

코퍼스를 이용한 효과적인 다의어 학습

김영미·서진희
(연세대학교)

1. 서론

의사소통중심 교수법에서는 다양한 상호활동들을 통해 의미전달능력, 즉 의사소통 능력을 키우게 되는데 의미전달에 있어 적절한 어휘의 선택과 그 의미에 대한 정확한 이해, 특히 이들 어휘가 사용되는 상황에 대한 올바른 이해가 중요하다. Wilkins(1972)는 '문법 지식 없이는 의미도출이 힘들고, 어휘지식이 없이는 어떤 의미도출도 불가능하다'라고 역설하였는데, 이는 의사소통능력에 주안점을 두고 있는 현행 교육에서 어휘교육이 갖는 중요성을 대변해 준다고 할 수 있다. 어휘를 문법적인 기능을 갖는 기능어와 실질적인 의미를 갖고 있는 내용어로 분류할 때, 내용어의 경우 상당수가 여러 개의 의미를 갖고 있는 다의어(polysemy)이고, 이들 다의어를 구성하는 여러 개의 의미 중 어떤 의미가 주된 의미(primary meaning)인지, 또 각각의 의미들이 어떤 문맥적 상황에서 실현되는지를 파악하는 것이 해당 어휘를 이해하는데 기본적인 일이라 할 수 있다.

본 연구는 우리나라 고등학교 1학년 영어 교과서를 대상으로 대표적인 다의어들을 선정하여 이들이 갖고 있는 여러 의미들의 구조를 파악하고, 이들의 분포를 분석하였다. 또 이 의미들의 분포를 영어권 또는 영어를 공용어로 하는 나라들에서 사용되고 있는 의미 분포와 비교하였다.

II. 이론적 배경

1. 다의어

어휘 교육에 있어 다의어는 교사와 학습자 모두에게 부담을 주는 대상이라고 할 수 있다. 다의어가 중의성을 내포하고 있어 부적절한 정보의 도출을 유발하기 때문이다 (Miller & Fellbaum, 1995). 게다가 가장 일반적으로 사용되는 어휘들이 가장 다의어적이라는 것이다. Webster 사전(Webster's 7th dictionary)상의 6만 단어 중 40%에 해당하는 21,488개의 단어가 둘이상의 관련된 의미를 갖는 다의어라는 사실은 다의어가 어휘 교육에서 차지하는 비중을 대변해 주는 숫자라 하겠다 (Byrd et al., 1987)(재인용, Ravin & Leacock, 2000).

다의어는 '여러 개의 관련된 의미를 갖는 단어'로 정의 된다. 외국어 다의어의 경우

다의어를 구성하는 여러 개의 의미들이 어떠한 형태로 구조화되는가를 파악하는 것이 해당 다의어를 이해하는 효과적인 방법일 것이다. 본 연구는 보다 설명력이 있다고 보여지는 의미망 구조 모형을 위주로 선정된 다의어들의 의미를 분류하였다.

의미망(meaning network) 구조는 한 낱말의 다양한 의미들은, 수평적으로는 원형(prototype) 의미와의 유사성을 통해 '확장(extension)' 관계를 맺으며 수직적으로는 공통성을 통하여 '도식(schematicity)' 관계를 맺는다는 이론이다.¹⁾

어휘교육에 있어 이러한 다의어의 의미구조에 대한 기본적인 이해는 사전에 나열되어 있는 여러 개의 의미들을 유기적으로 연계하고 나아가 사전에 나타나 있지 않은 의미들에까지 그 확장을 유추할 수 있는 기초를 제공한다는 점에서 의미가 있다고 하겠다.

다의어가 갖는 여러 개의 의미들 중 어떤 의미가 실제로 적용되었는가는 해당 다의어가 문장 내에서 어떻게 구성되느냐에 따라 그리고 어떠한 사회적 상황에서 사용되었느냐에 따라 결정된다. 이러한 현상에 대한 다양한 증거는 코퍼스로 대표되는 전산 언어학을 통해 발견할 수 있다. 코퍼스는 실제로 사용된 언어들의 표본을 모아놓은 것이기 때문이다.

2. 코퍼스 분석

흔히 '말뭉치'라고 번역되고 있는 코퍼스는 그 내용적인 측면에 있어 일반 언어학적 분석이 수행될 수 있는 문장 혹은 문장의 일부들의 모임(A collection of texts or parts of texts upon which some general linguistic analysis can be conducted)이라 정의된다(Meyer 2002). 코퍼스 분석 도구 중 콘코던서(concordancer)는 특정 단어가 그 단어가 포함되어 있는 실제 문장들과 함께 정렬되어 나타나 있는 결과를 일컬으며, 이를 통해 해당 단어가 문장 내에서 어떻게 사용되었는지를 살펴 볼 수 있다.

학교 수업에 있어 대부분의 경우 일차적인 학습 자료는 교과서이다. 경우에 따라 교과서가 유일한 학습자료일 수도 있다. 따라서 학교 교육에서 어휘 교육의 실태를 파악하기 위해 교과서를 분석하는 작업은 필수적이고 기본적인 절차라고 할 수 있다. 특히 어휘 교육에서 교과서가 차지하는 비중을 감안할 때 이들 교과서의 분석은 현행 교육의 문제점 및 개선방안을 제시하는데 중요한 역할을 할 것이다. 본 연구를 위해서 12종의 고등학교 1학년 영어교과서 전문을 코퍼스로 구축하고 영어권 국가에서 사용되는 표현을 실은 다른 코퍼스들과 비교 분석하였다.

III. 연구 방법 및 절차

1) 'tree' 와 같은 전형적인 다년생 나무의 개념으로 확장되고 수직적으로는 이들과 다른 형태를 갖는 ' (genealogical tree)' 까지 확대되는 경우는 의미망 구조로 설명되는 경우이다(, 1996).

1. 연구대상 및 도구

본 연구는 고등학교 1학년 영어교과서 12종을 연구 대상으로 실시되었다. 이들 교과서의 본문 전문을 코퍼스로 구축하였고 편의상 KHSEC(Korean High School English Corpus)로 명명하였다. 연구 도구로는 연구의 대상인 KHSEC 코퍼스를 비롯한 다른 코퍼스들과 이들 코퍼스를 분석하기 위한 도구인 WordSmith 4.0 소프트웨어이다. 다음은 이들 각각에 대한 설명이다.

1) KHSEC(Korean High School English Corpus)

법문사, 천재교육, 대한교과서, 현대영어사, 한국교육미디어, 중앙교육진흥연구소, 지학사, 장원교육, 금성출판사, 능률영어사(이찬승외), 능률영어사(이기동외), 시사영어사등에서 출간한 고등학교 1학년 영어 교과서 12종을 포함하며 약 10만 단어로 구성되어 있는 코퍼스이다.

2) ICE-Singapore(International Corpus of English-Singapore)

ICE-Singapore는 일백만개의 단어규모로 싱가포르에서 사용되고 있는 문어와 구어로 구성되어 있다. 본 연구에서는 ICE Singapore의 문어부분(45만 단어규모)이 사용되었다.

3) BNC(British National Corpus)

British National Corpus(BNC)는 1억 단어 규모의 코퍼스로 현대 영국 영어가 사용되고 있는 신문, 학술서적, 소설, 잡지 등 다양한 분야의 문어텍스트와(전체의90%) 비공식적인(informal) 대화내용, 자원자들에 의해 녹음된 내용, 정부 회의내용, 라디오방송 내용 등의 구어텍스트(전체의 10%)로 구성되어 있다. 본 연구에서는 약 1백만 단어로 구성된 BNC sampler(문어부분)(이하 BNC Sampler)를 사용하였다.

4) Collins COBUILD Learner's Dictionary

1987년에 출판되기 시작한 Collins COBUILD(Collins Birmingham University International Language Database) English Dictionary(이하 Collins COBUILD)는 코퍼스를 이용한 최초의 사전이라는 점에서 획기적이다. Collins COBUILD의 코퍼스인 ;The Bank of English는 2억 5천만단어규모이며 BNC와 마찬가지로 다양한 분야의 자료로 구성되어 있다.

5) WordSmith 4.0

WordSmith는 별도로 구축된 코퍼스들을 분석할 수 있는 도구들을 제공해 주는 소프트웨어로써 영국 Liverpool대학의 Mike Scott교수에 의해 개발되었다. 일반 컴퓨터에 탑재하여 사용할 수 있다. WordSmith 4.0의 자세한 기능은 WordSmith의 website(<http://www.lexically.net>)에서 알아 볼 수 있다.

2. 분석방법

1) 단어의 선정

먼저 WordSmith의 WordList기능을 이용하여 KHSEC내의 전반적인 단어의 분포에 대해 분석하여 보았다. 코퍼스를 이루는 각기 다른 단어는 7,155개에 이르며 이들 7,155개의 단어가 전문에 걸쳐 총 100,106번에 걸쳐 나타났음을 알 수 있었다. KHSEC의 TTR(type/token ratio) 지표는 7.2인데 이 수치는 전체 코퍼스를 100개의 단어(token)로 이루어진 것으로 가정할 때, 약 7개의 서로 다른 단어(type)들이 반복 사용되어 100개의 단어를 이룬다는 것을 의미한다. 분석대상 어휘를 선정하는 방법으로는 양적인 방법과 질적인 방법을 함께 사용하였다. 본 연구에서 적용하는 양적인 방법이라 함은 해당어휘의 발생 빈도수가 5회-50회 사이에 발생한 다의어중 하위의 미의 개수가 적어도 3개 이상인 다의어를 대상으로 하는 것을 말한다. 질적인 방법이라 함은 이렇게 추출된 다의어중 그 의미들 간의 구조에 대한 이해를 필요로 하는 다의어를 주관적으로 선택하는 것을 말한다.

2) 분석

(1) concordance 분석

선정된 어휘들의 사전적 정의 및 의미 분포를 Collins COBUILD에서 살펴보고 이를 기준으로 KHSEC, BNC Sampler, ICE-Singapore에 나타난 자료들을 비교분석한다. 선정된 단어들은 다의적인 의미를 갖고 있으므로 문맥에 따라 달리 해석될 수 있다. 따라서 각각의 의미가 사용된 빈도수에 따른 우선순위를 참조할 수 있다면 어휘 교육의 효율을 높일 수 있을 것이다. 예를 들면, reservation(s)의 경우 대다수의 학습자에게 예약이라는 의미로 인식되고 있는데 이러한 사실은 KHSEC 분석결과에도 그대로 반영되고 있다. 표 1에 나타나 있듯이 KHSEC내에서 추출된 콘코던스 결과는 3건이며 이들 모두 예약의 의미로 사용된 것이다.

1 out is your plane ticket. Always make a reservation in advance. Many discount
 2 discount fares require that you make a reservation 30days before your trip
 3 are usually \$15~25 a night reservations if you can. When you want

표 1. concordance :KHSEC ‘reservation(s)’

이에 반해 ;The Bank of English; Corpus를 기초로 작성된 Collins COBUILD에는 표 2 에서와 같이 ;reservation(s);이 의심 ;, 예약 ;,보호 구역 ;의 의미 순으로 정의 되어 있다.

res · · · **on** /rezə'veɪʃən/ **reservations**. 1 if you have reservations about something, you are not sure that it is entirely good or right. *My main reservation about his film was the ending.* 2 If you make a **reservation**, you arrange for something such as a table in a restaurant or a room in a hotel to be kept for you. 3A **reservation** is an area of land that is kept separate for a particular group of people to live in. *Sixty percent of the 700 Native Americans who live on the reservation are unemployed.* 4 See also **centralreservation**.

표 2. CollinsCOBUILD ‘reservation’ 3의

KHSEC, ICE-Singapore, BNC의 콘코던스 분석 결과를 정리하면 표 3과 같다.

Collins COBUILD	KHSEC	BNC Sampler	ICE Singapore	notes
1. 의심	0	12	1	without reservation
2. 예약	3	4	5	make a reservation
3. 보호구역	0	4	0	Indian reservation

표 3. ‘reservation(s)’ 3코퍼스별 의미분포

(2) 의미 구조 파악

코퍼스 분석과 함께 의미 구조의 분석이 이루어 졌다. 다의어로 선정된 ;free'의 경우 Collins COBUILD상에는 표 4 에서와 같이 17개의 의미로 정의되고 있다. 17개의 의미중 동사로 분류된 5, 9, 10, 14, 16은 제외하고2), 형용사로 분류된 12개만을 분석대상으로 선정하였다.3) 표 4와 같이 다의어를 구성하는 세부 의미들을 분류하고 나면 다의어 free는 Collins COBUILD 상에서 대략 ;no pay', 'no control', 'not present', 'not occupied'같은 포괄적인 개념의 순으로 분류 될 수 있다는 것을 알 수 있다.

2) 동사의 경우 다른 품사에 비해 문법적인 제약에 대한 의존도가 높기 때문이다.

3) 본 연구에서는 다의어의 범주를 단일 품사에 국한한다는 주장을 따른다.

순서	의미	예문	의미 분류	비고
1	공짜의, , 돈을 내지 않는	free of charge, free brochure	no pay	
2	대가 없는()	worked for free	no pay	
3	자유로운, 규칙이나 관습에 얽매이지 않는	be free to do something	no control	
4	허락하는 (permissive)	feel free to ask somebody	no control	
5	면제하다, 구하다	free		동사
6	특정한 물질을 포함하지 않은	free of fear, free from fear, salt-free diet	not present	
7	세금이나 이자가 없는	tax-free, duty-free	not present	
8	자유로운()	set free, a free man	no control	
9	놓아주다	free a prisoner		동사
10	자유롭게 하다, 다른일을 할 수 있게 하다	free up your computer for other tasks		동사
11	() , 한가한	free time	not occupied	
12	(,) 비어있는	this seat is free	not occupied	
13	매어있지 않은, 묶여있지 않은	pull his arm free	no control	
14	풀어주다()	cut through the driver belt to free him		동사
15	쓰지 않는, 사용하지 않고 있는	his hand groped for the switch on the bedside lamp	not occupied	
16	자유를 주다.	to give someone a free hand		동사
17	관대한, , 활수한	free with advice, free with money	no pay	

표 4. *free*의 의미 정리

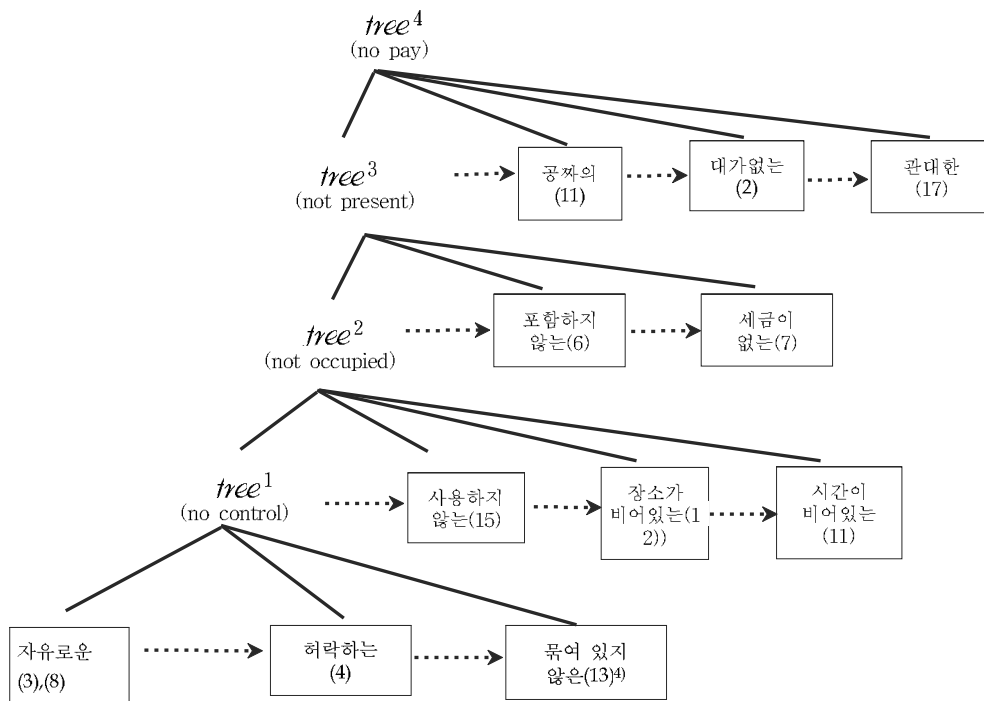


그림 1. *free*의 의미망 구조

그림 1은 표 4의 분석결과를 토대로 의미 구조를 분석해 본 것이다.

4) 괄호안의 숫자는 표 4

(3) 코퍼스별 어휘 분포

표 5 는 분석대상으로 선정된 6개 단어에 대한 코퍼스별 어휘 분포 이다.

대상단어 ()	Collins COBUILD	normalized ⁵⁾ (raw ⁶⁾) data			notes
		KHSEC	BNC Sampler	ICE Singapore	
reservation ()	1. 심	0(0)	12(12)	2(1)	without reservation
	2. 약	30(3)	4(4)	11(5)	make a reservation
	3. 호구역	0(0)	4(4)	0(0)	indian reservation
free ()	1. 짜의	60(6)	51(51)	53(24)	free night
	2. 유로운	70(7)	280(280)	133(60)	free trade
	3. 는	50(5)	47(47)	60(27)	smoke free
	4. 어있는	110(11)	7(7)	7(3)	free seat free time
race ()	1. 종	130(13)	79(79)	0(0)	competition
	2. 쟁	70(7)	19(19)	38(84)	people, tribe
smart ()	1. 련된	0(0)	10(10)	2(1)	fashionable
	2. ()	100(10)	10(10)	9(4)	street smart
	3. 퓨터의	0(0)	1(1)	9(4)	smart card
minute ()	1. ()	250(25)	229(229)	107(48)	60 minutes
	2. 은 시간	180(18)	72(72)	49(22)	just a minute
	3. ()	0(0)	2(2)	0(0)	meeting minutes
measure ()	1.	0(0)	3(3)	42(19)	amount
	2. / 표기준	0(0)	1(1)	24(11)	degree extent standard, criteria
	3. / 안	50(5)	35(35)	142(64)	particular action, bill
	4. / 측정도구	0(0)	9(9)	29(13)	act of measuring measuring device

표 5. 코퍼스별 어휘 분포

5) 비교하는 코퍼스들의 규모가 다르기 때문에, 집계된 수치를 정규화 (normalize)

BNC Sampler † 기준으로 하여 각 코퍼스로부터

6) KHSEC ICE Singapore

IV. 결론

분석 결과, 우리나라 영어 교과서에 나타난 다의어들의 의미 분포는 다른 영어권 국가에서의 의미 분포와 차이를 보였다. 본 논문은 이러한 차이가 학습자들의 어휘 교욱에 영향을 미칠 것이라는 전제하에 다음과 같이 제안하였다.

첫째, 교과서 내에서 사용하는 어휘의 선정에 있어 실제성(authenticity)여부의 검증이 선행되어야 한다.

둘째, 학생들이 다의어의 여러 의미에 노출될 수 있도록 교과서 내에 다양한 문맥을 포함시켜야 할 것이다.

셋째, 여러 가지 학습 활동을 통하여 학생들 스스로 어휘의 의미나 사용법을 발견할 수 있도록 교과서를 구성하도록 하여야 할 것이다. 코퍼스를 이용해 학생들 스스로가 실제 발화된 표현들을 통해 어휘의 의미를 파악하게 하는 것도 학습활동의 일부로써 활용될 수 있을 것이다.

참고문헌

- 고광윤. (2005). ;한국적 영어학과 영어 코퍼스의 활용; *영어학연구* 19: 21-35.
- 임지룡. (1996). ;다의어의 인지적 의미 특성; *언어학* 18: 229-261.
- Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus Linguistics*. Cambridge: Cambridge University Press.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar. Vol. 1: Theoretical Prerequisites*. Stanford, CA: Stanford University Press.
- Lewis, M. (2000). *Teaching Collocation: Further Developments in the Lexical Approach*. England: Language Teaching Publications.
- Meyer, C. F. (2002). *English Corpus Linguistics: An introduction*. Cambridge: Cambridge University Press.
- Wilkins, D. A. (1972). *Linguistics in Language Teaching*. London: Arnold.

김영미

e-mail : ymkz@naver.com

Mobile : 010-7147-0890

서진희

e-mail : jhsuh@yonsei.ac.kr

Mobile : 033-760-2741