 구조를 고려한 테이블이다.

리 순서를 보인다.

(10) 한영기계번역 시스템 구조도



4. 한영기계번역을 위한 연어구의 처리 방법

4.1 개관

연어구의 구성요소인 X와 Y는 한영기계번역 과정에서 분석을 해서 영어로 뜻에 맞는 대역어로 생성해야 할 문자열이다. 기계번역에서 분석의 과정이 문자열의 중의성을 해결해 나가고 구조를 결정하는 과정이라면, 분석의 방법 가운데 하나는 공기하고 있는 문자열과의 관계성을 따지는 일이다. 이 과정에서 문자열의 문법적 자질이나 의미적 자질을 탐색할 수도 있겠으나 특정 어휘의 탐색이 분석의 중요한 방법론이 될 수도 있다. 이것이 바로 연어(collocate)를 이용한 분석 방

법이며, 우리는 그 과정을 4.3절에서 자세히 설명할 것이다.

한편 기능적 언어관계에 있는 표현들은, 구를 재해석하는 과정이 기계번역의 알고리즘에 포함되어야 한다. 즉, 구 구성요소인 X와 Y가 독립적으로 문장에서 기능하는 것이 아니라 통합하여 문장에서 기능하는 것이기에 분석 과정이 시작되기 이전에 별도의 기능을 가진 통합 구성으로 재해석하는 것이 필요하다. 따라서 이와 같은 기능적 언어구는 관용구와 흡사하게 어휘부에 등재될 수 있다는 가능성을 제시하여 주는 것이고 관용구와 유사하게 처리될 수 있다.

그리고 의미적 언어관계에 속하는 것이기는 하나 번역대상어와 번역목표어의 어휘가 직접적으로 일대일 대응이 불가능한 경우도 기능적 언어구와 마찬가지로 구 구성요소 X와 Y를 통합하여 재해석하는 것이 필요하다. 예를 들면, 한국어 '배가 고프다'는 영어에서 '배'와 '고프다'의 개별 어휘 대응체를 구성할 수 없고, '배가 고프다' 전체 구성이 'hungry'라는 대역어를 갖게 된다.

이를 정리하여 보면, 한영기계번역 과정에서 언어구가 처리될 수 있는 방안은 아래와 같다.

(11) 한영기계번역에서 언어구의 처리 방안

가. 중의성해결 $\left(\begin{array}{l} \text{형태 중의성해결} \\ \text{범주 중의성해결} \\ \text{구조 중의성해결} \\ \text{의미 중의성해결} \end{array} \right)$

나. 구 구성의 어휘적 재해석

4.2 처리의 장소

언어구는 X와 Y라는 특정한 어휘 차원의 결합으로 만들어진 구 구성이기에 X와 Y의 최종적인 언어구 구성은 자질 차원이나 부류 차원에서 정해지는 것이 아니고, 특정 어휘 차원에서 결정될 문제다. 그런데 분석 대상 문자열의 자질과 같은 일반적인 자연어의 집합을 이용하여 형태소 분석이나 구문 분석을 행하는 것이 보통이다. 가령,

(12) 가. '떡고', '떡었다'

나. '떡까지', '떡도'

(12)의 ‘먹’은 명사의 ‘먹’과 동사의 ‘먹-’으로 중의적이다. 중의적인 정보는 모두 사전에 기록되어 있고, ‘먹’의 중의성을 해결하는 과정이 곧 형태소 분석 과정이다. 이 때는 문자열 ‘먹’을 인접 후행하는 문자열의 특수한 자질을 탐색하여 중의성을 해결한다. 인접 후행 문자열이 [v, aff]이면 ‘먹’이 동사로, 인접 후행 문자열이 [n, aff]이면 ‘먹’이 명사로 해결될 것이다. 이러한 처리 방식은 형태소 분석 원리로 구축되어 있고, 실제 분석 대상이 되는 모든 문자열([n, v] 중의성을 지니는 모든 문자열)에 적용되는 전국적인 성격을 지닌다.

그렇지만 언어 정보를 이용하여 형태소 분석이나 구문 분석을 행할 때는, 언어 정보가 위에서 언급한 [n, aff]나 [v, aff]와 같이 자질 차원의 자연류 집합이 아니므로 분석 대상이 되는 모든 문자열에 적용되는 것이 아니라 개별 어휘 내부에서 처리되어야 하기 때문에 국지적이다.¹⁰⁾ 따라서 언어 정보를 이용한 분석 방법은 형태소 분석 원리나 구문 분석 원리로 구축되지 못하고, 개별 어휘의 사전에 분석 과정이 명세화되어야 한다.

지금까지 살펴본 바는 (11가)의 중의성 해결에 쓸모가 있는 것이다. 반면 (11나)의 구 구성이 어휘적으로 재해석되어야 하는 언어구의 경우, 중의성 해결과는 상관없이 언어구의 핵이 되는 어휘와 언어관계로 결합하는 특정한 어휘를 좌우로 탐색하여 그 어휘가 발견되면 어휘적으로 재해석하는 절차를 언어구의 핵이 되는 어휘 사전 내부에 두어야 한다.

이제 언어구의 처리는 모두 사전에서 이루어져야 한다는 것을 알았다. 이제 남은 일은 언어구의 처리를 위한 사전의 구조와 실제 처리 방식에 대해서 알아보는 일이다. 실제 처리 방식은 4.3절에서 논하기로 하고, 지금은 사전의 구조에 대해서 살펴보기로 하겠다.

본 기계번역기의 사전에는 어휘항목의 등재와 그 어휘항목의 어휘내향 정보를 표시하기 위하여 {어근사전}, {접사사전}, {숙어사전}, {특수문장사전} 등을 두었다. 이 중 {숙어사전}과 {특수문장사전}은 단어보다 큰 단위가 어휘항목으로 등재될 경우를 위하여 마련해 둔 것이다. 그리고 어절 중의성 해결을 위해 {음절사전}을 두었으며¹¹⁾, 기계번역기 사용자 편의를 위하여 {전문어사전}과 {사용자사

10) 본 기계번역 시스템에서는 이러한 자연류의 집합이 자질로 표시되어 있고, 실제 분석 절차에서는 이 자질을 탐색하고, 이를 비교하여 자질을 쓰거나 지우고 그 결과를 저장하는 자질 연산의 방법을 사용한다. 이 점이 언어 정보를 이용한 분석 절차와 차이가 나는 점이다.

11) 한국어 형태소는 대체로 음절 단위로 실현되므로, 형태소의 분리는 음절 단위를 기본으로

진)을 따로 두었다.

이러한 사전 가운데 언어구가 처리되는 사전은 {어근사전}이다. 이 사전에는 어휘항목의 표제어 정보와, 이 어휘항목의 통사, 의미 범주적 특성이 자질로 명세화되어 있다. 그리고 별도의 창에 통사범주와 의미 범주의 하위 범주 자질이 목록화되어 있어 형태소 분석이나 구문 분석 과정에서 이용될 수 있도록 하였다. 또한 본고에서 설명하는 사전이 한영기계번역을 위한 것이므로 영어 대역어가 기록되어 있다.

한편, 어휘항목의 문자열이 중의성을 가지고 있으면, 이를 해결하기 위한 처리부가 사전에 마련되어 있다. 이 처리부에는 형태와 범주 중의성 해결을 위한 처리부와 의미적 중의성을 해결하는 데 쓰이는 처리부가 각각 마련되어 있다. 또한 특정한 어휘항목이 탐색되어 어휘항목을 재해석해야 할 경우를 위하여 별도의 처리부가 있다. 언어구와 관련된 처리가 이루어지는 곳이 바로 중의성 해결을 위한 처리부와 어휘항목을 재해석하는 처리부다.

그런데 중의성을 위한 처리부와 어휘항목의 재해석을 위한 처리부가 항상 활성화되어 있는 것은 아니다. 만일 어휘항목이 중의성이 없거나 이 어휘항목을 재해석할 필요가 없는 경우에는 비활성화되어 있어야 효율적이다. 이를 위해 사전 내부에는 활성화 정보가 표시되어 있어서 해당 어휘항목이 분석 대상 문자열에서 발견되면 이 활성화 정보를 통해 처리부의 내부로 들어가게 된다. 자세한 것은 4.3절에서 설명하기로 한다.

이제 마지막으로 기계번역 전 과정에서 언어구 처리의 순서를 살펴보아야겠다. 앞서도 말했듯이, 언어구는 모두 사전에서 처리되어야 한다고 했다. 이미 (10)의 그림을 통해 사전이 형태소 분석, 구문 분석, 영어 생성에 모두 관여한다는 것을 알고 있다. 따라서 형태소 분석에 관여하는 언어구는 형태소 분석 이전에 사전에서 처리되어야 하고, 구문 분석에 관여하는 언어구는 구문 분석 이전에 사전에서 처리되어야 하며, 영어 생성에 관여하는 언어구는 한국어의 구문 분석이 모두 끝나고 대역어를 선택하는 과정에서 처리되어야 한다.

한다. 그러나 둘 이상의 형태소가 하나의 음절로 융합되는 것도 가능하므로, 때에 따라서는 음소 단위의 분리도 고려해야 한다. 그런데 음소 단위로 분리하여 분석 후보가 과생성되는 경우가 발생할 수 있는데, 이러한 비효율성을 줄이려고 형태소 융합 가능성이 있는 음절에 대해 가능한 모든 음절 분리 정보를 부여하여 만든 사전이 바로 음절 사전이다(황화삼 2002 참조).

4.3 언어구의 처리 방안

4.3.1 중의성 해결

이 소절에서는 (11가)와 관련된 언어구의 처리 방식을 제안한다. 분석 대상인 한국어는 형태, 범주, 구조, 의미의 중의성 해결이 문제가 된다. 각각의 중의성 문제를 예로 들어보자.

- (13) 가. /갈/: ‘가다’, ‘갈다’
 나. /꼴/: ‘꼴다’, ‘꼴’
 다. /우리가 일상으로 구사하는 언어와 행동/
 라. /먹다/

(13가)의 ‘갈’은 ‘가다’에 관형형어미 ‘을’이 붙은 형태일 수도 있고, ‘갈다’에 관형형어미 ‘을’ 붙은 형태일 수도 있어 형태적으로 중의성이 있는 음절이다. (13나)의 ‘꼴’은 동사 ‘꼴다’에 관형형어미 ‘을’이 붙은 형태일 수도 있으나 명사 ‘꼴’일 수도 있어 동사와 명사의 범주 중의성이 있는 어절이다. (13다)의 구는 다음과 같이 구조적으로 중의적일 수 있는 것이다.

- (14) 가. [우리가 일상으로 구사하는 [언어와 행동]]
 나. [[우리가 일상으로 구사하는 언어]와 행동]

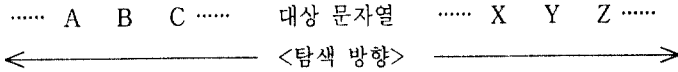
그리고 「표준국어대사전」(1999)에서 집필한 바에 따르면, (13라)의 ‘먹다’는 보조동사를 포함하여 총19개의 의미를 가진 단어다.

이렇듯, 분석 대상 문자열이 중의성을 갖는 것은 보편적이다. 따라서 기계번역의 성공 여부는 이러한 중의성의 문제를 잘 해결하는 데 달려있다고 해도 무방하다.¹²⁾ FCG의 기계번역기 시스템에서 채택하고 있는 중의성 해결 방법은 다음의 두 가지다.

12) 현재의 기계번역기 수준은 문장의 범위를 넘어서 텍스트 단위의 정보를 이용하거나 화용론적인 정보를 이용할 수가 없고, 거의 전적으로 문장 내부의 정보만을 이용하고 있다. 따라서 한국어 문장에 나타나는 모든 중의성을 해결할 수 없다는 게 현재 전반적인 기계번역 시스템의 한계이다.

- 선행 어휘항목의 자질이나 특정 형태 탐색하여 해결하는 방식
- 명제 테이블(P-table)의 어휘항목 형태나 자질을 탐색하여 해결하는 방식

첫 번째 방법은 분석 대상이 되는 문자열을 기준으로 선행하거나 후행하는 어휘항목의 자질이나 특정 형태를 탐색하여 중의성을 해결하는 것이다. 이는 구문 분석이 실제로 행해지기 전이나 구문 분석이 진행되고 있는 과정에서 택할 수 있는 방법이다.



예를 들어, ‘먹을’이라는 어절이 분석 대상 문자열이라 하자. 이 어절은 ‘먹[v, lex]+을[v, aff, unc]’의 결합일 수도 있지만, ‘먹[n, lex]+을[n, aff, thm]’의 결합일 수도 있어 범주 중의성을 보이는 경우다. 범주 중의성을 해결하기 위한 절차로 아래와 같은 스크립트 형태를 제시할 수 있다.

(15) ‘먹’의 중의성 해결 절차(1)¹³⁾

인접 선행하는 [unc, lex]가 있는지 확인하라.

ly : ‘먹’의 자질 중 [v, amb]를 삭제하라.

ln : ……

한편, 두 번째 방법은 논항 분석을 비롯한 구문 분석 과정이 끝나 그 분석 결과가 명제 테이블에 표시가 되면, 그것을 이용하여 중의성을 해결하는 방식이다. 예를 들어, ‘철수가 물을 맛있게 먹었다’의 문장에서 ‘먹다’의 의미는 ‘먹다’의 대상이 무엇이나에 따라 그 대역어가 결정된다. 먼저 분석 결과가 표시된 P-table을 제시하고 이 결과를 이용하여 중의성을 해결해 나가는 과정을 스크립트의 형태로 제시하면 아래와 같다.

13) ‘을’도 어미와 조사 자질을 [v, n, aff]처럼 한꺼번에 가지고 있어 중의적이나, ‘을’의 인접 선행 문자열이 [n]나 [v]로 정해지면 자동적으로 ‘을’의 중의성이 해결된다.

(16) '철수가 물을 맛있게 먹었다'의 P-table

		S	V	DO	
Head		철수	먹	물	
Mod			맛있게		

(17) '먹'의 중의성 해결 절차(2)

.....

P-table의 DO에 [n, liquid]항이 있는지 확인하라.

ly : '먹'의 대역어를 '/drink/drinks/drank/drun/drinking/'으로 결정하라.

In :

지금까지 중의성 해결의 예는 선후행 문자열의 자질과 P-table의 자질을 탐색하여 처리한 것이다. FCG에서는 자질뿐만 아니라 특정 문자열을 참고로 중의성 해결이 가능하다. 이것이 바로 연어 정보를 참고하는 방식이다. 우리는 (13나)와 (13라)예의 처리를 보일 것이다.

(13') 나. /골/: '골다', '골'

라. /먹다/

(13나)의 '골'의 형태는 동사 '골다'의 활용형일 수 있고 명사 '골'의 어근형이기도 하여 범주 중의적이다. 따라서 어근 사전에는 '골'이 최대한 가질 수 있는 범주 자질을 다 명시해 준다. 그리고 이 '골'이 분석 과정에서 중의성이 해결되어야 할 어휘항목이라는 명세화를 자질로 해 주어야 한다.

(18) /골/: [amb, n, v, lex, act, v3]

형태소 분석에 들어가기에 앞서 문자열 탐색을 하는 수순에서 문자열의 자질 중 [amb]를 발견하게 되면 곧바로 해당 어근 사전의 내부에서 중의성 해결의 절차가 있는 중의성 처리부로 들어간다. 즉, [amb]가 중의성 처리부로 인도하는 활성화 자질이다.

(19) '골'의 중의성 해결 절차

.....

선행하는 [n, lex]가 있고, 그 [n, lex] 중에서 '코'가 있는지 확인하라.

1y : [amb, n]을 삭제하라.

1n : 인접 후행하는 [v, aff]가 있는지 확인하라.

2y : [amb, n]을 삭제하라.

2n :

위의 스크립트 가운데 2y로 결과 표시되는 것은 이미 우리가 보았던 바이고, 1y로 결과 표시되는 것은 대상어 '골'을 기준으로 선행하는 언어 정보를 활용한 것이다. 이 언어 정보를 활용하여 범주 중의성을 해결할 수 있다.

다음으로 동사 '먹다'의 경우다. 가령, '철수는 욕을 많이 먹었다'에서 '먹다'는 이미 (17)에서 보인 대로 의미 중의성 해결 절차가 사전 내부에서 필요하다. 단, 이 처리 수순은 (15)의 중의성 해결과는 다르게 구문 분석이 끝난 이후에 분석 결과가 쓰여진 P-table을 참조하여 이루어지는 것이다. 그리고 의미 중의성 해결은 '골'의 [amb]와 유사한 [p_semy] 자질을 가지고 있어서, 구문 분석이 끝난 다음 대역어 선택 수순에서 참고한다. 이 때는 [p_semy]가 활성화 자질이다.

(20) '먹'의 중의성 해결 절차(3)

.....

P-table의 DO에 [n, lex]인 '욕'이 있는지 확인하라.

2y : '먹'의 대역어를 '/get/gets/got/gotten/getting/'으로 결정하라.

2n :

4.3.2 구 구성의 어휘적 재해석

이 소절에서는 (11나)와 관련된 언어구의 처리 방안을 제안한다. 기계번역 과정에서 번역대상어의 어휘가 번역목표어의 어휘로 그대로 대응되지 못하는 경우가 많다. 가령,

(21) 가. 나는 그 사람이 매우 두려웠다.

나. 나는 몹시도 배가 고프다.

(21가)의 ‘두렵다’는 번역목표어가 영어라면 두 단어 이상으로 번역되어야 할 한국어 어휘항목이고, 반대로 (21나)에서 ‘배가 고프다’는 한국어에서는 두 어휘항목 이상이지만 영어에서는 한 어휘항목으로 번역해야 하는 예다. 따라서 (21가)의 유형은 영어의 연어구의 예이고, (21나)의 유형은 한국어의 연어구의 예가 된다. 본고가 한국어 분석을 중심으로 논의하는 것이기에 여기서는 (21나)의 유형만 살피기로 한다.

(21나)의 ‘배가 고프다’는 한국어에서는 구 구성이지만 번역목표어인 영어의 구조 때문에 구 구성이 어휘로 재해석되어야 하는 예다. 이러한 연어구의 특징은 다음과 같이 정리해 볼 수 있다.

(22) 가. 두 어휘항목 이상이 하나의 독립된 대역어를 갖는다.¹⁴⁾

나. 연어구 구성요소 사이에 어휘항목이 개입할 수 있다. (예: 나는 배가 몹시도 고프다.)

‘배가 고프다’의 구가 VP이기 때문에 우리는 이 예를 ‘고프다’의 어근 사전에서 처리할 것이다.¹⁵⁾ ‘고프다’의 어근 사전에는 ‘고프다’의 고유 자질이 포함되어 있고, 아울러 ‘배가 고프다’의 구를 어휘적으로 재해석하기 위하여 활성화 자질이 표시되어 있다. 아래 (23)에서 [id_nc]가 활성화 자질이다.

(23) /고프다/: [v, lex, vl, id_nc]

14) 이러한 성격은 관용 표현과 같다. 본 기계번역기에서 관용 표현의 처리 방법은 세 가지다. 첫째, 전체 문장이 관용 표현일 때는 특수문장 사전에 통째로 등재한다. 둘째, 관용 표현의 구성요소가 항상 연결하여 출현하는 경우다. ‘눈감쌀 사이’가 예가 되겠고 이 관용 표현은 속어 사전에서 별도 처리한다. 셋째, 관용 표현의 구성요소가 연결하지 않고 떨어져 출현할 가능성이 있는 경우다. 이 경우는 앞으로 설명할 연어구의 처리와 똑같으나, 중의성 해결에는 관여하지 못한다.

15) ‘고프다’를 핵으로 삼는 것은 ‘배가 고프다’가 전체적으로 VP라는 통사론적 이유이기도 하겠지만, 만일 ‘고프다’를 현문자열에서 삭제하면 ‘고프-’에 인접하여 후행하는 어미를 처리할 수 없다는 형태론적 이유도 있다.

이제 구 구성을 어휘로 재해석하는 처리부에서 '배가 고프다'가 처리되는 과정을 살펴보겠다. 우리는 (22)에서 제시된 두 가지 특성으로 '배가 고프다'류의 언어구를 처리하기 위해서 언어구의 비핵이 되는 요소를 삭제하는 처리 방식을 제안한다. 즉, '배가 고프다'에서 '배'를 삭제하는 방식이다. 이와 같은 처리의 장점은, 어휘항목 대 어휘항목의 일대일 대응을 유지하게 하고, 언어구 구성요소 사이에 어휘항목이 개입되어도 처리가 가능하다는 점이다. 그리고 '배'를 삭제하고 남은 문자열 '고프다'에 대해서는 원래 '배가 고프다'가 가질 수 있는 범주 자질과 대역어 정보를 기록하여 재해석한다. 이를 스크립트 형태로 보이면 아래와 같다.

(24) '배가 고프다'의 언어구 처리 절차¹⁶⁾

선행하는 문자열 '배'가 있는지 확인하라.

ly : 현문자열에서 '배'를 삭제하라. 그리고 '고프다'의 자질을 [v, lex, vl, stal]로 치환하라. 그리고 '고프다'의 대역어로 '/hungry/'를 할당하라.

ln : ……

이와 같은 처리 방식은 문법적 언어구에도 그대로 적용된다. 우리가 살펴볼 예는 '-어야 하다'다. 이 예의 핵어는 통사론적 이유와 형태론적 이유에 의해서 '하다'가 되겠고, 곧 '하다'의 어근 사전에서 '-어야 하다'를 어휘적으로 재해석해야 할 것이다. 따라서 '하다'의 어근 사전에는 '하다'의 고유 자질과 더불어 '-어야 하다'의 언어구를 처리하기 위하여 [id_nc]가 활성화 자질로 표시되어 있어야 한다. '-어야 하다'의 언어구를 처리하는 방식을 스크립트 형태로 보이면 다음과 같다.¹⁷⁾

16) 이러한 처리 방식은 대부분의 문법적 언어구에도 적용된다.

17) 한 익명의 심사자가 아래의 예를 들어 '-어야 하다'를 규칙적으로 처리할 가능성에 대해 질문하였다.

(가) 길을 먼저 닦아야 상하수도룰 놓을 수 있다.

(나) 길을 먼저 닦아야 한다.

본 기계번역 시스템에 의하면 (나)의 '-어야'는 후행하는 '하다'가 없기 때문에 언어구로 처리되지 않는다. 또한 '하다'가 포함된 (가)와 마찬가지로 (나)의 '-어야'의 의미는 영어로 *must*나 *should*로 번역될 수 있어 같다고 볼 수 있다. 하지만 (가)에서 '-어야 하다'를 언어구로 처리하지 않는다면, (나)에서처럼 복문으로 처리될 수밖에 없습니다.

(가') [[길을 먼저 닦아야] 상하수도룰 놓을 수 있다]

(25) '어야 하다'의 언어구 처리 절차

선행하는 문자열 '어야'가 있는지 확인하라.

ly : 현문자열에서 '어야'를 삭제하라. 그리고 '하다'의 자질을 [aux]로 치환하라. 그리고 '하다'의 대역어로 '/must/'를 할당하라.

ln :

5. 요약 및 남은 문제

지금까지 우리는 FCG를 기반으로 한 한영기계번역기에서 언어구를 어떻게 처리하고 있는지를 살펴보았다. 언어구는 X와 Y가 언어관계를 이루는 표현이었고, 언어관계는 X와 Y라는 특정한 어휘(문법적 언어관계이면 X나 Y 중 어느 하나가 특정한 문법 형태소) 차원의 결합이라는 게 핵심이었다. 이러한 언어관계는 다시 언어구가 독립적인 기능을 표시하는가, 아니면 결합 어휘의 개념 구조를 하위명세화하는가에 따라 기능적 언어관계와 의미적 언어관계로 분류하였다.

이러한 언어관계의 분류는 한영기계번역기의 처리 방식에 영향을 미친다. 이를 정리하여 보이면, 다음과 같다.

- 한영기계번역에서 언어구의 처리 방안

가. 중의성 해결	(형태	중의성 해결
		범주	중의성 해결
		구조	중의성 해결
		의미	중의성 해결

(나') [[길을 먼저 댕어야] 하다]

(가')와 (나') 모두 두 개의 절을 가지고 있기 때문에 격 정보 분석을 두 번 거치게 된다. 이 과정에서 '하다'의 논항구조를 따져야 하고, 문장에 적합한 영어 대역어도 선택해야 하며, 생략된 논항 정보가 있으면 영어로 생성할 때 복원을 해 주어야 한다. 따라서 기계번역 처리 과정의 효율성이 떨어질 뿐만 아니라 올바른 번역을 보장하기도 힘들다. 그래서 언어구나 관용구를 최대한 코퍼스에서 많이 확보하여 사전에 저장해 주는 것이 기계번역의 성공률을 높이는 데 일조할 수 있다. 이는 기계번역을 실현하는 컴퓨터의 저장 공간이 확대되고 처리 속도도 빨라져서 예전에 가급적 규칙으로 처리하려고 하였던 구문이라든가 그 속에서 고정적인 표현(언어구나 관용구)이 발견되면 번역기의 사전에 저장하여도 문제가 되지 않기 때문이다.

나. 구 구성의 어휘적 재해석

언어 정보를 이용한 분석 방법은 자질 따위의 자연류 집합을 이용하는 형태소 분석 원리나 구문 분석 원리로 구축되지 못하고, 개별 어휘의 사전에 분석 과정이 명세화되어야 함을 알 수 있었다. 그리고 4.3절에서는 몇 가지 예를 들어 어근 사전에서 언어구가 처리되는 모습을 보였다.

그러나 우리가 제시한 언어구의 처리 방식이 완벽한 것은 아니다. 이는 (나)의 언어구 구성을 어휘적으로 재해석할 때 문제가 있다. (나)의 경우는 언어구의 비핵이 되는 어휘를 삭제하는 방식을 택한 것이다. 그런데 비핵이 되는 어휘가 언어구의 핵이 되는 어휘를 제외한 기타의 어휘와 문법적인 관계를 맺을 때는 삭제할 수가 없다. 가령, '배가 고프다'에서 '내 배가 고프다'처럼 '배'를 수식하는 어휘가 있으면 '배가 고프다'는 언어구로 처리되지 못한다.¹⁸⁾ 이는 실제 코퍼스상에서는 발견할 수 없었으나, 출현 가능성은 상존하는 것이기 때문에 별도의 처리가 필요하다.

참고문헌

- 강현화(1997). "[체언+용언]폴의 언어 구성에 대한 연구", 『사전편찬학연구』 8, 191-224, 연세대.
- 고창수(2002). 『자질연산문법이론』, 서울: 도서출판 월인.
- 김원경(2000). 『한국어 격 정보와 자질 연산 문법』, 고려대 박사학위논문.
- 金鎮海(2000). 『國語 連語 研究』, 慶熙大 博士學位論文.
- 서상규(2002). "한국어 정보 처리와 언어 정보", 『國語學』 39, 321-353, 국어학회.
- 宋正根(2002). 『명사와 동사 결합 구성의 유형론과 기계 번역』, 서울대 碩士學位論文.
- 유혜원(2002). 『국어의 격 교체 구문의 연구: 한영 기계 번역 시스템을 중심으로』, 고려대 박사학위논문.

18) <세종계획 21>의 일환으로 구축된 550만 어절 규모의 코퍼스를 고려대학교 민족문화연구원 전자텍스트연구소(<http://ikc.korea.ac.kr/cgi-bin/kwic/kwic.cgi>)에서 온라인으로 검색할 수 있다. '고프다'의 코퍼스를 검색한 결과, '배'의 수식어는 단 하나도 나오지 않은 것으로 보아, 실제 용례를 한영기계번역기로 구현할 때는 큰 문제가 되지 않을 것이다.

- 이동혁(1998). 「국어의 연어적 의미 연구」, 고려대 석사학위논문.
- 임근석(2002). “현대 국어의 어휘적 연어 연구”, 1-120. 『國語研究』 167.
- 최호철·이정식(1998). “자연 언어 처리를 위한 전자 사전 구축 방안”, 『어문논집』 37, 411-438, 고려대.
- 임홍빈(2002). “한국어 연어의 개념과 그 통사·의미적 성격”, 『國語學』 39, 279-311, 국어학회.
- 홍종선(2002). “임홍빈(2002)에 대한 토론”, 『國語學』 39, 355-360, 국어학회.
- 홍종선·강범모·최호철(2000). “한국어 연어 정보의 분석 응용에 관한 연구”, 『한국어학』 11, 73-158.
- 홍종선·강범모·최호철(2001). 「한국어 연어 관계 연구」, 서울: 월인.
- 황화상(2002). “한국어 형태소 분석기의 설계와 구현: 다음질 접사 정보와 음절 사진의 활용”, 『한국어학의 오늘과 내일』(2002 한국어학회 국제 학술대회 발표논문집), 701-711, 한국어학회.
- Benson et al.(1986). *The BBI Combinatory Dictionary of English*. Amsterdam: John Benjamins Publishing Company.
- Cowie, A. P. (eds.)(1998). *Phraseology: Theory, Analysis, and Applications*. Oxford: Clarendon Press.
- Dorr, B. J. et al.(1999). “A Survey of Current Paradigms in Machine Translation”, *Advances in Computers*, Vol 49, M. Zekowitz (Ed), London: Academic Press, 1-70.
- Fillmore, C. J.(1968). “The Case for Case”, *Universals in Linguistics Theory*. Holt: Rinehart and Winston, 1-88.
- Hausmann, F. J.(1985). “Kollokationen im deutschen Wörterbuch: Ein Beitrag zur Theorie des lexikographischen Beispiels”, in H. Bergenholtz and J. Mugdan (eds.), *Lexikographie und Grammatik*. Tübingen: Max Niemeyer, 118-129.
- Jackendoff, R.(2002). *Foundation of Language: Brain, Meaning, Grammar, Evolution*. Oxford and New York: Oxford University Press.
- Kim, Y. et al.(2001). “Collocation Dictionary Optimization Using WordNet and k-Nearest Neighbor Learning”, *Machine Translation* 16, Netherlands: Kluwer Academic Publishers, 89-108.
- Lipka, L.(1992). *An Outline of English Lexicology*. Tübingen: Max Niemeyer

Verlag GmbH & Co.

Mel'čuk, I.(1996). "Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon", in Wanner, L. (eds.), 37-102.

Mel'čuk, I.(1998). "Collocation and Lexical Function", in Cowie, A. P. (eds.), 23-54.

Mel'čuk, I. et al.(2001). "Towards a Lexicographic Approach to Lexical Transfer in Machine Translation (Illustrated by the German-Russian Language Pair)", *Machine Translation* 16, Netherlands: Kluwer Academic Publishers, 21-87.

Wanner, L. (eds.)(1996). *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam: John Benjamins Publishing Co.

부록

아래에 나열하는 자질의 목록은 본 한영기계번역 시스템에서 사용하고 있는 자질을 모두 나열한 것이 아니라 이 글의 논의 과정에서 사용한 자질만 목록화한 것이다.

- (1) 명사 어간 [n, lex]
- (2) 동사 어간 [v, lex]
 - 동작동사 [v, lex, act]
 - 상태동사 [v, lex, sta]
- (3) 관형사 [unc, lex]
- (4) 조사 [n, aff]
 - '을'류 조사 [n, aff, thm]
- (5) 어미 [v, aff]
 - 관형형어미 [v, aff, unc]
- (6) 조동사 [aux]

212 언어과학연구 제24집

이 동 혁

136-701 서울시 성북구 안암동 5가 1번지
고려대학교 민족문화연구원 기계번역연구실
전화번호: (02)3290-2499
전자우편: metacog@ikc.korea.ac.kr

투고논문접수일	2003년 2월 13일
논문심사일	2003년 2월 17일
심사완료일	2003년 3월 3일