

코퍼스와 영어교육*

오선영
(서울대학교)

Oh, Sun-Young. (2004). Corpus and English language education. *Foreign Language Education Research*, 7, 1-38.

Recently, the use of corpus has been very popular in almost all areas of language research, as a means of exploring the actual patterns of language use. The main goal of this paper is to show the possibilities of and advocate exploiting corpus for English language education. For this goal, it first provides a basic understanding of corpus by describing the types of corpus, issues in corpus design, and the use of a concordancing program in analyzing corpus. This is followed by the discussion of three different ways in which corpus can contribute to English language education. First, the content of English teaching can be improved by the more accurate and detailed description of the target language that is provided by corpus-based research. Secondly, corpus analysis itself can be adopted as a language teaching methodology, in which learners discover various facts about English by themselves. Finally, analysis of learner corpus can help us diagnose and solve the problems that the learners typically have.

I. 서론

최근 들어 언어 사용의 실제 양식을 탐구하는 수단으로서 코퍼스 언어학(Corpus Linguistics)이 크게 각광을 받고 있으며, 이와 함께 외국어 교육 분야에서도 코퍼스를 이용하여 교수 및 학습의 효율성을 증진시키는

* 본 연구는 2004년도 서울대학교 교육종합연구원 지원금으로 이루어졌음.

방안에 대한 관심이 급증하고 있다. 여기서 코퍼스란 특정 원칙에 입각하여 수집된 뒤 컴퓨터에 저장되어 있는 방대한 양의 문어(written language) 혹은 구어(spoken language)의 언어자료를 가리키는 말이다. 흔히 코퍼스 연구는 이러한 전자 코퍼스가 사용되기 시작한 1960년대 초반부터 이루어지기 시작했다고 생각하는 경우가 많다. 그러나 사실은 그보다 훨씬 오래 전부터 비전자(non-electronic) 코퍼스에 기초한 언어 분석의 전통이 다양한 형태로 존재해 왔다 (Kennedy, 1998). 예를 들면 18세기부터 성서를 코퍼스로 하여 이에 대한 주석을 달거나 비평을 하기 위해 콩코던스(concordance), 즉 용어 색인을 만들었으며, 이 후에는 여러 가지 학문적 목적으로 셰익스피어 같은 저명한 작가들의 작품에 대해서도 콩코던스가 만들어졌다. 뿐만 아니라 17세기 초반부터 사전 편찬시 단어의 의미를 예시하기 위해 작가들이 쓴 실제 문장들을 이용하였으며 (e.g., Samuel Johnson의 Dictionary of the English Language), 71년이란 긴 세월이 걸려 1928년에 비로소 완성된 옥스퍼드 영어사전 (Oxford English Dictionary) 역시 고전 문학 작품들로 이루어진 코퍼스에 기반을 두고 만들어졌다. 20세기 전반에 들어서는 언어 교육적인 목적을 위해 코퍼스가 이용되기도 하고 (Thorndike, 1921; Thorndike & Lorge, 1944), 모국어 습득 연구들에서도 어린이와 부모들의 실제 발화를 코퍼스로 이용하였다.

이러한 코퍼스 연구는 1950년대 Chomsky의 등장과 함께 언어학의 방향이 실증주의(empiricism)에서 이성주의(rationalism)로 선회하면서 큰 타격을 입게 된다 (McEnery & Wilson, 1996). Chomsky는 코퍼스가 언어학자에게 결코 유용한 도구가 될 수 없다고 주장했는데, 그 이유는 언어학자는 언어수행(linguistic performance)보다 언어능력(linguistic competence)을 연구의 대상으로 삼아야 한다고 보았기 때문이다. 언어능력은 잠재적이고 내재적인 언어 지식인 반면 언어수행은 특정 상황에서의 언어사용, 즉 언어능력의 외적 증거라 할 수 있다. 이러한 입장에서 보면 코퍼스는 본질적으로 구체적인 발화의 수집물로서 수행 자료이고, 따라서 인간의 내재적인 언어능력을 밝히는 데에는 큰 도움이 될 수 없다는 것이다. 그리하여 Chomsky를 따르는 언어학자들은 코퍼스 대신 내성(introspection)이나 직관(intuition)을 언어능력에 대한 증거로서 사용하게 된다.

그러나 이러한 Chomsky의 비판에도 불구하고 코퍼스 연구의 맥이 완

전히 끊긴 것은 아니었다. 예를 들면 음성학 분야나 모국어 습득을 연구하는 쪽에서는 계속해서 자연적으로 관찰된 자료가 이론을 뒷받침하는 주요한 증거로 사용되었다 (McEnery & Wilson, 1996). 내성적 자료는 사실상 여러 가지 한계점을 지니고 있는데 (McEnery & Wilson, 1996), 무엇보다도 내성적 판단은 관찰이 불가능하므로 입증하기가 어렵다. 또한 내성적 자료는 인위적이라서 실제 코퍼스에서 발견되는 증거와는 동떨어진 경우가 많으며, 특정 단어나 문법 구조의 빈도(frequency)에 대해서 구체적인 정보를 제공해 줄 수도 없다. 반면에 코퍼스는 보다 과학적인 접근법으로서 그 결과를 객관적으로 입증할 수 있고 다양한 양적(quantitative) 정보를 제공해 줄 수 있다는 장점을 가지고 있다. 초기 코퍼스 연구는 수작업으로 이루어져야 했기 때문에 시간과 비용 면에서 비효율적이고 오류 발생 비율도 높았다. 하지만 1960년대 이후로 컴퓨터 및 소프트웨어의 발달과 보급으로 인해 이러한 제약점을 극복할 수 있게 되었고, 이는 코퍼스 언어학 부활에 박차를 가하게 되었다. 1961년 Francis와 Kučera는 최초의 전자 코퍼스로 불리는 Brown Corpus를 구축하였고, 이후 많은 전자 코퍼스들이 계속 구축되면서 코퍼스 언어학은 활기를 되찾게 되었다.

코퍼스를 다양한 컴퓨터 도구와 함께 활용하면 이전에는 불가능했던 언어 사용에 관한 여러 가지 문제들을 연구할 수 있다. 본 논문은 코퍼스에 대한 기본적인 이해를 도모하고 코퍼스가 영어교육에 어떻게 기여할 수 있는지 논의하는 것을 주 목적으로 한다. 이를 위해 먼저 2장에서 코퍼스의 유형, 코퍼스 설계 시 고려할 문제 및 콩코던스 프로그램 이용 방법 등을 간단히 소개한 뒤, 3장에서 본격적으로 코퍼스가 영어 교육에 활용될 수 있는 방식을 교육 내용, 교수 방법, 그리고 학습자 언어 연구라는 세 가지 측면에서 살펴보도록 하겠다.

II. 코퍼스의 유형과 설계 및 이용

코퍼스는 그 구축 목적과 포함된 텍스트의 유형에 따라서 크게 일반 코퍼스(general corpus)와 전문 코퍼스(specialized corpus)로 양분될 수 있다. 일반 코퍼스는 언어를 폭넓게 대표하려는 목적으로 만들어지기 때문

에 서로 다른 유형의 텍스트를 많이 포함하고 있으며, 구어와 문어 둘 중 하나로만 이루어진 경우도 있고 둘 다를 포함할 수도 있다, 일반 코퍼스는 일반적인 언어 자질들을 연구하고 언어교육이나 번역을 위한 참고 자료를 만들거나 전문 코퍼스와 비교 시 기준선을 제공하기 위해 활용된다. 앞서 소개된 Brown Corpus를 비롯하여 Lancaster-Oslo/Bergen (LOB) Corpus, British National Corpus (BNC) 등의 코퍼스들이 모두 이에 속한다.¹⁾ 반면 전문 코퍼스는 그 명칭에서 알 수 있듯이 전문적이고 특수한 연구 목적을 염두에 두고 설계된 코퍼스로서 특정한 유형의 텍스트들만으로 구성된다. 전문 코퍼스의 예로는 영어의 역사적 변천 과정을 연구하기 위해 750년부터 1700년까지의 문어 텍스트를 수집하여 구축한 Helsinki Corpus나 미국 대학 환경 내에서의 구어만으로 이루어진 Michigan Corpus of Academic Spoken English (MICASE), 그리고 모국어 습득하는 어린이들의 언어를 담고 있는 Child Language Data Exchange System (CHILDES) Database 등을 들 수 있다. 이외에 영어가 사용되는 여러 국가들의 지역 방언 연구를 위해 구축된 International Corpus of English (ICE)도 전문 코퍼스에 속한다고 할 수 있다.

한편 코퍼스는 원시 코퍼스(raw corpus)와 주석 코퍼스(annotated corpus)로 분류해 볼 수도 있는데, 전자는 텍스트를 수집한 뒤 다른 언어 정보를 첨가하지 않은 코퍼스이고, 후자는 여러 가지 형태의 정보를 추가로 첨가시킨 코퍼스이다. 가장 흔하게 첨가되는 정보는 텍스트 내 각 단어의 품사인데, 이렇게 품사 태깅(tagging)이 이루어진 코퍼스는 둘 이상의 품사로 기능하는 단어의 한 가지 품사로서의 사용만을 검색하고자 할 때,²⁾ 혹은 어떤 문법 구조들이 함께 발생하는지 알아보는데 매우 유용하다. 품사뿐만 아니라 음성학적, 혹은 음운적 주석을 달수도 있고 또 문장

1) Brown Corpus는 각 2,000 단어로 이루어진 500개의 문어 미국영어 표본이 합하여 전체 100만 어절을 이루고 있으며, 크게 정보제공을 위한 산문(informative prose)과 상상적 산문(imaginative prose)이라는 두 범주에 속하는 총 15개 텍스트 유형으로 분류된다. LOB Corpus는 Brown Corpus와 동일 기준을 적용하여 구축한 문어 영국영어 코퍼스이고, 영국영어를 대표하기 위해 1990년대에 구축된 BNC는 1억 어절 규모로 문어(90%) 뿐만 아니라 구어(10%)에 속하는 다양한 장르의 텍스트를 포함하고 있다.

2) 예를 들어 'like'라는 단어는 동사, 전치사, 부사, 담화표지 등으로 다양하게 사용될 수 있는데 이 중에서 동사로 쓰인 'like'만 검색하여 연구하고 싶은 경우가 이에 해당된다.

의 통사적인 정보를 첨가하는 파싱(parsing)도 있는데 이는 그 과정의 복잡성으로 인해 태깅에 비해서 그리 많이 이루어지지 못하였다.

코퍼스 언어학에서 가장 중요한 요소 중 하나는 바로 코퍼스 설계(design)라고 할 수 있다 (Biber, 1990). 코퍼스 설계는 코퍼스를 통해 이루어지는 모든 분석 및 결과의 타당성과 신뢰도에 지대한 영향을 미치기 때문이다. 코퍼스 설계와 관련하여 고려해야 할 중요한 요인들로는 코퍼스의 크기, 대표성(representativeness)과 균형(balance), 영속성(permanance) 등이 있다 (Hunston, 2002; Kennedy, 1998). 먼저 코퍼스의 크기는 연구 목적과 밀접한 관련이 있다. 예를 들어 어휘 관련 연구가 목적이라면 많은 양의 단어들과 각 단어의 서로 다른 의의(sense)들이 잘 드러나 있도록 코퍼스의 크기가 커야 한다. 백만 어절의 코퍼스는 빈도수가 낮은 어휘항목들에 대해서는 신뢰할 만한 정보를 제공해줄 수 없다. 반면 문법 연구를 위해서라면 대부분의 경우 이 정도 크기의 코퍼스만으로도 충분한데, 이는 어휘 항목에 비해 문법 구조들이 훨씬 더 종류가 한정되어 있고 상대적으로 빈도수는 높기 때문이다. 대표성이란 코퍼스에 포함된 텍스트가 한 언어 또는 장르를 얼마나 잘 대표하는가 하는 것으로, 대표성 있는 코퍼스에 근거한 분석은 그 언어나 장르 전체로 일반화될 수 있다 (Leech, 1991). 대표성은 여러 가지 범주에서 생각해 볼 수 있다. 즉 서로 다른 언어사용역 (소설, 비소설, 일상적 대화, 방송 등), 담화 방식 (독백, 양자간 대화, 다자간 대화), 주제 (예술/과학, 국가적/지역 뉴스), 화자나 필자에 대한 정보 (국적, 성별, 나이, 교육 수준, 사회 계층, 모국어/방언) 등 다양한 차원에서의 대표성을 고려하여 코퍼스에 포함시킬 텍스트를 선정해야 한다 (Reppen & Simpson, 2002). 또한 이렇게 서로 다른 범주들 사이에 균형이 이루어져야 한다. 그러나 각 범주에 속하는 텍스트의 양을 동일하게 한다고 해서 꼭 코퍼스의 균형이 이루어지는 것은 아니며 (Kennedy, 1998),³⁾ 실제 코퍼스 구축을 할 때에는 여러 가지 요인들을 복합적으로 고려해서 결정을 내리게 된다. 영속성이란 코퍼스가 정적(static)

3) 일반 코퍼스의 경우 문어와 구어 텍스트간의 균형을 생각해 볼 수 있는데, 보통 우리가 하루에 접하는 언어의 양은 문어보다는 구어가 훨씬 더 많으므로 문어보다 구어가 더 많이 포함되어야 한다고 주장할 수 있다. 그러나 신문과 같은 문어 텍스트는 일상 대화에 비해 훨씬 더 많은 사람들이 접하게 된다는 점에서 반대의 주장도 가능하다.

인가 동적(dynamic)인가와 관련된 개념인데, 대부분의 코퍼스는 일단 구축되고 난 뒤에는 그 구성의 변화가 없는 정적 코퍼스이다. 반면 동적 코퍼스는 시간이 흐름에 따라 언어 사용양식이 어떻게 변화하는지 감시할 수 있도록 새로운 텍스트가 정기적으로 추가되는 코퍼스로서 모니터 코퍼스라고도 한다 (Sinclair, 1991).⁴⁾ 영속성과 관련된 결정 역시 코퍼스 설계 단계에서 이루어져야 한다.

영어교육 분야에서 원어민(native speaker)의 언어 사용을 연구하기 위해 코퍼스를 이용한다면 많은 경우 코퍼스를 직접 구축할 필요는 없다. 연구자가 아주 큰 부담 없이 구입할 수 있는 일반 코퍼스와 전문 코퍼스들이 이미 상당수 구축되어 있으므로 이를 이용하면 된다 (부록 참조). 그러나 연구목적이 특수하여 기존의 코퍼스를 사용할 수 없다면 연구자가 직접 코퍼스를 구축해야 할 것이고 (예를 들어 특수 목적으로 영어를 배우는 학습자들을 위해 그들이 주로 접하게 될 언어자료를 수집한다거나 또는 특정 집단 학습자의 언어 분석을 위해 그들의 글이나 말을 수집하여 코퍼스로 만드는 경우), 이런 경우 그 연구 목적에 부합하도록 코퍼스 설계에 신중을 기해야 할 것이다. 코퍼스를 직접 구축하지 않는 경우에도 위에서 살펴본 코퍼스 설계와 관련된 문제들을 잘 고려하여 사용할 코퍼스를 선택해야 한다.

코퍼스 연구에서는 컴퓨터 도구들을 이용하게 되는데, 컴퓨터는 주로 특정 단어나 단어 연쇄, 문법 구조, 혹은 같은 품사에 속하는 단어들을 코퍼스 내에서 모두 검색해 주는 기능을 담당한다. 다시 말해 연구 대상인 단어나 구조의 모든 예들을 문맥과 더불어 찾아주는 것인데 이런 과정들은 콩코던스 프로그램을 이용함으로써 가능해진다. 일반인들이 사용하기에 편리한 콩코던스 프로그램으로는 MonoConc, WordSmith Tools 등이 있다 (관련정보는 부록 참조). 이러한 프로그램을 통해서 얻을 수 있는 가장 기본적인 정보는 단어목록(word list)인데, 이것은 코퍼스 내에서 발견되는 모든 단어들의 목록을 말하며 알파벳순 혹은 빈도순으로 정렬될 수 있다. [그림 1]은 MonoConc 프로그램으로 도출해낸 브라운 코퍼스의 빈도순 단어목록 화면인데 이로부터 'the, of, and, to, a' 가 가장 많이 사용

4) 대표적인 모니터 코퍼스로는 The Bank of English가 있으며 이 코퍼스는 2001년 기준으로 4억 어절 규모이다 (Hunston, 2002).

된 다섯 단어임을 알 수 있다.

[그림 1: 단어목록 화면]

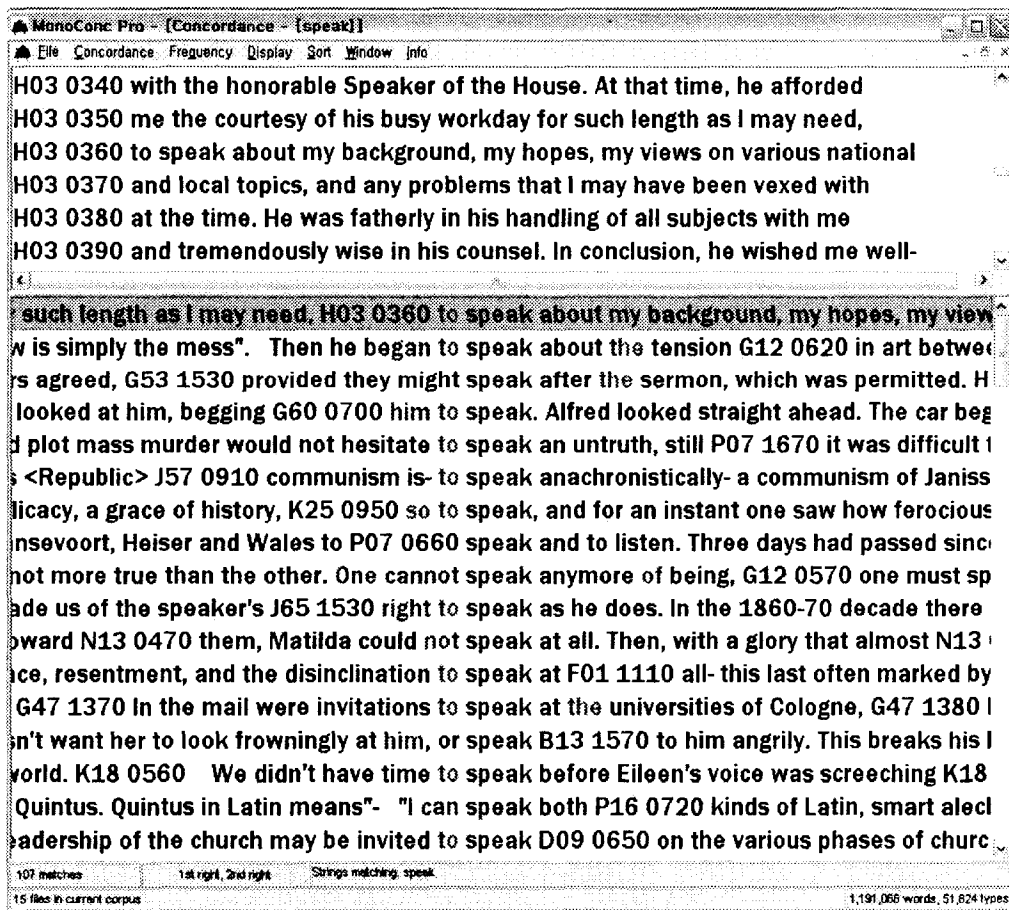
Count	Pct	Word
69377	5.8248%	the
36146	3.0348%	of
28708	2.4103%	and
25926	2.1767%	to
23321	1.9580%	a
21222	1.7818%	in
10517	0.8830%	that
9993	0.8390%	is
9755	0.8190%	was
9491	0.7968%	he
9435	0.7921%	for
8655	0.7267%	it
7244	0.6082%	with
7211	0.6054%	as
6955	0.5839%	his
6688	0.5615%	on
6325	0.5310%	be
5351	0.4493%	at
5272	0.4426%	by
5135	0.4311%	i
5114	0.4294%	had
5110	0.4290%	this

15 files in current corpus 1,191,056 words, 51,824 types

[그림 1]에서 볼 수 있듯이 각 단어의 발생 빈도에 대한 정보도 얻을 수 있다 (즉 ‘the’는 69377번, ‘of’는 36146번 사용되어 각각 코퍼스 전체 어절의 5.8%와 3%를 차지하고 있다). 코퍼스로부터 도출된 단어목록은 어휘 교육이나 평가 개발 등에 유용하게 이용될 수 있다 (Reppen & Simpson, 2002). 예를 들면 단어목록을 참조하여 특정 빈도 범위 안-예를 들어 100만 단어 당 열 번에서 열 다섯 번까지-에서 발생하는 단어들을 선택하여 교수요목(syllabus)이나 평가 항목에 포함시킬 수 있다. 뿐만 아니라 서로 다른 코퍼스의 단어목록을 비교하면 각 코퍼스가 대표하는 언어사용역간의 기본적인 어휘상의 차이점을 알아낼 수도 있다.

단어목록 외에 콩코던스 프로그램을 이용하여 어떤 단어들이 코퍼스 내에서 함께 발생하는가, 즉 어휘 동시발생(co-occurrence) 양식에 대한 정보도 얻을 수 있다. 이러한 양식은 콩코던스 선(concordance line)을 통해 잘 드러나는데, 콩코던스 프로그램의 검색 창을 띄우고 목표 단어나 어구를 입력하면 프로그램이 코퍼스 내의 텍스트를 탐색하여 모든 목표 단어/어구의 예를 목록 형태로 제공해 준다. 이러한 화면표시를 문맥 내 핵심어 (Key Word In Context: KWIC)라고 부르는데 이는 목표 단어의 다양한 사용이나 의미를 조사하는 데 이용될 수 있다. [그림 2]는 목표 단어 'speak'를 브라운 코퍼스에서 검색하여 얻어진 KWIC 화면이다.

[그림 2: 콩코던스 KWIC 화면]



화면의 윗부분에서는 화면 아래 부분의 여러 예들 중 연구자가 선택한 예(이 경우는 첫 번째 예)의 문맥을 제공해 주는데 이 창의 크기와 문맥

의 양은 조절될 수 있다. 이렇게 검색된 목표 단어의 예들 역시 다양한 방식으로 정렬할 수 있다. 예를 들면 목표 단어의 오른쪽 혹은 왼쪽에 나타나는 단어의 알파벳순으로 정렬할 수 있으며 ([그림 2]에서는 오른쪽 첫 번째 단어의 알파벳순으로 정렬되어 있다), 또는 목표 단어의 주변에서 발생하는 특정 단어들을 찾아서 정렬할 수도 있다. 이처럼 어떤 단어들이 목표 단어 주변에서 가장 자주 나타나는지 알아낼 수 있고 이는 연어(collocation) 분석에 큰 도움을 줄 수 있다. 연어 분석은 개별 어휘 항목의 문법적, 의미적 사용 양식에 대하여 중요한 정보를 제공해 준다 (Sinclair, 1991). 같은 뜻을 지닌 것처럼 보이는 단어나 문법 구조들일지라도 그 쓰임을 자세히 살펴보면 실제로는 다른 단어들과의 연합 양식이나 특정 구조에 대한 선호 경향에 있어서 차이를 보이는 경우가 많다.⁵⁾ 코퍼스와 콩코던스 프로그램을 이용하면 이러한 차이를 보다 쉽게 인지할 수 있다.

지금까지 코퍼스의 유형과 코퍼스 설계시 고려해야 할 문제들, 그리고 콩코던스 프로그램의 사용법 및 이를 통해 얻을 수 있는 정보에 대해서 간단히 살펴보았다. 다음 장에서는 코퍼스가 어떻게 영어교육에 활용될 수 있는지 구체적으로 논의해 보도록 하겠다.

III. 영어교육 분야의 코퍼스 활용

코퍼스가 영어 교육 분야에서 활용될 수 있는 방식은 크게 세 가지로 구분해 볼 수 있다. 첫째, 코퍼스 분석은 목표어(target language)인 영어가 실제로 사용되는 양식들을 밝힘으로써 영어교육의 내용 향상에 기여할 수 있다. 둘째, 코퍼스 분석 활동 자체가 영어교육의 한 방법론(methodology)으로 채택되어 사용될 수 있다. 마지막으로 학습자 코퍼스(learner corpus)를 이용한 학습자 언어 분석을 통해서 특정 집단 영어 학습자가 가진 문제점을 진단하고 해결해 나가도록 도와줄 수 있다.

5) 예를 들어 유의어 'begin'과 'start'는 자동사와 타동사 둘 다로 쓰일 수 있는 동일한 문법적 가능성을 지니고 있다. 그러나 코퍼스 분석을 통해 살펴보면 'start'는 (특히 학문적인 글에서) 자동사 양식을 훨씬 더 선호한다 (Biber, Conrad & Reppen, 1998).

1. 교육 내용

먼저 코퍼스는 무엇을 가르칠 것인가, 즉 교육의 내용면에서 기여할 수 있다. 이는 다시 크게 둘로 양분하여 생각해 볼 수 있는데, 첫째, 코퍼스 분석을 통해 목표어의 다양한 측면들을 보다 정확하고 세밀하게 조사하고 기술(description)함으로써 학습자들이 배우는 내용의 질적 향상을 가져올 수 있으며, 둘째, 교과서나 교수요목에 포함시킬 내용을 직관이 아닌 구체적인 코퍼스 증거에 근거하여 결정할 수 있다.

실제 사용되는 언어에 대한 보다 정확한 기술이라는 측면에서 코퍼스의 활용이 가장 두드러진 분야는 어휘와 문법이다. 특히 어휘 기술과 관련하여 사전 편찬 분야에서는 코퍼스 활용이 거의 절대적인 위치를 차지하고 있어서 현재 영미권에서 출판되는 영어 사전 중에는 코퍼스에 기반을 두지 않은 경우가 드물다.⁶⁾ 코퍼스를 이용하면 수백만, 혹은 수천만 단어로 이루어진 텍스트 내에서 특정 단어나 어구가 쓰인 모든 예들을 수초 내에 불러들일 수 있다. 이처럼 많은 양의 자연적인 예들을 조사할 수 있으므로 정의(definition)가 이전보다 더 완전하고 정확해 졌으며, 또 사전이 훨씬 더 빨리 생산, 개정될 수 있다. 새로운 언어자료가 정기적으로 계속 첨가되는 모니터 코퍼스는 신조어나 단어의 의미 변화를 보다 신속하게 사전에 반영하도록 할 수 있기 때문에 사전 편찬에 특히 큰 역할을 한다. 그러나 다른 일반 코퍼스들 역시 단어들에 대한 양적인 정보를 제공해 준다는 점에서 어휘 연구에 중요한 역할을 할 수 있다. 코퍼스는 단어의 다양한 의의 및 특정 유형의 쓰임을 확인하고 각각의 상대적 빈도를 알아내는데 사용된다.⁷⁾ 또한 서로 다른 단어들이 얼마나 자주 사용되는지, 어떤 단어들이 특히 자주 사용되고 어떤 것들이 잘 사용되지 않는지를 밝혀준다. 의미나 빈도뿐만 아니라 단어의 사용 양식상의 특징도 역시 코퍼스를 통해 알아낼 수 있는데, 이러한 특징은 그 단어의 구체적인 의미와도 밀

6) 사전 편찬을 위해 콩코던스 프로그램을 이용하여 문맥 내에서의 단어의 쓰임 및 언어 사용을 연구한 것은 코빌드 프로젝트(Cobuild Project)가 최초이며 그 첫 결과물이 코빌드 영어 사전(Collins Cobuild English Dictionary) (1987)이다.

7) 예를 들어 'good'이라는 단어는 50만 어절의 영국영어 구어 코퍼스인 London-Lund Corpus에서 800회 이상 발견되며, 20개가 넘는 의의를 가지고 있다 (Kennedy, 1998).

접한 관련이 있는 경우가 많다. 흔히 동의어로 간주되는 단어들의 실제 쓰임을 관찰해 보면 예상치 못했던 차이가 드러나고, 이러한 차이에 근거하여 각 단어의 의미를 보다 정확히 파악할 수 있다. 예를 들어 ‘결과적 계사(resulting copulas),’ 즉 어떤 힘이나 행동의 결과로 생기는 상태 변화를 나타내는 ‘become, turn, go, come’은 흔히 사전에서 ‘-이 되다’의 의미를 지닌 동의어로 제시하고 있다. 대부분의 사전은 이 네 단어간의 의미 차이에 대해 아무런 단서도 제공하지 않는 반면, 코퍼스 연구는 이 단어들이 전형적인 사용 맥락에 있어서 매우 큰 차이가 있음을 보여 준다 (Reppen & Simpson, 2002). 즉 ‘turn’은 색깔이나 외형의 변화를 가리킬 때 사용되며 (‘the sky turned red’), ‘go’는 부정적인 상태로의 변화 (‘go crazy’, ‘go bad’, ‘go wrong’), 그리고 ‘come’은 좀 더 활동적인 상태로의 변화를 기술한다는 것이다 (‘come awake’, ‘come alive’) (Biber et al., 1999). 이처럼 걸보기에는 뜻이 비슷한 단어들의 경우 코퍼스를 통해서 자세히 분석을 해 보면 여러 가지 차원에서 상이한 점이 발견되는 경우가 많다.⁸⁾ 이러한 유의어간의 차이는 코퍼스에 기초한 연구 없이는 밝히기가 힘들다. 그리고 이러한 연구들의 결과는 유의어들의 실제 쓰임에 있어서의 미묘한 차이를 알기 힘든 외국어 학습자들에게 유용하게 사용될 수 있다. 한 단어의 사용 양식을 제대로 이해하는 것은 외국어 학습자들에게 매우 중요한 일인데, 원어민들의 직관은 그러한 양식을 예측하는데 크게 도움을 주지 못하는 경우가 많다 (Reppen & Simpson, 2002). 전통적인 사전 역시 이런 경우에 별 도움을 주지 못하지만 코퍼스는 이러한 정보를 보다 직접적으로 눈에 드러나게 보여줄 수 있기 때문에 외국어 학습에 큰

8) 한 가지 예를 더 들자면, Biber, Conrad와 Reppen (1998)은 유의어인 ‘big, large, great’을 분석하였는데 그 결과 각 형용사는 선호하는 언어형(collocates)이나 의미뿐 아니라 언어사용역별 분포 역시 달랐다. ‘Big’은 물리적 크기를 가리킬 때 흔히 사용되는 반면, ‘large’는 양을 가리킬 때 사용된다. ‘Great’ 역시 양을 가리키지만 특히 ‘great deal’이라는 언어형태로 사용되며, 의미 영역이 다른 두 형용사에 비해 더 넓어서 강도(intensity), 크기, 그리고 가족관계까지도 포함한다. 이러한 의미 선호도 차이는 소설과 학구적 산문이라는 두 언어사용역에서의 이들 형용사의 빈도 분포 차이를 설명해 준다. 소설에서는 ‘big’이 훨씬 더 많이 쓰이는 반면 양을 다루는 학문적 글에서는 ‘large’가 많이 발견되었다. 두 언어사용역 모두 ‘great’를 사용하지만 소설에서는 그 의미가 좀 더 다양하게 나타남으로써 학문적 글과 비교했을 때 좀 더 다양한 기술(skill)이나 관심을 반영하고 있다.

도움을 줄 수 있다. 특히 연어와 같은 단어 결합 양식에 대한 정보는 영어 학습자들이 단어 사용의 문맥을 정교화하고 좀 더 원어민과 같은 언어 사용형태로 발전해 나가도록 도울 수 있다.

어휘뿐 아니라 문법 분야에서의 코퍼스 활용 또한 활발하게 이루어지고 있다. 문법 연구는 어휘 연구와 더불어 코퍼스를 이용해서 가장 많이 이루어진 연구 유형이다. 코퍼스를 사용한 문법 연구는 연구자의 직관이나 내성이 아닌 실제 언어 사용을 토대로 하며 규범문법(prescriptive grammar)보다는 기술문법(descriptive grammar)을 지향한다. 전자 코퍼스가 존재하기 이전에도 비전자형태의 코퍼스는 기술문법의 구축을 위한 주 자료였고, 20세기 전반의 주요 기술문법들은 문법 특징이나 구조를 예시하기 위해 흔히 신문이나 소설을 이용했다.⁹⁾ 가장 중요한 비전자 코퍼스로는 1953년에 시작되어 1987년에 이르러서야 완성된 Survey of English Usage (SEU) Corpus를 꼽을 수 있는데 이 코퍼스는 이후 만들어진 많은 코퍼스 설계의 기초를 확립했다 (Kennedy, 1998). 지금도 가장 종합적이고 체계적인 문법서 중 하나로 꼽히는 *A Comprehensible Grammar of the English Language* (Quirk et al., 1985)가 바로 이 코퍼스를 이용해서 만들어짐으로써 이후 코퍼스에 기초한 문법서들의 효시가 되었다. 1990년대의 대표적인 코퍼스 문법서로는 *Longman Grammar of Spoken and Written English* (Biber et al., 1999)를 꼽을 수 있는데, 이 문법서는 여러 가지 언어 사용역에서의 문법 사용을 분리하여 양적 정보와 함께 자세히 설명하고 있다.

Kennedy (1998)는 코퍼스 문법 연구들을 크게 형태소 및 단어 단위의 연구와 문장 단위의 연구 둘로 나누어 소개하고 있다. 이 중 형태소나 단어 단위의 연구라면 시제(tense)와 상(aspect), 서법(mood), 태(voice) 등과 관련된 연구를 들 수 있다. 예를 들어 영어에서 가장 자주 사용되는 동사들과 그 동사들의 시제 및 상 사용을 코퍼스를 통해 살펴본 연구들이 있는데 (Francis & Kučera, 1982; George, 1963; Joos, 1964; Ota, 1963), 이 연구들은 서로 다른 코퍼스를 사용했지만 비슷한 결과들을 보여 준다. 그 중 한 가지 결과는 구어체 영어에서는 단순 현재 시제가, 문어체 영어(특히 이야기)에서는 단순과거가 훨씬 더 자주 사용되며, 완료상이나 진행

9) 이 당시 코퍼스를 문법 연구에 이용한 대표적인 학자로는 Jespersen이나 Fries를 꼽을 수 있다 (Kennedy, 1998).

상은 상대적으로 잘 쓰이지 않는다는 것이다. 태와 관련된 연구 중 Francis와 Kučera (1982)는 브라운 코퍼스 내 여러 장르에서의 능동태와 수동태 사용 비율을 분석한 결과, 상상적 산문(4.2%)보다 정보제공 산문(14.6%)에서 수동태가 더 자주 사용된다고 보고하고 있다. Svartvik (1966)은 SEU 코퍼스에서의 수동태 사용을 조사하였는데, 3,600개가 넘는 수동태 중에서 80%가 행위자 없이 쓰이고 있음을 발견했다. 이는 수동태 문장을 그에 대응하는 능동태 문장으로부터 도출하면서 능동태 문장의 주어를 수동태에서 'by + 명사(구)'의 행위자 구문으로 전환하도록 지도하는 기존 수동태 교수 방법에 문제가 있음을 시사한다. 한편 문장 단위의 문법연구들은 조건문, 명사수식, 인과관계 표현법, 부정문, 분열문(clefting) 등의 주제를 다루고 있다. Wang (1991)은 Brown 코퍼스와 LOB 코퍼스를 이용하여 조건문을 연구했는데, 이에 따르면 if-조건절의 약 절반이 사실적(factual) 의미로 (e.g., If you boil water, it vaporizes; If I wash the dishes, my husband dries them), 25%가 미래 예측의 의미로 (e.g., If it rains, I'll stay home) 사용되었다. 그리고 나머지가 상상적(imaginative) 조건문이었는데, 14% 정도는 있을 법하지 않은(improbable) 조건을 (e.g., If I had the time, I would go to Japan) 표현하고 그 나머지는 사실과 반대되는(counterfactual) 조건문 (e.g., If my mom were alive today, she would be so happy)이었다. 다시 말해서 실제에 입각한 조건문이 상상적 조건문보다 약 세배나 더 자주 쓰인다는 것인데, 이러한 코퍼스 연구 결과를 참고하여 전자보다 후자를 훨씬 더 강조하는 현 문법 교육의 방향을 재고해 보아야 한다.

이처럼 코퍼스를 통해 이전보다 더 정확하게 언어 사용을 파악하게 되면 학습자들이 배우게 될 내용을 향상시킨다는 점에서 의의가 있다. 보다는 사전이나 문법서가 개발되면서 학습자뿐만 아니라 교사들도 이로부터 크게 도움을 받을 수 있다. 또한 어휘나 문법항목의 빈도에 관한 정보는 교수요목의 선정 및 순서 결정에 있어서 중요한 역할을 할 수 있다. 코퍼스에 기초한 연구들은 어떤 언어 자질과 구조들이 중요한지, 그리고 다양한 자질과 구조들이 어떻게 사용되는지에 대한 증거를 제공함으로써 영어 교사들에게 도움을 줄 수 있다. 따라서 교사들은 코퍼스 정보에 근거하여 교수(instruction)를 구체화할 수 있으며 가르칠 언어 항목에 대한

정보를 얻기 위해서 코퍼스 연구들을 참조할 수 있다. 예를 들어 만약 교수의 초점이 대화체 영어라면 교사들은 구어에 대한 코퍼스 연구들을 읽고 어떤 자질들과 문법 구조들이 대화체 영어의 특징인지 결정한 후 구체적인 교수내용을 준비할 수 있다. 만약 교수의 초점이 특정 문법 구조라면, 코퍼스 연구들은 그 구조와 연관된 어휘 및 화용적 동시발생 양식을 밝혀줌으로써 그 문법 구조의 사용 범위에 대하여 구체적인 정보를 제공할 수 있다. 특수 목적으로 영어를 학습하고 있는 학생들의 경우(English for Specific Purposes)에는 이들이 주로 접하게 될 언어 상황을 대표할 수 있는 전문 코퍼스를 이용하여 이들에게 특히 필요한 어휘와 문법 구조들을 선정하여 가르친다면 훨씬 더 효과적이 될 수 있을 것이다. 이러한 예로서 미국에서 대학과정을 이수하게 될 외국 학생들을 위해서 Biber, Conrad, Reppen, Byrd와 Helt (2002)는 미국 대학환경에서의 구어 및 문어 텍스트로 이루어진 TOEFL 2000 Spoken and Written Academic Language Corpus를 분석함으로써, 이러한 환경에서의 구어와 문어 언어 사용역이 근본적으로 차이가 있다는 것을 밝혀냈다. 학구적인 담화(academic discourse)에 대한 이전 연구들은 대부분 문어 담화를 대상으로 하였기 때문에 일상적인 구어 담화와의 차이에만 집중을 했던 반면, Biber 외 4인(2002)의 연구는 대학 교실에서의 수업이 여러 측면에서 일상 대화와 유사하다는 점 등을 새로이 보여주면서, 이런 특수 목적을 가진 학생들을 위한 교재 개발 시 이러한 사실들을 반영해야 한다고 지적하고 있다.

한편 현재 사용되고 있는 영어 교과서나 교재의 내용을 코퍼스에 나타난 실제 언어 사용과 비교함으로써 전자의 진정성(authenticity) 여부를 점검해 볼 수 있다. 실제로 교과서와 원어민 코퍼스를 비교한 연구들에 따르면 영어 교과서는 코퍼스와 매우 상이한 측면을 보여주곤 한다. 특히 전통적인 교재나 교수요목은 원어민 코퍼스에서 매우 빈번하게 나타나는 언어 자질들을 무시하고 상대적으로 빈도수가 매우 낮은 자질들을 더 강조하는 경향이 있는 것으로 나타났다.¹⁰⁾ Holmes (1988)는 영어교과서가

10) Widdowson (1991)이 지적하듯이 교육 내용을 선정하는 데 있어서 빈도수보다 생산성(productivity)이나 학습가능성(learnability)이 더 중요할 수도 있겠지만, 빈도수 역시 분명히 중요한 요소이며, 특히 읽기와 듣기 같은 수용적 기술(receptive skill)과 관련해서는 더욱 그러하다 (Aston, 2000).

인식 서법(epistemic modality) 표현에 있어서 서법 조동사(modal verbs)에만 지나치게 치중함으로써 원어인 코퍼스에서 나타나는 다양한 언어 책략들을 무시하고 있다고 비판하고 있다. Carter (1998) 역시 교과서의 대화가 실제 코퍼스 대화에서 매우 빈번하게 발견되는 담화표지(discourse marker), 생략(ellipsis), 모호하고 애매한 표현(vague language, hedges) 등을 포함하지 않고 있다고 보고한다. 이 외에 Hughes와 McCarthy (1998)는 대과거(past perfect) 표현과 관련하여, Mindt (1996)는 조건문과 조동사, Mindt (1997)는 ‘any,’ ‘will/would,’ 불규칙 동사들과 관련하여 마찬가지로 영어 교과서의 문제점을 제기하였다. 코퍼스는 진정성 있는 텍스트를 제공해준다는 점에서 의사소통적 언어교육(communivative language teaching)에도 기여할 수 있다. 의사소통적 언어교육의 특징 중 하나는 언어 교육 목적으로 특별히 제작되거나 수정되지 않은 텍스트, 즉 진정성 있는 텍스트를 선호한다는 것이다.

이는 문맥 속에서 의사소통을 위해 실제로 사용되는 언어를 보여줌으로써 실제 언어 양식을 왜곡하는 위험을 피할 수 있으며, 이러한 이유로 인해 학습자들을 더 동기화할 수 있다고 여겨진다 (Aston, 2000). 실제 의사소통을 위해 사용된 언어의 수집물인 코퍼스가말로 진정성 있는 자료의 풍요로운 원천이 될 수 있다.¹¹⁾

2. 교수 방법

위에서 코퍼스가 교육 내용에 어떻게 기여할 수 있는지를 살펴보았다. 코퍼스가 영어교육에서 이용될 수 있는 두 번째 방법은 학습자들로 하여금 직접 코퍼스와 상호작용하도록 하는 것이다. 이러한 방법론적 차원에서 코퍼스 활용이 본격화된 데에는 자료가 이끄는 학습(Data-Driven Learning: DDL)을 주창한 Tim Johns의 역할이 매우 크다. 그는 “연구는 연구자들에게만 맡기기에는 너무 중요하다 (Research is too important to be left to the researchers) (1991, p. 2)”고 주장하면서, 학습자들이 “언어

11) Exploring Spoken Language (Carter & McCarthy, 1997)는 실제 코퍼스 자료를 근거로 만든 구어체 영어 교재로서 폭넓게 사용되고 있다.

탐정 (language detectives) (1997, p. 101)”이 되어 목표어에 관한 다양한 사실들을 실제 예들로부터 스스로 발견하도록 하는 것이 중요하다고 역설한다. 따라서 DDL은 콩코던스 선의 형태로 주어지는 코퍼스 자료를 학습자들 스스로 연구함으로써 언어에 대한 질문에 스스로 답해 나가도록 하는 상황을 만드는 데 그 핵심이 있다.

DDL은 교실 상황에서 주로 두 가지 방식 중 하나로 이루어지게 된다. 만일 교실의 컴퓨터 설비가 충분하다면 학습자들이 직접 콩코던스 프로그램을 사용하여 보다 능동적으로 코퍼스 탐험에 참여할 수 있다. 학생은 자신의 학습 과정 중에 생겨난 질문, 예를 들어 “이 문맥에서는 A와 B중 어느 것이 더 적합한가?” “A와 B의 차이는 무엇인가?” 등의 질문들을 스스로 만들어내고 코퍼스를 통해 이에 대한 해답을 찾게 된다.

이렇게 학생들이 직접 컴퓨터를 이용하게 하면 훨씬 더 학습자 중심의 접근이 가능하며, 또 코퍼스 사용 책략을 기쁨으로써 교실 밖으로까지 학습이 이어질 수 있다. 그러나 만일 충분한 컴퓨터 시설이 갖추어지지 못했다면 교사가 코퍼스 탐색의 결과물을 교실로 가져갈 수도 있다. 이 경우 교사는 특정 집단 학생들에게 문제가 되거나 필요하다고 여겨지는 항목들에 대해 코퍼스로부터 자료를 도출한 뒤 그 자료를 선택, 편집하여 학생들로 하여금 이를 탐구하도록 한다. 이 경우 편집된 코퍼스 자료를 종이에 인쇄하여 전체 학급과 함께 이용할 수 있다 (Willis, 1998 참조). 첫 번째 경우에는 학생들의 동기가 최대화된다는 장점이 있는 반면, 교사의 통제력이 떨어진다는 단점이 있다. 사실 좀 더 근본적인 문제는 학생들이 스스로 연구할 수 있도록 컴퓨터 시설을 충분히 갖춘 교육환경이 그리 많지 않다는 것이다. 반면 두 번째 방식을 이용하게 되면 교사는 학생들이 접하게 될 정보를 미리 선택할 수 있으므로 좀 더 통제력을 지니게 된다. 또한 컴퓨터에 익숙하지 않은 학생들의 정서적 여과기(affective filter)를 낮춰줄 수 있고 교사가 미리 준비를 해 오기 때문에 수업 시간을 단축할 수 있다. 그리고 보다 많은 학생들에게 자료를 제공해 줄 수 있는 면도 있다. 단점이라면 학생들이 찾으려는 또는 찾아낸 정보를 기억하려는 동기가 상대적으로 낮을 수 있다는 것이다 (Hunston, 2002; Reppen & Simpson, 2002).

DDL을 하기 위한 교육 자료를 만들려면 먼저 기본적인 도구와 자원

(컴퓨터, 코퍼스, 콩코던스 프로그램)을 사용할 수 있어야 한다.¹²⁾ 그 외에는 특별한 기술 없이 단지 단어목록과 콩코던스 선만을 가지고도 여러 가지 학습/교수 활동들을 제작할 수 있다 (Donley & Reppen, 2001; Tribble & Jones, 1997). 목표 단어의 콩코던스는 그 단어의 의미와 용법에 대해 보다 더 잘 이해하도록 하는데 이용될 수 있다. 처음에는 콩코던스를 통해 단어의 의미를 발견하도록 할 수 있는데, 이런 경우 콩코던스 KWIC 화면에서 목표 단어만 모두 생략하여 빈 칸으로 두거나 또는 목표 단어를 인위적으로 만든 단어로 모두 대치한 후 학습자들이 실제로 들어가 할 표현을 맞추는 활동도 가능하다. 같은 단어를 동시에 여러 문맥 안에서 제시하게 되면 그 의미를 추론하는 과정이 호기심을 자아내는 문제 해결 과업의 성격을 띠게 된다 (Tribble & Jones, 1997). 따라서 DDL은 특정 언어 항목에 대한 교수와 학습에 도움이 될 뿐만 아니라, 이러한 일련의 과정을 통해 문맥을 이용하여 단어의 의미를 추론해내는 기술을 신장시키는 데에도 도움이 될 수 있다. 텍스트에서 모르는 단어를 마주치게 될 때 주변 문맥으로부터 그 의미를 유추해 내는 것은 읽기 과정뿐만 아니라 어휘 학습에 있어서도 매우 중요한 능력이며, 이러한 능력의 개발은 전반적인 언어 능력의 향상으로까지 이어질 수 있다. 한 단어의 문법적 특징 (예를 들어 'interested'나 'depend' 뒤에 어떤 전치사가 사용되는지)이나 유의어간의 차이점 등도 코퍼스 활동을 통해 학습시키면 효과적이다.

코퍼스 자료를 이용한 교실 활동의 예를 하나 소개해 보겠다. 만약 'think about'과 'think of'의 차이에 대하여 학생이 질문을 한다면 교사는 [표 1], [표 2]와 같이 코퍼스에서 추출한 콩코던스 선을 제공함으로써 그 차이를 관찰하도록 할 수 있다 (Reppen & Simpson, 2002, pp. 110-111).

12) 학습자들이 직접 코퍼스를 이용하는 방식으로 DDL을 하려면 학습자들에게도 이러한 도구들을 생산적으로 사용할 수 있도록 훈련시켜야 할 것이다.

[표1: THINK OF]

stank. Then as he was trying to	think	of	something to say to her (all
yes, wedding presents. We must	think	of	something. You probably don't
racking my brains for three hours to	think	of	something. I simply cannot last
a second catastrophe. I tried to	think	of	something to say myself, but my
offered frills. Nicandra tried to	think	of	something pleasing to say:
only you were here, then we could	think	of	something to do. 'Christopher
groaning quietly, perhaps trying to	think	of	something that summed up what
let said nothing. He had tried to	think	of	something to say, but the only
lunch?' 'Ah me, the young! You	think	of	nothing but your stomachs.
sympathy and collusion. But I can	think	of	nothing to say. Perdie says,
she tried to speak, but she could	think	of	nothing, and her mother, shifting
anything so familiar, and he could	think	of	nothing on earth to say. It
man in the world.' 'As he could	think	of	nothing else, Martin repeated
But try as she might, she could	think	of	nothing to say like that, fierce
listening. 'Can we ourselves	think	of	nothing that needs to be done?
'what an idiot I was not to	think	of	it before! You all right Elfie?
... no, wait a minute, come to	think	of	it you 're finding. hmm.
or him, on other occasions, come to	think	of	it. We've been aware of each
happened to those kids. And come to	think	of	it, Hamelin's rats and children
like that five years ago, come to	think	of	it, or even ten. It 's the
wash his feet, he had seen, come to	think	of	it, the moon not too remote from
probably cheaper than Selina, come to	think	of	it, what with the hotel mark
could have. I didn't happen to	think	of	it then. 'And when did you
her pregnant. Better not even to	think	of	it. Just go on hating him,
and done with. Don't let us ever	think	of	it again. My family always
'How nice. What did you	think	of	it?' Patrice held her breath,

[표2: THINK ABOUT]

You wouldn't just	think	about	it it 's just gone isn't it
Well that 's a good way, if you	think	about	it he's got, he's got four
more, I mean they can wear, if you	think	about	it they were suits in the
When you	think	about	it, yeah he was So what '
it seems easier that way when you	think	about	it dunnit? Mm it's a lot be
does that come from? Oh when you	think	about	it Pledge, why do they call
wasn't the money really when you	think	about	it because at end of day,
more. I mean they can wear if you	think	about	it they were suits in the
week! And why, they don't need to	think	about	it, they can talk you out of
penetrating as lasers. 'We might	think	about	that, 'I say at last.
I'll have to start and	think	about	that train, Dwight.
see it. That's they wa I like to	think	about	that sort of place. It's
another way, but I don't want to	think	about	that for a while. 'Timothy
get eight to twenty-five. Now	think	about	that. The district attorney

이런 콩코던스 선을 비교 조사하게 되면 두 표현의 사용상의 차이를 보다 확실히 인지할 수 있게 된다. Reppen과 Simpson이 설명하듯이, 'think of'와 'think about'의 가장 큰 차이는 전자는 부정 지시대상(indefinite references) (e.g., something, nothing, 특정 지시대상이 없는 it)와 함께 사용되는 경우가 대부분인 반면 후자는 보통 특정한 지시대상과 함께 쓰인다 (e.g., 선행 텍스트에 특정한 지시 대상이 있는 it이나 that)는 것이다.¹³⁾

DDL의 장점은 학습자들이 자신들 스스로 발견해낸 것은 보다 더 잘 기억하도록 동기화되기 때문에 학습이 촉진된다는 것이다 (Hunston, 2002). 뿐만 아니라 코퍼스 자료는 이전에는 미처 눈치 채지 못하고 지나갔던 언어 사용양식을 드러내줄 수 있기 때문에, 교사가 간과하거나 교과서가 다루지 못한 부분까지도 학습자 스스로 주목할 수 있다는 장점이 있다. 학습자들에게 코퍼스를 직접 탐색해 나가도록 지도함으로써 이후에도 학습자의 자율적(autonomous) 언어 학습을 장려한다는 면에서도 의의가 있다 (Aston, 2000).¹⁴⁾

3. 학습자 언어 연구

지금까지 살펴본 코퍼스 연구나 코퍼스를 이용한 교수/학습 활동은 영어를 모국어로 하는 사람들의 말이나 글로 이루어진, 즉 원어민 코퍼스를 대상으로 한 것들이다. 원어민 코퍼스가 영어교육의 향상에 도움이 될 수 있다는 것에 대해서는 의심의 여지가 없을 것이다. 이미 여러 번 지적한

13) 이것은 위에 제시된 제한된 양의 문맥만으로는 파악하기 힘들 수도 있겠지만 실제로 콩코던스 프로그램을 사용하는 경우에는 목표 단어 주변 문맥의 범위를 확장시킴으로써 이러한 지시양식을 확인할 수 있다.

14) 교실에서의 콩코던스 과업 사용은 아직 어느 정도 논란의 대상이 되고 있다. 자료로부터 학습을 이끌어내는 귀납적인 접근법을 선호하는 쪽에서는 이를 강력히 옹호하고 있는 반면 (Johns, 1994), 학생들이 방대한 양의 언어 예들을 분석하도록 적절하고 효과적으로 지도하는 것이 어렵다는(Cook, 1998) 등, 코퍼스를 외국어 교실에서 이용하는 것에 대해 문제점을 제기하는 연구자들도 있다 (Owen, 1996; Widdowson, 2000). 분명 외국어 학습자들을 코퍼스와 콩코던스 도구에 노출시키는 것이 얼마나 효과적인지에 대한 연구와 실험이 필요하다 (Aston, 2000; Reppen & Simpson, 2002).

대로 코퍼스는 원어민의 전형적인 언어 사용을 원어민 직관보다도 더 잘 드러내 주기 때문에 코퍼스 연구들은 이러한 사용을 제대로 반영하도록 교육과정 및 교재 개발 면에 기여할 수 있다. 코퍼스를 활용한 교수 방법 역시 학습자들이 코퍼스를 통해 원어민들의 언어 사용 양식을 스스로 발견해 내도록 하는 데 주안점이 있었다. 그러나 외국어 교육에 있어서는 원어민들의 전형적인 언어 사용 양식을 잘 파악하여 가르치고 배우는 것도 중요하지만, 특정 집단 학습자들이 목표어를 배우면서 전형적으로 겪는 어려움을 이해하여 이를 교육에 반영하는 것 역시 매우 중요하다. 학습자의 모국어와 목표어를 대조하는 것만으로는 이러한 학습상의 어려움을 예측하는 데 충분치 않으므로 이를 보완하기 위해서는 학습자들이 생산한 언어를 원어민의 언어와 비교 분석하는 것이 필요하다. 다시 말해서 원어민 코퍼스는 특정 목표어 자질의 상대적인 어려움이나 학습 가능성에 대한 정보는 제공해 줄 수 없기 때문에, 이러한 이유들로 인해 학습자 코퍼스(learner corpus), 즉 외국어를 배우는 사람들의 말이나 글로 이루어진 코퍼스의 구축에 점점 더 많은 관심이 생겨나고 있다. 학습자 코퍼스는 체계적이고 컴퓨터화된, 학습자들이 생산한 텍스트의 수집물로 정의될 수 있는데, 체계적이라는 것은 코퍼스에 포함된 텍스트들이 여러 기준(학습자 수준, 모국어 등)에 근거하여 선택되었으며, 또 이러한 선택이 대표성을 가지고 있고 균형 또한 이루어져 있다는 것을 의미한다(Nessedhauf, 2004).

학습자 코퍼스는 원어민 코퍼스와 비교하여 상대적으로 매우 짧은 역사를 가지고 있다. 대부분의 학습자 코퍼스는 1990년대에 들어서 구축되기 시작했다. 미국이나 영국의 주요 출판사들에서는 학습자를 위한 사전과 교재 개발에 이용하기 위해 학습자 코퍼스를 구축해 왔으며 (e.g., Longman Learners' Corpus, Cambridge University Press Learner Corpus),¹⁵⁾ 이 밖에 유럽 및 아시아 일부 지역(중국, 일본 등)에서도 연구 목적으로 이미 여러 영어 학습자 코퍼스가 구축되었거나 구축 중에 있

15) 학습자 코퍼스 분석의 결과를 토대로 만들어진 최초의 사전은 Longman Language Activator (1993)이며, 이 외에 Longman Dictionary of Contemporary English (1995), Longman Essential Activator (1997), Cambridge International Dictionary of English (1995) 등도 학습자 코퍼스 분석에 바탕을 둔 사전들이다 (Nessedhauf, 2004).

다.16) 가장 널리 알려진 학습자 코퍼스는 벨기에의 Sylviane Granger가 중심이 되어 만든 International Corpus of Learner English (ICLE)이다. 이 코퍼스는 초기에 참여한 유럽 11개 각 국가에서 20만 어절의 코퍼스를 구축한 뒤 이를 합친 형태로서 현재 전체 2백 5십만 어절로 이루어져 있는데, 중국과 일본 등 8개 국가 학습자 코퍼스가 더 추가되고 있는 중이다. ICLE 코퍼스는 영어를 외국어로 배우는 학습자들이 쓴 약 500단어의 논설문(argumentative essays)으로 이루어져 있다 (Granger, 1998). 이 학습자들은 대학에서 영어나 영문학을 공부하고 있는 높은 능숙도(proficiency) 수준의 학습자들임에도 불구하고, 이 학습자 코퍼스를 이에 상응하는 원어인 코퍼스와 비교한 연구들은 다양한 차이점을 밝혀냈다.17) 학습자 코퍼스를 통해 이루어지는 연구들은 대부분 대조분석(contrastive analysis)의 형태를 취하는데, 이미 언급한 대로 원어인 코퍼스와 대조를 할 수도 있고, 또는 다른 모국어 배경을 가진 학습자들의 중간언어(interlanguage)를 비교할 수도 있다.

최근 들어 학습자 코퍼스가 외국어 교육에 기여할 수 있다는 인식이 확산되고 있으며 (e.g, Aston, 2000; Granger, Hung & Petch-Tyson, 2002), 학습자 코퍼스를 분석한 연구들의 숫자도 빠르게 증가하고 있다.18) 문법

16) 현재까지 구축된 최대 규모의 학습자 코퍼스는 중국 대학생들의 자료로 이루어진 2천 5백만 어절의 Hong Kong University of Science and Technology Learner Corpus이다. 이 외에도 홍콩 학생들을 대상으로 한 3백만 어절의 TeleNex Student Corpus, 백이십만 어절의 Chinese Learner English Corpus, 스웨덴 대학교 학생들의 작문으로 이루어진 백만 어절의 Uppsala Student English Project 등이 있다.

17) 학습자 코퍼스 연구의 핵심은 비교, 즉 학습자 코퍼스와 원어인 코퍼스 간의 비교, 혹은 서로 다른 학습자 집단에 의해 생산된 코퍼스간의 비교에 있다. 따라서 코퍼스 간의 비교 가능성이 중요하다. ICLE 코퍼스에 상응하는 원어인 코퍼스는 LOCNESS (Louvain Corpus of Native English Essays)이며, 30만 어절의 이 코퍼스는 미국과 영국의 대학생들이 쓴 비슷한 형식의 논설문으로 이루어져 있다.

18) 제2언어습득 연구에서 학습자 코퍼스가 사용되기 시작한 것은 비교적 최근의 일이다. 다른 언어관련 분야에 비해 이 분야의 코퍼스 활용이 상대적으로 늦어진 이유에 대해 Leech (1998)는 언어교육 분야의 재정적인 어려움과 함께 이론적인 문제를 지적한다. 의사소통적 교육과 학습이 생산물(product)보다 과정(process)을 중시함에 따라 학습자들이 생산해 낸 결과물에 대한 자세한 연구가 별로 이루어지지 않았고, 또 제2언어습득을 정신적인 과정, 즉 언어능력의 습득으로 파악함으로써 관찰 가능한 학습자의 언어수행에는 큰 관심을 두지 않았다는 것이다. Leech (1998)는 이제 비로소 제2언어습득 연

(e.g., 보어절: Biber & Reppen, 1998, 시제: Granger, 1999, 서법: Aijmer, 2002), 어휘 (Altenberg & Granger, 2001; Kim, 2002; Ringbom, 1998a), 어법 (e.g., De Cock, 1998; Milton & Freeman, 1996), 담화 (e.g., Altenberg & Tapper, 1998; Cheng & Warren, 2000; Flowerdue, 1998; Petch-Tyson, 1998) 등 다양한 분야에서 연구가 이루어지고 있다. 제2언어 습득 문제들에 대한 연구도 소수 있는데, Tono (2000)는 일본인 영어 학습자 코퍼스를 분석하여 문법 형태소 관련 오류 비율이 제2언어 습득 연구자들이 제안한 형태소 습득 순서와 대체로 일치한다는 것을 밝혔다. Nesselhauf (2004)는 지금까지 이루어진 학습자 코퍼스 분석 연구들에서 반복적으로 나타나는 결과를 다음 세 가지로 요약하여 제시한다. 첫째, 학습자의 모국어가 학습자 언어의 여러 가지 측면에 흔히 생각하는 것보다 훨씬 더 큰 영향을 미친다는 것을 많은 연구들이 증명하고 있다. 둘째, 원어민의 글과 비교할 때 학습자들의 글은 여러 면에서 부적절하다는 것이 드러났다. 예를 들면 학습자들의 글에서는 주장을 완화하는 표현 (e.g., 'seem, suggest, indicate' 등의 동사나 'it would appear'와 같은 구문)이 원어민의 글에 비해 덜 사용되는 반면 (Flowerdue, 2000), 어떤 표지들 (e.g., 'obviously, certainly')은 원어민보다 훨씬 더 자주 사용되기도 한다 (Milton & Hyland, 1999). 또한 학습자 코퍼스에서는 작은 범위의 어휘 항목이 원어민 코퍼스에 비해 더 자주 사용되는데, 예를 들어 동사 'think'는 원어민 코퍼스에서는 만 단어당 여섯 번밖에 사용되지 않는데 비해 학습자 코퍼스에서는 21-30차례나 사용되었다 (Ringbom, 1998b). 또한 학습자들은 'people'이나 'thing'과 같은 포괄적인 어휘 항목들을 매우 빈번하게 사용하는 경향이 있다 (Ringbom, 1998b). 셋째, 학습자들은 학구적인 글을 쓰면서도 구어체 영어에 가까운 자질들(e.g., 대명사, 강조 접사, 상황 지시사 등)을 많이 사용한다는 것이 여러 차례 지적되었다 (Arts & Granger, 1998; Biber & Reppen, 1998; Granger & Rayson, 1998; Petch-Tyson, 1998). Altenberg와 Tapper (1998)는 스웨덴 영어학습자들이 'but'과 같은 소수의 구어체 접속사를 자주 사용하는 반면, 'therefore, thus, however, though, yet' 등과 같은 접속사들은 원어민에 비해 훨씬

구 패러다임에 균형이 회복되어 학습자들이 자연적으로 생성해낸 자료에 관심을 기울이게 되었다고 주장한다.

덜 사용한다는 것을 발견했다.

학습자 코퍼스는 중간언어 발달에 대한 정보를 제공해 줄 수 있는데 학습자들이 저지르기 쉬운 전형적인 오류들을 파악, 분류하고, 또 원어민들에 비해 더 많이, 혹은 덜 사용하는 자질들을 밝히는 등¹⁹⁾ 어떤 영역에서 가장 원어민과 다른 언어 수행을 보이는지를 밝혀냄으로써 이러한 부분에 교육을 집중할 수 있는 효과가 있다.²⁰⁾ 원어민에 비해 덜 사용되는 항목들의 경우에는 이러한 항목들에 초점을 맞추는 교재를 제작할 수 있을 것이다. 또한 원어민들이 사용하지 않는 부적절한 단어 결합 양식(e.g., 'absolutely easy')이 학습자 코퍼스에서 발견된다면, 원어민 코퍼스로부터 적절한 언어 형태들을 얻어내어 제시함으로써 학습에 도움을 줄 수도 있다.²¹⁾ 배경이 다른 학습자 집단간에 언어 사용이 어떤 차이를 보이는가를 조사하면 모든 학습자들에게 공통으로 나타나는 언어 습득 관련 현상과 특정 배경을 가진 학습자들에게만 보이는 현상을 구분할 수 있다. 이러한 연구 결과는 영어교재, 특히 쓰기 교재 개발에 유용하게 사용될 수 있다 (Granger, 1998; Granger, et al., 2002). 그리고 발달 단계에 대한 연구 결과는 목표어 자질들을 어떤 순서로 가르쳐야 할지 결정하는데 중요한 참고자료가 될 것이다.

한편 콩코던스 프로그램을 통하여 학습자 코퍼스를 보다 직접적으로 교수/학습 활동에 이용할 수도 있다. 즉, DDL의 교수방법을 학습자 코퍼스에 적용하는 것인데, Granger와 Tribble (1998)이 처음 이 방법을 제시

19) 원어민과 비교하여 학습자들이 더 많이 사용하는 (overuse) 언어자질과 덜 사용하는 (underuse) 자질들을 체계적이고 정밀하게 연구하는 것은 코퍼스 사용 이전에는 불가능했던 일이다 (Leech, 1998).

20) Hunston (2002)은 학습자 코퍼스 연구 결과들이 흥미롭기는 하지만 이를 교육적으로 어떻게 응용할 것인가는 쉽지 않은 문제임을 지적한다. 예를 들어 'thing'이라는 단어를 너무 많이 사용하는 학습자 집단에게 단순히 이 단어를 덜 사용하라고 말하는 것은 의미가 없으며, 교사는 원어민들이 언제 'thing' 대신 다른 표현을 선택하는지, 또 그 표현이 무엇인지 알아야 한다는 것이다. 따라서 교육적인 응용을 위해서는 좀 더 정밀한 연구가 필요하다.

21) 예를 들어 Kennedy (2003)는 British National Corpus를 이용하여 'extremely, greatly' 등 정도부사들의 언어 관계를 조사하고 그 결과를 영어교육에 반영할 것을 제안하고 있다.

한 이후 소수 연구자들이 이를 시도했다 (Flowerdue, 2001; Horváth, 2001; Milton & Hyland, 1999; Ragan, 2001, Seidlhofer, 2000). 가장 기본적인 방법은 오류가 발생할 것으로 예상되는 항목을 원어민 코퍼스와 학습자 코퍼스 각각에서 검색하여 얻은 콩코던스 선을 비교함으로써 학습자들이 원어민과 비원어민 언어 사용간의 차이점을 인지할 수 있도록 하는 것이다. Granger와 Tribble (1998)은 'accept'라는 동사 패턴의 예를 제시하는데, 원어민 코퍼스에서는 'accept' 뒤에 명사 (e.g., learn to accept *their traditions*) 혹은 that-절 (e.g., Hugo cannot accept *that the party line has changed*)이 따르는 반면 학습자 코퍼스에서는 to 부정사가 'accept'를 뒤따르고 있다 (e.g., He could never accept *to be inferior*). 교사는 이처럼 같은 언어 항목의 쓰임에 있어서 두 코퍼스가 어떻게 다른지 학습자 스스로 발견하게 하고 그에 따라 자신들도 보다 원어민에 가까운 언어 사용을 하도록 지도할 수 있다. 이러한 부정적 증거(negative evidence)는 특히 고급 수준의 학습자들에게, 그리고 화석화 되었거나 되고 있는 형태에 대해서 집중적으로 제공해주면 매우 효과적이다 (Granger, 1996). 학습자들에게 오류를 찾게 하거나 학습자와 원어민 언어 간의 차이를 발견하도록 하는 것은 학습자 자율성을 기르고 스스로 그러한 차이에 민감해지도록 훈련시킬 수 있다. 뿐만 아니라 학습자들은 이런 식으로 실수를 확인하고 분석하는 것에 동기가 유발된다고 한다 (Fan et al., 1999, Nessedhauf, 2004에서 재인용).

여러 가지 언어 영역 중에서 학습자 코퍼스 분석을 하기에 좋은 부분은 단어들의 동시발생양식인데, 특히 함께 쓰이는 단어들이 바로 옆에 붙어서 나타나는 경우, 즉 동사와 전치사, 형용사와 명사, 부사와 형용사, 동사의 보문(complementation) 같은 경우이다. 특정 단어를 잘못 사용한 경우나 너무 많이, 혹은 덜 사용한 경우들도 원어민 코퍼스와의 비교를 통해 밝힐 수 있다. 반면 문법 분야는 학습자 코퍼스를 통한 연구를 하기가 쉽지 않은데, 그 이유는 문법 구조들은 개별 단어에 비해 검색하기가 더 어렵고 또 바로 옆 문맥만 가지고서는 그것이 올바르게 사용된 것인지 아닌지 알기 힘들기 때문이다. 때로는 주어진 코퍼스 내에 알아보고자 하는 항목의 예가 전혀 없거나 또는 부족할 수도 있다. 따라서 만일 학생들과 함께 학습자 코퍼스 분석 활동을 하려면 교사는 미리 충분한 준비를 해야 한

다. 학습자 코퍼스의 한계점으로는 학습자의 수용적 능력을 연구할 수 없으며 텍스트에서 나타나지 않는 단어나 양식의 경우에는 학습자가 그것을 아는지 모르는지 알 방법이 없다는 것을 들 수 있다 (Nessedhauf, 2004). 따라서 매우 드물게 나타나는 현상을 연구하려면 실험적인 방법을 사용해야 한다.²²⁾

지금까지 이루어진 학습자 코퍼스 분석 연구들은 여러 가지 유용한 결과들을 보여주고 있지만 앞으로 개선되어야 할 여지가 많이 남아있다. 예를 들면 대부분의 연구들이 코퍼스에서 자동적으로 검색, 추출하여 쉽게 연구할 수 있는 현상들(즉, 개별 단어나 소수 문법 구조들)에 집중한 경향이 있다. 앞으로는 분석대상의 범위가 좀 더 확대되어야 하며 보다 종합적인 연구들이 이루어져야 할 것이다. 또한 많은 연구들이 관찰된 현상에 대해 깊이 있는 분석을 하기보다는 양적인 정보를 제공하는 데 그치는 경우가 많다. 그러나 학습자 언어의 특징을 제대로 파악하기 위해서는 양적 분석과 함께 수준 높은 질적 연구가 반드시 병행되어야 한다. 뿐만 아니라 다양한 수준의 학습자들, 그리고 서로 다른 담화 양식들에 대한 비교 연구도 필요하다.

IV. 결론

코퍼스를 이용하는 것은 이제 언어와 관련된 모든 분야에서 이루어지고 있다고 해도 과언이 아니다. 어휘나 문법 연구뿐만 아니라 담화 분석, 화용론, 언어 변이 연구, 역사언어학, 사회언어학, 심리언어학, 문체론 등의 분야에서도 전자 코퍼스를 이용한 연구들이 활발히 진행되고 있다. 영어 교육 분야에서도 코퍼스의 유용성이 인식되면서 코퍼스 연구들의 영향은 이미 사전, 문법서와 기타 교재들을 통해 교육 현장에 미치기 시작하였다. 또한 코퍼스와 코퍼스 탐색을 위한 컴퓨터 도구들이 더 쉽게 사용 가능해지면서 외국어 교실에 큰 도움을 줄 수 있게 되었다. 하지만 앞서 지적한

22) 언어 적성이나 동기 등 학습자 특성의 역할에 대한 연구도 학습자 코퍼스로는 이루어지기 힘들며, 특정한 교수법이 습득에 미치는 영향에 대한 연구 역시 실험적 절차에 의해서 더 쉽게 밝혀질 수 있을 것이다. 따라서 학습자 언어 분석을 위한 최선의 방법은 코퍼스 분석과 실험 방식의 결합일 수도 있다 (Nesselhauf, 2004).

대로 우리나라의 경우 영어교육 분야에서 코퍼스에 대한 인식과 관심이 상대적으로 부족한 실정이다. 따라서 우리나라 영어교육 전문가들이 코퍼스의 활용가치에 대해 제대로 인식하고 그 효용가치를 극대화하기 위해 노력하는 것이 필요하다. 본 논문은 이러한 필요성을 부각시키기 위한 목적으로 코퍼스가 영어교육에 활용될 수 있는 여러 가지 방식들을 소개하였고, 이제 앞으로의 과제 몇 가지를 제시하면서 이 글을 끝맺고자 한다.

먼저 교육 내용면에서 교사는 영어의 실제 사용 양식을 알기 위해 기존의 코퍼스 연구들을 참조하거나 또는 코퍼스를 직접 사용할 수 있어야 한다. 그리고 교육적인 결정을 내리는 데 있어서 더 이상 직관이나 관행에만 의존하지 않고 다양한 상황(구어/문어, 공식적/비공식적 상황 등)에서의 실제 언어 사용 양식에 근거하여 이러한 결정을 내려야 할 것이다. 교과서 집필자나 교재 개발자들은 코퍼스를 참조하여 진정성 있는 내용을 포함시키도록 해야 한다. 앞으로 남은 중요한 과제 중 하나는 어떻게 하면 코퍼스를 통해 발견된 (또 앞으로 더 발견될) 새로운 정보, 즉 빈도에 대한 정보와 언어 사용 양식에 대한 지식을 영어교육 관련 교재 개발 및 교수요목 선정에 잘 포함시킬 것인가 하는 것이다.

코퍼스를 외국어 수업에서 교수방법으로 이용하는 것은 아직 초기 단계에 머물러 있긴 하지만, 이와 관련하여 효율적인 코퍼스 활용 방안에 대해 여러 가지 제안들이 있다 (Carter, Hughes, & McCarthy, 1998; Fox, 1998; Johns, 1997; Tribble & Jones, 1997; Willis, 1998). 외국어 교실 상황에서 코퍼스를 사용해 보도록 교사들을 독려하는 연구자들이 많이 있는데 (Barnbrook, 1996; Flowerdew & Tong, 1994; Johns, 1994; Simpson & Swales, 2001; Wichmann et al., 1997), 한국에서는 아직까지 코퍼스를 교수나 학습 방법 차원에서 이용하고 있는 경우가 거의 없다. 이 분야가 비교적 최근에 개발되어 많은 사람들에게 생소하다는 것이 아마도 가장 큰 이유일 것이다. 따라서 현직 혹은 미래 교사들에게 코퍼스에 대해 소개하고 그 활용 가치와 실제 사용법을 널리 알리는 것이 필요하다. 외국의 경우 교사교육 프로그램에서 코퍼스 사용법을 가르칠 뿐만 아니라 미래 교사들의 비판적 언어 인식을 기르는 데 코퍼스를 이용하기도 하는데 (O'Keefe & Farr, 2003), 우리나라에서도 사범대학의 교과과정과 교원연수 과정에 이러한 내용을 포함시키는 것이 바람직할 것이다. 코퍼스를 직

접적인 교육의 수단으로 활용하는 방법들이 얼마나 효과적인가 하는 것은 많은 교사들이 실제로 코퍼스를 교육 도구로 사용해 본 뒤, 시간이 지나야만 분명해질 것이다. 학습자들로 하여금 직접 콩코턴스 프로그램을 사용하게 하는 방식과 교사가 인쇄물로 제공하는 방식 중 어느 것이 학습에 더 효과적인지, 또 이러한 학습의 효과는 학습자의 능숙도나 학습 스타일과 어떻게 상호작용하는지 등에 대한 정보도 필요한 만큼 이를 알아보기 위한 연구들도 요청된다.

학습자 코퍼스는 분명히 영어교육에 중요한 기여를 할 수 있다. 무엇보다도 특정 집단의 학습자들이 전형적으로 어려움을 느끼는 부분을 밝힘으로써 교재 향상과 교육과정 개발에 기여할 수 있다. 교실에서는 학습자 코퍼스가 학습자들에게 부정적 증거를 스스로 발견하는 기회를 제공해 줌으로써 역시 유용하게 사용될 수 있다. 그런데 다른 국가 영어 학습자들의 코퍼스를 바탕으로 한 연구 결과는 우리에게 유용한 시사점을 제공해 줄 수는 있겠지만 결코 그대로 적용될 수는 없다. 따라서 한국인 학습자들을 대상으로 하는 학습자 코퍼스의 구축과 분석이 매우 중요하다. 앞서 소개한 대로 이미 유럽과 아시아 일부 국가에서는 대규모 영어 학습자 코퍼스가 구축되고 있으나 한국의 경우 그렇지 못하다. 물론 개별 연구자들은 나름대로 여러 형태의 학습자 코퍼스를 구축하여 연구에 이용하고 있을 것이다. 하지만 앞으로는 특정 원칙에 입각해 좀 더 체계적으로 대규모 학습자 코퍼스를 설계, 구축하는 것이 필요하다고 생각된다. 한편 기존의 학습자 코퍼스는 대개 학습자들이 쓴 글을 모아서 이루어진 문어 코퍼스이다.²³⁾ 이는 학습자들의 문어 자료가 구어 자료보다 수집하기 훨씬 더 편리하고 시간적, 경제적으로도 절약된다는 면을 고려할 때 당연한 결과라고 할 수 있다.²⁴⁾ 하지만 앞으로는 구어 학습자 코퍼스 역시 구축되어

23) 구어 학습자 코퍼스는 일본 학습자들의 인터뷰 자료인 Standard Speaking Test Corpus와 다양한 국가 학습자들의 인터뷰로 이루어진 Louvain International Database of Spoken English Interlanguage 등 소수만이 존재한다.

24) 학습자 코퍼스 뿐만 아니라 원어민 코퍼스의 경우에도 구어 코퍼스는 문어 코퍼스에 비해 그 수가 훨씬 더 적다. 그 주 원인은 코퍼스 구축에 드는 비용, 적절하고 대표성 있는 구어 자료 상황에 접근하기가 쉽지 않다는 것, 녹음의 질, 그리고 전사(transcription)에 드는 시간 등이다 (McCarthy, 1998).

야 한다. 원어민 코퍼스 연구 결과들은 언어 사용 양식이 구어와 문어 등 언어사용역에 따라 현저하게 다를 수 있다는 것을 보여 준다. 따라서 학습자들의 언어 사용이 이러한 담화 양식별 특성을 제대로 반영하고 있는지를 확인하기 위해서는 학습자 코퍼스에 포함되는 자료 역시 다양해질 필요가 있다. 그리고 초급부터 고급 수준까지 다양한 능숙도에 속하는 학습자 코퍼스가 있어야 영어학습의 발달단계에 대한 연구가 가능해질 것이다. 만일 오랜 시간에 걸쳐 개별 학습자들의 각 학습단계별 자료를 얻어 내어 통시적(longitudinal) 학습자 코퍼스가 구축될 수 있다면 더욱 바람직할 것이다. 물론 위에서 언급한 코퍼스의 구축은 결코 쉬운 일이 아니다. 이를 위해서는 시간적, 재정적 뒷받침이 선결되어야 할 과제이며, 코퍼스 규모가 커질수록 여러 연구자들의 협력과 교육 관련 기관(대학이나 국가 교육 담당 부서 등)의 지원이 더욱 요청될 것이다. 하지만 이런 여러 가지 어려움에도 불구하고 구축이 이루어지기만 한다면 영어교육을 위하여 매우 효과적이고 구체적인 정보를 제공할 수 있으리라 기대된다.

코퍼스는 원어민의 직관으로 인지할 수 없는 다양하고 새로운 언어 사용의 양식들을 밝혀냄으로써 언어에 대한 우리의 기존 지식을 수정하거나 보다 정교하게 다듬어 가고 있다. 코퍼스 연구들이 발견해 낸 이러한 증거들은 영어교육의 내용면에서 이미 긍정적인 영향을 미치기 시작하였으며 그 영향력은 앞으로도 계속될 것이다. 무엇보다도 코퍼스는 교사와 학습자들이 목표어가 의사소통 상황에서 실제로 사용되는 방식을 스스로 탐구해 볼 수 있도록 기회를 제공해 줌으로써 보다 효과적인 교수와 자율적인 학습에 크게 기여할 수 있다. 이와 더불어 학습자 코퍼스 분석을 통해 한국 학습자들만의 고유한 학습 과정이나 어려움을 파악하여 이들에게 가장 적합한 교재와 교육 과정을 선택하고 개발할 수 있다면 코퍼스의 가치는 더욱 커질 것이다. 이렇게 우리 앞에 펼쳐진 다양한 가능성을 얼마나 적극적으로 잘 활용하여 한국의 영어교육을 한 단계 더 발전시키느냐 하는 것은 이제 우리들의 몫이다.

참 고 문 헌

- Aarts, J., & Granger, S. (1998). Tag sequences in learner corpora: A key to interlanguage grammar and discourse. In S. Granger (Ed.), *Learner English on computer* (pp. 132-142). London: Longman.
- Aijmer, K. (2002). Modality in advanced Swedish learners' written interlanguage. In Granger et al. (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 55-76). Amsterdam/Philadelphia: John Benjamins.
- Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of *make* in native and non-native student writing. *Applied Linguistics*, 22(2), 173-194.
- Altenberg, B., & Tapper, M. (1998). The use of adverbial connectors in advanced Swedish learners' written English. In S. Granger (Ed.), *Learner English on computer* (pp. 80-93). London: Longman.
- Aston, G. (2000). Corpora and language teaching. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 7-17). Berlin: Peter Lang.
- Barnbrook, G. (1996). *Language and computers: A practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.
- Biber, D. (1990). Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing* 5, 257-269.
- Biber, D., & Reppen, R. (1998). Comparing native and learner perspectives on English grammar: A study of complement clauses. In S. Granger (Ed.), *Learner English on computer*

- (pp. 145-158). London: Longman.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman Grammar of Spoken and Written English*. London: Longman.
- Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *TESOL Quarterly*, 36(1), 9-48.
- Cambridge International Dictionary of English* (1995). Cambridge: Cambridge University Press.
- Carter, R. (1998). Orders of reality: CANCODE, communication, and culture. *ELT Journal* 52(1), 43-56.
- Carter, R., & McCarthy, M. (1997). *Exploring spoken language*. Cambridge: Cambridge University Press.
- Carter, R., Hughes, R., & McCarthy, M. (1998). Telling tails: Grammar, the spoken language and materials development. In B. Tomlinson (Ed.), *Materials development in language teaching* (pp. 67-86), Cambridge: Cambridge University Press.
- Cheng, W., & Warren, M. (2000). The Hong Kong Corpus of Spoken English: Language learning through language description. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 133-144). Berlin: Peter Lang.
- Collins Cobuild English Language Dictionary*. (1987). London: Collins.
- Cook, G. (1998). The use of reality: A reply to Ronald Carter. *English Language Teaching Journal*, 52, 57-63.
- De Cock, S. (1998). A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English. *International Journal of Corpus*

Linguistics, 3(1), 59-80.

- De Cock, S., Granger, S., Leech, G., & McEnery, T. (1998). An automated approach to the phrasicon of EFL learners. In S. Granger (Ed.), *Learner English on computer* (pp. 67-79). London: Longman.
- Donley, K. M., & Reppen, R. (2001). Using corpus tools to highlight academic vocabulary in sustained content language teaching. *TESOL Journal* 10(2/3): 7-12.
- Fan, M., Greaves, C., & Warren, M. (1999). Identifying characteristic patterns in students' writing using a corpus of learner data. In R. Berry, B. Asker, & K. Hyland (Eds.), *Language analysis, description and pedagogy* (pp. 147-161). Hong Kong: Language Centre HKUST.
- Flowerdue, L. (1998). Integrating 'expert' and 'interlanguage' computer corpora findings on causality: Discoveries for teachers and students. *English for Specific Purposes* 17(4), 329-345.
- Flowerdue, L. (2000). Investigating referential and pragmatic errors in a learner corpus. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 145-154). Berlin: Peter Lang.
- Flowerdue, L. (2001). The exploitation of small learner corpora in EAP materials design. In M. Ghadessy, A. Henry, & R. Roseberry (Eds.), *Small corpus studies and ELT: Theory and practice* (pp. 363-379). Amsterdam: John Benjamins.
- Flowerdew, L., & Tong, A. (Eds.) (1994). *Entering text*. Hong Kong: Language Center.
- Fox, G. (1998). Using corpus data in the classroom, In B. Tomlinson (Ed.), *Materials development in language teaching* (pp. 25-66). Cambridge: Cambridge University Press.
- Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.

- George, H. V. (1963). *Report on a verb-form frequency count*.
Monograph 1. Hyderabad: Central Institute of English.
- Granger, S. (1996). Exploiting learner corpus data in the classroom: Form-focused instruction and data-driven learning. Paper presented at TALC 1996, Lancaster, 9-12 August 1996.
- Granger, S. (Ed.) (1998). *Learner English on computer*. London: Longman.
- Granger, S. (1999). Use of tense by advanced EFL learners: Evidence from an error-tagged computer corpus. In H. Hasselgard & S. Oksefjell (Eds.) *Out of corpora: Studies in honour of Stig Johansson* (pp. 191-202). Amsterdam: Rodopi.
- Granger, S., & Rayson, P. (1998). Automatic profiling of learner texts. In S. Granger (Ed.), *Learner English on computer* (pp. 119-131). London: Longman.
- Granger, S., & Tribble, C. (1998). Learner corpus data in the foreign language classroom: Form-focused instruction and data-driven learning. In S. Granger (Ed.), *Learner English on computer* (pp. 199-209). London: Longman.
- Granger, S., Hung, J., & Petch-Tyson, S. (Eds.) (2002). *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam: Benjamins.
- Holmes, J. (1988). Doubt and certainty in ESL textbooks. *Applied Linguistics* 9(1), 21-44.
- Horváth, J. (2001). *Advanced writing in English as a foreign language: A corpus-based study of processes and products*. Pécs: Lingua Franca Csoport.
- Hughes, R., & McCarthy, M. (1998). From sentence to discourse: Discourse grammar and English language teaching. *TESOL Quarterly* 32(2), 263-287.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

- Johns, T. (1991). Should you be persuaded Two samples of data-driven learning materials. In T. Johns & P. King (Eds.), *Classroom Concordancing [ELT Journal, 4]* (pp. 1-16).
- Johns, T. (1994). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.) *Perspectives on pedagogical grammar* (pp. 293-313). Cambridge: Cambridge University Press.
- Johns, T. (1997). Contexts: The background, development and trialling of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 100-115). London: Longman.
- Joos, M. (1964). *The English verb: Form and meaning*. Madison: University of Wisconsin Press.
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London: Longman.
- Kennedy, G. (2003). Amplifier collocations in the British National Corpus: Implications for English language teaching. *TESOL Quarterly* 37(3), 467-487.
- Kim, M.-H. (2002). Uses of *make* in Korean EFL learner writing: A corpus-based study. *English Teaching*, 57(4), 297-314.
- Leech, G. (1991). The state of the art in corpus linguistics. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 8-29). London: Longman.
- Leech, G. (1998). Preface. In S. Granger (Ed.), *Learner English on computer* (pp. xiv-xxii). London: Longman.
- Longman Language Activator* (1993). London: Longman.
- Longman Dictionary of Contemporary English* (1995). London: Longman.
- Longman Essential Activator*. (1997). London: Longman.
- McCarthy, M. (1998). *Spoken language and applied linguistics*.

- Cambridge: Cambridge University Press.
- McEnery, T., & Wilson, A. (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- Milton, J., & Freeman, R. (1996). Lexical variation in the writing of Chinese learners of English. In C. E. Percy, C. E. Meyer, & I. Lancashire (Eds.), *Synchronic corpus linguistics: Papers from the sixteenth international conference on English language research on computerized corpora (ICAME 16)* (pp. 121-131). Amsterdam: Rodopi.
- Milton, J., & Hyland, K. (1999). Assertions in students' academic essays: A comparison of English NS and NNS student writers. In R. Berry, B. Asker, & K. Hyland (Eds.), *Language analysis, description and pedagogy* (pp. 147-161). Hong Kong: Language Centre HKUST.
- Mindt, D. (1996). English corpus linguistics and the foreign language teaching syllabus. In J. Thomas & M. Short (Eds.), *Using corpora for language research: Studies in the honour of Geoffrey Leach* (pp. 232-247). London: Longman.
- Mindt, D. (1997). Corpora and the teaching of English in Germany. In G. Knowles, T. McEnery, S. Fligelstone, & A. Wichman, (Eds.) *Teaching and language corpora* (pp. 40-50). London: Longman.
- Nessedhauf, N. (2004). Learner corpora and their potential for language teaching. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 125-152). Amsterdam/Philadelphia: John Benjamins.
- O'Keeffe, A., & Farr, F. (2003). Using language corpora in initial teacher education: pedagogic issues and practical applications. *TESOL Quarterly* 37(3), 389-418.
- Ota, A. (1963). *Tense and aspect of present-day American English*. Tokyo: Kenkyusha.

- Owen, C. (1996). Do concordances require to be consulted? *ELT Journal* 50(3), 219-224.
- Petch-Tyson, S. (1998). Writer/reader visibility in EFL written discourse. In S. Granger (Ed.), *Learner English on computer* (pp. 107-118). London: Longman.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*. London: Longman.
- Ragan, P. H. (2001). Classroom use of a systemic functional small learner corpus. In M. Ghadessy, A. Henry, & R. Roseberry (Eds.), *Small corpus studies and ELT: Theory and practice* (pp. 207-236). Amsterdam: John Benjamins.
- Reppen, R., & Simpson, R. (2002). Corpus linguistics. In N. Schmitt (Ed.), *An introduction to applied linguistics* (pp. 92-111). New York: Oxford University Press.
- Ringbom, H. (1998a). High-frequency verbs in the ICLE corpus. In A. Renouf (Ed.), *Explorations in corpus linguistics* (pp. 191-200). Amsterdam: Rodopi.
- Ringbom, H. (1998b). Vocabulary frequencies in advanced learner English: A cross-linguistic approach. In S. Granger (Ed.), *Learner English on computer* (pp. 41-52). London: Longman.
- Seidlhofer, B. (2000). Operationalizing intertextuality: using learner corpora for learning. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 207-223). Berlin: Peter Lang.
- Simpson, R., & Swales, J. (Eds.) (2001). *Corpus linguistics in North America: Selections from the 1999 symposium*. Ann Arbor, MI: University of Michigan Press.
- Sinclair, J. (1991). *Corpus, concordance, and collocation*. Oxford: Oxford University Press.
- Svartvik, J. (1966). *On voice in the English verb*. The Hague: Mouton.

- Thorndike, E. L. (1921). *Teacher's wordbook*. New York: Columbia Teachers College.
- Thorndike, E. L., & Lorge, I. (1944). *The Teacher's wordbook of 30,000 words*. New York: Columbia Teachers College.
- Tono, Y. (2000). A computer learner corpus based analysis of the acquisition order of English grammatical morphemes. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 123-132). Berlin: Peter Lang.
- Tribble, C., & Jones, G. (1997). *Concordances in the classroom: A resource book for teachers*. London: Longman.
- Wang, S. (1991). *A corpus study of English conditionals*. Unpublished MA thesis. Victoria University of Wellington.
- Wichmann, A., Fligelstone, S., McEnery, T., & Knowles, G. (Eds.) (1997). *Teaching and language corpora*. London: Longman.
- Widdowson, H. G. (1991). The description and prescription of language. In J. Alatis (Ed.), *Georgetown university round table in language and linguistics. Linguistics and language pedagogy: The state of the art* (pp.11-24). Washington DC: Georgetown University Press.
- Widdowson, H. G. (2000). On the limitations of linguistics applied. *Applied Linguistics*, 21(1), 3-25.
- Willis, J. (1998). Concordances in the classroom without a computer. In B. Tomlinson (Ed.), *Materials development in language teaching* (pp. 44-66). Cambridge: Cambridge University Press.

부록: 코퍼스 관련 웹 사이트

인터넷상에서 사용할 수 있는 코퍼스:

LDC (Linguistic Data Consortium) Online

<http://www ldc.upenn.edu/ldc/online/>

문어 코퍼스: Brown Corpus, Wall Street Journal, New York Times 등.

구어 코퍼스: Switchboard transcripts (전화 대화)

MICASE (Michigan Corpus of Academic Spoken English)

<http://www.lsa.umich.edu/eli/micase/index.htm>

British National Corpus Online (제한된 양만 검색 가능)

<http://sara.natcorp.ox.ac.uk/lookup.html>

Virtual Language Center

<http://vlc.polyu.edu.hk/concordance/>

주요 코퍼스 웹사이트:

British National Corpus

<http://www.natcorp.ox.ac.uk/>

American National Corpus (구축 중)

<http://americannationalcorpus.org/>

International Corpus of English

<http://www.ucl.ac.uk/english-usage/ice/>

International Corpus of Learner English

<http://www.fltr.ucl.ac.be/FLTR/GERM/ETAN/CECL/cecl.html>

Longman Learners' Corpus

<http://www.longman-elt.com/dictionaries/corpus/lclearn.html>

코퍼스 보관 및 판매 기관:

LDC

<http://ldc.upenn.edu/>

ICAME (International Computer Archive of Modern and Medieval English)

<http://helmer.aksis.uib.no/icame.html>

콩코던스 프로그램:

MonoConc

<http://www.athel.com/mono.html#mono>

Wordsmith Tools

<http://www.lexically.net/wordsmith/>

무료 콩코던스 프로그램:

WinConcord

<http://www.linglit.tu-darmstadt.de/downloads/wconcord.zip>

AntConc

<http://www.f.waseda.jp/anthony/software/antconc2.5.3.exe>

코퍼스 언어학 관련 학술지:

International Journal of Corpus Linguistics

http://www.benjamins.com/cgi-bin/t_seriesview.cgi?series=IJCL

ICAME Journal

<http://helmer.hit.uib.no/journal.html>