

# 우리나라 중·상급학습자 영어의 특징 : 말뭉치 언어학적 분석\*

안성호(한양대학교)

이영미(한양대학교)\*\*

Ahn, Sung-Ho and Lee, Young-Mi. 2004. **Characteristics of Intermediate/Advanced Korean Inter-Englishes: A Corpus-Linguistic Analysis.** *Korean Journal of English Language and Linguistics* 4-1, 83-102. The purpose of this paper is to find out some major characteristics of intermediate-advanced Korean learners' English by corpus- linguistically analyzing their essays in comparison with native speakers'. We construct a corpus of CBT TOEFL essays by Korean learners, NNS1 (94076 words in 402 texts), and its sub-corpus, NNS2 (14291 words in 45 texts), and then a corpus of model essays written or meticulously edited by native speakers, NS (14833 words in 35 texts). We compare NNS1 and NNS2 with NS, and with some other corpora, in terms of high-frequency words, and show that Korean learners' writings have more features of informal writing than those of formal writing, which is in accord with the reports in Granger (1998) that EFL writings by European advanced learners are characterized by informality.

**Key Words:** learner English, corpus, informality, Korean inter-English, inter-language

## 1. 서론

Meyer(2002, p. xi)에 따르면, 말뭉치(Corpus)는 “일반적 언어분석

---

\*이 논문의 일부는 The 2003 KASELL International Conference(6월 25-26, 한양대학교)에서 발표된 바 있다. 발표 논문에 대한 논평을 담당해주신 김용진교수(숭실대)와 정영국교수(국제영어대학원대학)께, 본 원고의 수정에 많은 도움을 주신 익명의 《영어학》 논평자 세 분께, 이 연구를 해가도록 격려와 논의의 시간을 아끼지 않으신 한양대학교 영어교육과 동료 교수들께, 그리고 함께 연구와 논의를 해 온 이은연 선생께 감사를 드린다. 그러나 물론 모든 잘못은 저자들의 책임임을 밝혀두는 바이다.

\*\*제1저자(안성호); 공동저자(이영미)

의 기반이 되는 텍스트(혹은 그 일부분)의 모듈”으로 이해될 수 있다. 말뭉치를 토대로 한 언어분석은 그 언어의 구조와 사용에 대한 정확하고 신뢰성 있는 기술을 가능하게 하기 때문에, 오래 전부터 참고 문법서의 저술이나(Jespersen, 1909-49; Curme, 1947), 옥스퍼드영어사전<sup>1)</sup>과 같은 사전의 편찬 등에 널리 이용되어 왔다.

현대적 의미의 말뭉치 언어학에서는 전자화된 대용량의 말뭉치를 컴퓨터 프로그램을 통하여 처리함으로써, 단어나 단어 연쇄의 빈도수가 보여주는 패턴을 통하여 언어사용과 관련된 새로운 측면을 보여줄 수 있다(Quirk et al., 1972, 1985; Collins COBUILD English Dictionary). 또, 연구목적에 맞게 말뭉치를 이용하거나 구축함으로써 영어의 전반적 특징은 물론이고 문어와 구어, 장르별, 연령별, 성별, 사회 문화적 배경에 따른 영어의 특징에 대한 분석도 가능해졌다.

특히 ICLE<sup>2)</sup>와 같은 학습자영어 말뭉치를 통하여, 대용량의 학습자 영어 자료와 원어민의 영어 자료를 상호 비교 분석할 수 있게 되었다. 이로써 학습자 영어에 나타난 모국어의 영향이나 모국어와 관계없이 나타나는 발달단계의 특징 등, 학습자 영어의 새로운 측면 뿐 아니라, Granger(1998b)가 강조한 바와 같이, 학습자 언어의 연구에 있어서 오류분석방법이 다를 수 없었던 학습자의 회피(Avoidance) 전략 사용 여부도 밝혀 낼 수 있게 되었다는 점이 특기할만하다 하겠다.

Granger(1998)에 보고된 여러 연구들은 유럽의 상급영어학습자 말뭉치인 ICLE를 바탕으로, 모국어가 다른 유럽의 학습자영어와 원어민의 영어를, 고빈도 어휘(high-frequency words)를 중심으로 연어나 품사별 빈도수 등을 다각적으로 분석, 비교하고 있다. 특히, Ringbom(1998)과 Granger & Rayson(1998)은 유럽 여러 나라의 상급학습자들이 격식체 글쓰기에서 비격식체, 구어체에 가까운 어휘

<sup>1</sup>옥스퍼드영어사전(Simpson & Weiner, 1989)에는 약 5백만 개의 인용문구가 있으며 모두 수작업으로 수집, 정리, 분석되었으며, 완성하는데 50여년이 걸렸다.

<sup>2</sup>ICLE(International Corpus of Learner English)는 EFL 상급영어학습자들의 논쟁적 에세이 말뭉치를 구축하려는 프로젝트로서 현재 19개 나라 학습자 영어의 하위 말뭉치들이 구축되고 있다고 한다. 각 나라 당 20만 어절을 담고 있다.

와 문체를 사용하고 있음을 밝히고 있다. 그들은 이 경향이 여러 나라 영어학습자들에게 나타나므로 모국어의 영향이라기보다는 영어습득단계에서 나타나는 학습자 영어의 공통적인 특징이라는 가설을 제안하고 있다.

본 연구의 목적은 Granger(1998)에서 보고된 연구방법론을 따라서 우리나라 중·상급 영어학습자들이 쓴 에세이를, 첫째, 원어민의 에세이와, 둘째, 다른 언어를 모국어로 하는 영어학습자들의 에세이와, 말뭉치 언어학적으로 비교, 분석함으로써, 한국인 학습자 영어를 대상으로 하여 Ringbom(1998)과 Granger & Rayson(1998)의 위가설을 점검하는 것이다.

제 2절에서 보고되는 바와 같이, 우선, 한국인 학습자 영어의 말뭉치와 그에 상응하는 원어민 영어의 말뭉치를 구축한다. 구축된 말뭉치의 연구 과정 및 결과는(제 3절) 다음과 같다. 우선, 어휘 빈도수를 중심으로 두 말뭉치에 나타난 언어적 특징을 상호 비교한다. 둘째, 그 결과를 기존의 연구가 밝힌, 타 언어를 모국어로 하는 학습자 영어의 특징과 비교함으로써, 관사, 전치사 및 부정사적 to, 인칭대명사, 본동사, 축약 등의 측면에서 한국학습자들의 영어가 타 언어 학습자 영어의 전반적 특성과 비격식적이라는 점에서 유사함을 밝힌다. 제 4절에서는 요약 및 결론을 제시하고, 교육적인 시사점에 대하여 논한다.

## 2. 말뭉치의 원천 및 세부사항

본 연구에서 자료로 택한 학습자 영어 자료는 CBT TOEFL의 에세이 작문을 준비하는 한국 학습자들의 모의시험 글이며, 인터넷 사이트 <http://www.cbttoefl.co.kr>(2002. 8 - 2002. 11)에서 수집한 것이다.

각각의 글쓰기에 대한 CBT TOEFL의 제한은 다음과 같다: (가) 실제 TOEFL에서 요구하는 정도의 길이(약 300단어 수준)보다 지나치게 길거나 짧지 않아야 한다; (나) 제한 시간은 30분이다; (다) 주어진 시간 안에 작성자는 짜임새 있고 의미전달이 명확하게 글을 발전, 전개시켜야 한다.

그리고 이것은 4가지 유형으로 분류된다: '동의 여부'(agreeing or

disagreeing), '선호서술'(stating a preference), '설명하기'(giving an explanation), 그리고 '주장하기'(making an argument)가 그것이다.

수집한 학습자 영어 자료는 13개의 과제 각각에 대한 300단어 내외의 에세이 14-48개로 이루어져 있으며, 총합은 402개의 에세이 텍스트의 94,076개 단어이다(부록 1 참조). CBT TOEFL의 평가 점수는 0점에서 6.0까지로 0.5점씩 차이가 나며, 5.0점일 경우는 상위 약 9%를 차지하며 4.0점일 경우는 상위 약 40%를 차지한다. 수집된 자료는 모의시험 에세이 중 CBT TOEFL의 평가 기준에 근거하여 채점한 결과, 적어도 4.0이상의 글이며 대부분은 5.0 이상을 받은 글이다. 따라서 그들의 영어 수준을 중·상급으로 규정하고자 하는 것이다. 우리는 이렇게 수집된 비원어민의 말뭉치를 편의상 NNS1(Non-Native Speaker1)이라고 하겠다.

NNS1과 동일한 형식의 원어민 자료를 수집하고자, 미국에 기반을 둔 인터넷 사이트 <http://www.testmagic.com>에서 CBT-TOEFL의 모범 에세이를 수집하였다. 수집에 한계가 있어 22개의 과제에 대하여 각각 1-5개의 에세이로 구성하였으며, 34개 텍스트의 14,833 단어로 이루어져 있다(부록 2 참조). 우리는 NNS1과 동일한 형식의 원어민영어 말뭉치를 NS(Native Speaker)<sup>3)</sup>라고 부르겠다.

NS가 약 15,000단어로 NNS1에 비해 상당히 적기 때문에, NS와 가장 비슷한 자료를 만들기 위해 NNS1의 13개의 과제 각각에 대하여 글 3-4편을 무작위로 추출하였다. 45개의 text의 14,291 단어의 하위 말뭉치인 이것을 NNS2라고 하겠다. NS와 NNS2는 크기 뿐 아니라 내용면에서도 유사하게 하기 위해 동의 여부, 선호서술, 설명, 주장 등 각각의 유형의 편수를 맞추었다.

이 연구를 위해 구축된 말뭉치를 표로 정리하면 다음과 같다.

표 1. 구축된 말뭉치의 크기

	NS	NNS1	NNS2
과제종류	22	13	13
어휘수	14,833	94,076	14,291

<sup>3)</sup>이 말뭉치는 원어민의 글과, 원어민 전문가가 수정, 편집한 영어학습자들의 글로 이루어져 있다. 모든 텍스트가 원어민의 글이 아니라는 한계가 있지만, 원어민 글쓰기 전문가에게 수정, 편집되어 모범 에세이로 제시되고 있어 원어민의 영어와 동일한 성격을 지닌 것으로 간주하였다.

### 3. 연구 과정 및 결과

#### 3.1. 연구 진행 과정

본 연구에서는 신뢰성을 위해 NS와 크기가 유사한 NNS2를 주 분석대상으로 삼아 분석을 실시한 후, 그 결과를 NS와, 그리고 기존의 다른 말뭉치의 분석 결과와 비교한다. 그 진행과정은 다음과 같다. 우선, 소프트웨어 WordSmith를 사용하여 NNS1, NNS2, NS, ICE-GB<sup>4)</sup> 중 대학생 에세이 모음, Brown<sup>5)</sup>의 어휘 빈도수를 비교하고 주요한 단어들의 주석을 만들어 조사한다. 둘째, 빈도수가 높은 어휘를 중심으로 우리나라 학습자들이 원어민에 비해 과도사용(overuse) 혹은 과소사용(underuse)되는 주요어휘를 비교 분석한다. 여기에서는 주로 NNS2의 결과를 기반으로 한다. 셋째, 과도사용 혹은 과소사용하는 어휘들을 관사, 전치사와 부정사의 to, 인칭대명사, 본동사, 축약 등 유형별로 분류하여 NS의 대응하는 품사 및 어휘들과 비교 분석한다. 넷째, ICE-GB 등의 논쟁적 에세이 혹은 격식체 글의 말뭉치와 비교하여 우리나라 학습자들의 언어특징을 살펴보고, 필요시 타당한 해석을 위해 Brown, LLC<sup>6)</sup>와도 부분적인 비교를 한다.

#### 3.2. 결과 및 해석

각각의 말뭉치에 나타난 상위 15위까지의 어휘들이 표 2에 제시되어 있다. NS와 비교했을 때 NNS1, NNS2에서 과도사용(overuse)되는 단어는 a, to, I, my, can, be, their, they 등이며, 과소사용(underuse)되는 단어는 the, of, in, and 등이다. 빈도수가 비교적 비슷한 단어들은 is, it, with, are, not, people 등이다.

빈도수에서 차이를 보이는 단어들을 관사, 전치사 및 부정사적

<sup>4</sup>ICE-GB는 International Corpus of English-Great Britain의 약자로 200개의 영문 텍스트와 300개의 영어 구어 텍스트를 지니고 있다. 이 말뭉치를 사용하도록 해 준 박명관 교수(동국대)께 감사사를 드린다.

<sup>5</sup>Brown 대학에서 만들어진 최초의 근대적 균형잡힌 말뭉치로 100만 어절을 담고 있다.

<sup>6</sup>LLC는 London-Lund Corpus로 100개의 텍스트에 50만어절로 이루어진 영국 구어영어 말뭉치이다. 더 자세한 내용은 <http://khnt.hit.uib.no/icame/manuals/LONDLUND/INDEX.HTM>를 참조하기 바란다.

to, 인칭대명사, 본동사 등 품사별로 비교 분석함으로써, 우리나라 학습자의 전반적인 언어적 특징들이 무엇인지 살펴보기로 하자.

### 3.2.1. 관사

a와 the에 대한 기존의 연구를 보면, 격식체 글에서는 정관사 the가 많이 사용되는 반면 부정관사 a는 적게 사용됨을 알 수 있다. Johansson(1985)은, **LOB(The Lancaster-Oslo/Bergen Corpus)**의 어휘 빈도수를 분석한 결과, 학문적인 글(Category J--learned texts)에서 정관사의 빈도수가 가장 높고, 부정관사의 빈도수가 가장 낮은 사실을 밝혔다.

표 2에서 관사들을 살펴보자.

표 2. 각각의 말뭉치에 나타난 상위 20위의 어휘들

순 위	NS	NNS1	NNS2	ICE-GB
1	the(4.21%)	to(3.83%)	to(3.76%)	the(7.12%)
2	to(3.47%)	a(3.37%)	the(3.46%)	of(4.33%)
3	and(3.24%)	the(3.09%)	a(2.93%)	and(2.86%)
4	of(2.66%)	I(2.51%)	and(2.41%)	to(2.70%)
5	in(2.42%)	and(2.21%)	of(1.99%)	in(2.30%)
6	a(2.18%)	of(1.98%)	I(1.98%)	a(1.96%)
7	that(1.69%)	in(1.81%)	in(1.76%)	is(1.57%)
8	is(1.63%)	is(1.71%)	is(1.57%)	that(1.22%)
9	I(1.22%)	that(1.37%)	that(1.39%)	as(0.97%)
10	for(1.18%)	for(1.29%)	for(1.32%)	for(0.74%)
11	are(1.02%)	can(1.15%)	be(1.18%)	was(0.74%)
12	people(0.99%)	their(1.01%)	are(1.15%)	this(0.71%)
13	have(0.91%)	it(1.00%)	their(1.10%)	it(0.68%)
14	it(0.86%)	have(0.98%)	can(1.08%)	by(0.67%)
15	not(0.79%)	my(0.97%)	it(0.99%)	be(0.66%)
16	can(0.73%)	be(0.93%)	they(0.97%)	which(0.54)
17	their(0.71%)	they(0.93%)	people(0.91%)	are(0.52)
18	with(0.71%)	are(0.92%)	with(0.85%)	not(0.47)
19	more(0.71%)	not(0.83%)	my(0.80%)	with(0.46)
20	be(0.69%)	with(0.81%)	not(0.80%)	from(0.44)

부정관사 a는 NS에서 2.18%(상위6위), NNS1에서 3.37%(상위 2위),

NNS2에서 2.93%(상위 3위)로 나타난다. 이는 우리나라 학습자들이 원어민과 비교해서 a를 과도사용함을 보여주는 것이다. 반면 정관사 the는 NS에서 4.21%(상위1위), NNS1에서 3.09%(상위 3위), NNS2에서 3.46%(상위2위)로, 우리나라 학습자들이 원어민에 비해 이를 과소사용하고 있음을 알 수 있다.<sup>7)</sup> 이러한 결과를, 우리는 우리나라 학습자들이 격식체보다는 평상체에 더 가까운 글쓰기를 하고 있음을 시사한다고 해석하고자 한다.<sup>8)</sup>

우리나라 학습자들의 이러한 경향은, Ringbom(1998), Granger & Rayson(1998)의 연구에서 나타난 유럽 영어학습자들의 경향과 매우 유사하다. Ringbom(1998)은 ICLE에 속한 서유럽 7개국의 상급 영어학습자의 말뭉치를 서로 비교하고, 그 결과를 또 원어민의 논쟁적 에세이 말뭉치인 LOCNESS<sup>9)</sup>와 비교한 연구에서 7개국의 학습자 모두가 원어민에 비해 a를 과도사용하고, the를 과소사용한다고 보고한다. 또한 Granger & Rayson(1998)는, 불어 ICLE와 LOCNESS를 품사별로 비교했을 때, 전자에서 부정관사 a의 과도사용과 정관사 the의 과소사용이 나타난다고 보고하고 있다.

### 3.2.2. 전치사 및 부정사적 to

<sup>7)</sup>원어민 대학생들의 에세이인 ICE-GB와 비교해 보아도 NNS1과 NNS2에서는 a가 과도사용되고 the가 과소사용되고 있다.

<sup>8)</sup>이러한 해석에서 고려할 점은 학습자 언어의 발달단계에서 나타나는 a, the 사용의 오류이다. 실제 텍스트를 살펴보자.

- (i) a. **(The)** second reason , **(the)** purpose of earning money, After **(the)** economic crisis in 1990's, for **(the)** student's future, **(a)** more important thing, **(a)** house is the basic requirement for **the** well being
- b. If you have **a** same program to watch,
- c. I wholeheartedly disagree with **the** idea.

학습자들은 (ia)의 괄호 속에서의 같이 a, the를 생략한다든가, 밑줄 친 부분과 같이 첨가하기도 하고, (ib)의 경우처럼, a, the를 혼동하여 사용하는가 하면, (ic)의 경우처럼 that 대신에 the를 사용하는 등 다양한 오류를 보이고 있어 위의 해석에 신중함이 요구된다고 할 수 있다. 즉 a의 과도사용과 the의 과소사용이 학습자들의 오류로 인한 것인지, 격식체 글쓰기의 기준에서 벗어나기 때문에 나타나는 특징인지 고려할 필요가 있다고 하겠다.

<sup>9)</sup>LOCNESS는 총 324304 단어의 영국과 미국 학생들의 에세이로 이루어진 말뭉치이다.

위의 표 1에서 NS와 NNS1, NNS2의 전치사를 비교해보면, 그 순위에서는 to, of, in, for, with 순으로 NS과 NNS1·NNS2에서 별 차이가 없음을 알 수 있다. 그러나, 그 백분율에서는 후자에서 of와 in이 다소 과소사용되고 있음을 알 수 있다. NS에서는 to(3.47%, 상위 2)<sup>10</sup>, of(2.66%, 상위 4), in(2.42%, 상위 5), for(1.18%, 상위 10), with(0.71%, 상위 18) 순이고, NNS1에서는 to(3.83%, 상위 1), of(1.98%, 상위 6), in(1.81%, 상위 7), for(1.29%, 상위 10), with(0.81%, 상위 20), NNS2에서는 to(3.76%, 상위 1), of(1.99%, 상위 5), in(1.76%, 상위 7) for(1.32%, 상위 10), with(0.85%, 상위 18) 순이다.

원어민의 전치사의 사용경향은 어떠한가? Brown 말뭉치에서 정보적 산문(A-J)과 상상적 산문(K-R)을 비교해보면 흥미로운 사실을 알 수 있다. 전자의 경우 of(3.39%, 상위 2), to(2.19%, 상위 4), in(1.92%, 상위 5)순이고, 후자의 경우는 to(2.15%, 상위 3), of(1.92%, 상위 5), in(1.35%, 상위 8)순이다. of의 경우 정보적 산문에 비해 상상적 산문에서 상당히 적게 사용되고 있으며, in의 경우도 상상적 산문에서 그 사용이 적다. 이는 비격식체 글에서보다 격식체 글에서 of와 in의 사용이 더 많다고 할 수 있다.

이와 같은 사실을 살펴 볼 때, 우리나라 학생들이, 아주 뚜렷하지는 않지만, of와 in을 대체적으로 과소사용한다는 사실은 그들이 대체적으로 평상체와 구어체에 더 가까운 글쓰기를 한다는 해석과 적어도 양립한다고 할 수 있겠다.

다음으로, to의 경우 NS, NNS1, NNS2 세 말뭉치에서 ICE-GB와 Brown 말뭉치에서보다 빈도수가 높게 나타나고 있어, 우리는 to의 용례를 NS와 NNS2에서 구체적으로 살펴보았다. NS에는 515개, NNS2에서는 538개의 용례가 있었으며, 각 말뭉치의 to의 용례 중에서 두드러진 차이를 보이는 것은 to부정사의 to이었다. 그 중 want\*<sup>11</sup> to, try\*/trie\* to, like\* to, need\* to의 쓰임에서 그 차이가 심하게 나타났다.

<sup>10</sup>이는 전치사의 to와 부정사의 to를 모두 포괄하는 수치이다.

<sup>11</sup>want\* to의 \*가 의미하는 것은 동사의 변형인 wants, wanted, wanting을 포함하는 것으로, 용례에 모두 포함되어 나온다. 마찬가지로 try\*/trie\* to는 try, trying, tries, tried를 포함하며, lik\* to는 like, likes, liked, liking을 포함한다. 이 경우 함께 찾아지는 likely는 수동으로 제외하였다.



우리나라 학습자들은 want\* to와 try\*/trie\* to를 지나치게 과도사용한다. want\* to는 NS에서 9회이지만 NNS2에서는 35회로 나타난다. try\*/trie\* to는 NS에서 2회, NNS2에서 11회로 나타난다. 반면 lik\* to와 need\* to는 과소사용하는데, 전자는 NS에서 14회, NNS2에서 2회이며, 후자는 NS에서 15회, NNS2에서 7회로 나타나고 있다.

Brown 말뭉치에서 정보적 산문과 상상적 산문을 비교하면, want\* to와 try\*/trie\* to의 사용이 후자에서 훨씬 많이 나타남을 알 수 있다. 표 3을 살펴보자. 정보적 산문의 어휘수가 상상적 산문에 비해 3배정도 많지만 want\* to와 try\*/trie\* to의 용례의 수는 거의 비슷하여, 격식체 글에서는 이러한 표현이 적게 쓰임을 알 수 있다.

이와 같은 사실에 근거하여, 우리는 우리나라 학습자들의 want\* to, try\*/trie\* to의 과도사용 경향 또한 평상체에 더 가까운 글쓰기를 한다고 해석할 수 있겠다.<sup>12)</sup>

표 3. Brown 말뭉치에 나타난 글의 종류별 to의 분포

Brown Corpus	정보적 산문	상상적 산문
어휘수	900,759	297,002
to의 용례	16,000	6,373
want* to의 용례	133	153
try*/trie* to의 용례	173	144
lik* to의 용례	74	42
need* to의 용례	89	20

그러나 문제가 그렇게 간단하지는 않다. 표 3의 Brown 말뭉치의 정보적 산문과 상상적 산문에서 lik\* to와 need\* to의 빈도수를 보면, lik\* to는 후자에서, need\* to는 전자에서 좀 더 많이 나타난다. 즉, 원어민은 비격식체 글에서는 lik\* to를 격식체 글에서는 need\* to를 더 자주 사용한다는 것이다. 우리나라 학습자들이 비격식체 글쓰기를 하고 있을 것이라는 지금까지의 가정은 그들이 need\* to를

<sup>12)</sup> 이는 학습자들이 친숙하고 안전한 어휘(Lexical teddy bears)에 더욱 의존하는 경향이 있다는 Hasselgren(1994)의 “테디 곰인형 원리(the teddy bear principle)”와도 관계된다고 할 수 있다. 그는 노르웨이 상급 영어 학습자들이 어려운 단어를 사용하지 않고 익숙한 단어를 사용한다고 밝힌 바 있다.

과소사용한다는 사실과는 양립하지만, lik\* to를 과소사용하는 경향과는 상반되는 것이다.

왜 우리나라 영어학습자들이 lik\* to를 과소사용하는 것일까? 두 가지 원인을 생각해 볼 수 있을 법하다. 우선, 입력의 불균형 때문일 수가 있겠다. 제6차 교육과정 영어교과서 1종을 살펴보면 본문에서 want\* to는 31회, try\*/trie\* to는 19회가 등장하지만, lik\* to는 전혀 등장하지 않는다.<sup>13)</sup> 또 다른, 더 큰 가능성은 모국어의 영향 때문이라는 것이다. like에 상응하는 우리나라 말의 “좋아하다”는 [\_\_NP]로 잘 쓰이나 [\_\_IP]로는 자연스럽게 쓰이지 않는 듯하기 때문이다.<sup>14)</sup>

이와 같이 특수한 사정을 고려하여 lik\* to를 제외한다면, 부정사적 to의 분포도 우리나라 학습자들이 비격식체에 더 가까운 글쓰기를 함을 보여준다고 해석할 수 있겠다.

### 3.2.3. 인칭대명사

인칭대명사의 사용이 글의 성격에 따라 차이가 있음은 잘 알려져 있다. 예를 들어, Biber(1988, 1995)는 1·2인칭대명사는 학문적인 글 등의 정보적 언어 산출(Informational Production)에서보다는 대화나 사적인 편지 등 개인의 감정을 강조하는 참여적 영어 산출(Involved Production)에서 더 많이 나타난다고 관찰한 바 있다(Poole & Field, 1976; Chafe, 1982; Chafe & Danielewics, 1987 참고).<sup>15)</sup>

여러 기존의 말뭉치를 기반으로 원어민의 경향을 분석해보아도, 1, 2인칭 대명사의 사용이 비격식체 글에서 훨씬 많음을 알 수 있다. Brown 말뭉치의 정보적 산문(A-J)과 ICE-GB의 논쟁적 에세이와 같은 격식체 글에서는 I, you의 사용이 상당히 낮다. 특히 LOCNESS같은 논쟁적인 글에서는 글의 객관성을 유지하기 위해 I,

<sup>13</sup>그러나, 다른 교과서에서도 꼭 이럴 것이라는 보장이 없으므로 좀더 세심한 조사가 필요하다 하겠다. 이는 나중 연구로 미룬다.

<sup>14</sup>이와 관련하여, 취미를 진술하는 글에서 우리나라 영어학습자들은 “My hobby is...”를, 원어민들은 “I like/love to VP” 혹은 “I like/love V-ing ...”를 주로 사용한다는 Ahn & Lee(2003)의 보고는 흥미롭다.

<sup>15</sup>Biber(1995)를 알게 해 주신 김용진 교수(숭실대)께 깊이 감사를 드리는 바이다.

you의 사용이 낮고 3인칭 *it, they*의 사용이 두드러진다(Granger & Rayson 1998).

반면에 Brown(K-R)과 LLC와 같은 비격식체 글이나 구어에서는 *I, you*의 사용이 상당히 많아진다. 그리고 다소 일반적인 주제에 대한 개인의 동의 여부나 선호를 밝히는 TOEFL 에세이에서도 *I, we, you*의 사용이 훨씬 높게 나타난다.

즉, 학문적, 논쟁적인 글에서는 개인의 참여가 객관성을 저하시키기 때문에 1, 2인칭대명사의 사용이 제한되지만, 비격식체 글이나 구어 영어에서는 그 사용의 빈도가 두드러진다는 것이다.

이와 같은 사실을 염두에 두고, 우리는 NS, NNS1·NNS2에 나타난 인칭대명사의 빈도를 조사하였는데, 그 결과는 표 4와 같다.

표 4. 인칭대명사의 빈도수<sup>16)</sup>

인칭 대명사	NS	NNS1	NNS2	ICE-GB
<i>I</i>	1.22%(9)	2.51%(4)	1.98%(6)	0.18%(53)
<i>it</i>	0.86%(14)	1.00%(13)	0.99%(15)	0.68%(13)
<i>they</i>	0.96%(22)	0.93%(17)	0.97%(16)	0.30%(35)
<i>we</i>	0.54%(25)	0.72%(23)	0.65%(22)	0.10%(117)
<i>he</i>	0.38%(38)	0.25%(79)	0.17%(108)	0.30%(34)
<i>you</i>	0.16%(100)	0.27%(70)	0.30%(57)	0.02%(500이하)
<i>she</i>	0.07%(233)	0.14%(130)	0.15%(114)	0.12%(98)

먼저 NS과 NNS1·NNS2를 살펴보면, 7개의 인칭대명사의 순위가 거의 비슷함을 알 수 있다. 대체로 *I, it, they, we, he, you, she* 순서로 나타나고, 각각의 말뭉치에서 *I*가 상위 1위를 차지한다. 그 원인을 추측해보면, TOEFL 에세이 글들이 성격 상 개인적인 선호를 서술하거나 동의여부를 밝히는 글들이기 많기 때문일 것이다. 이 경우에는 *I*의 사용이 다른 글에서보다 많아질 수밖에 없다고 본다. TOEFL 에세이의 과제는 Do you agree or disagree--? Do you support or oppose--? In your opinion, which is better? 와 같은 형식으로 제시되고 있다.

이런 유사점에도 불구하고 우리나라 학습자들은 원어민에 비해 *I*

<sup>16)</sup>이 표에서 -( )는 빈도수 순위이다.

를 과도사용하고 있다. NNS1에서는 전체에서 2.51%로 상위 4위이며, NNS2에서도 1.91%로 상위 6위이다. 하지만 NS에서는 1.23%로 상위9로 나타나 다소 큰 차이를 보이고 있다. 2인칭 대명사 you도, NS에서는 0.16%(상위100위)이지만, NNS1과 NNS2에서는 각각 0.27%(상위66)와 0.30%(상위42)로, 우리나라 학습자들이 다소 과도 사용한다고 하겠다. 반면 he의 경우 우리나라 학습자들은 NS와 ICE-GB에 비해 과소사용하고 있어, 인칭대명사의 경우 3인칭인 he 보다는 I, you 같은 좀더 개인적인 표현을 더 많이 사용하고 있음을 알 수 있다.

이와 같은 사실에 비추어 볼 때, 우리는 인칭대명사의 사용 면에서도 우리나라 학습자들의 글쓰기가 원어민에 비해 더 개인적이고 비격식적이며 구어체의 모습을 띄고 있다고 결론을 내릴 수 있겠다. 이는 프랑스 영어학습자들에 대한 Granger & Rayson(1998)의 보고와도 일맥상통한다. 그들은 ICLE 프랑스 학습자영어 말뭉치에서 1, 2인칭 대명사(I, you, we)의 과도사용이 두드러짐을 보이면서, 그들의 글이 구어체의 특징을 지님을 지적한 바 있다.

#### 3.2.4. 본동사

비교 대상이 되는 NS와 NNS1·NNS2의 고빈도 어휘 목록에서 본동사를 선택하여 그 순위를 살펴보면 표 5와 같다.<sup>17)</sup> NS와 비교해서 우리나라 학습자들의 가장 주목할 만한 특징은 think의 과도사용이다. NS에서는 17회로 0.11%, NNS1에서는 380회로 0.40%, 그리고 NNS2에서는 52회로 0.36%를 차지하고 있다.

그리고 NNS1에서 상위 1위를 차지하는 buy가 에세이 과제의 직접적인 영향을 받은 것으로 보고 고려대상에서 제외시켰을 때, 우리나라 학습자들이 want, get 역시 과도사용한다고 볼 수 있다. want의 경우 NS에서 11회로 0.07%, NNS1에서 349회로 0.37%, NNS2에서 35회로 0.24%로 나타난다. get의 경우 NS에서 19회로 0.13%, NNS1에서 222회로 0.24%, NNS2에서 36회로 0.25%로 나타난다. 즉, 우리나라 학습자는 원어민에 비해 특히 think와 want, get 등을 과도사용 한다고 하겠다.

---

<sup>17)</sup>like의 경우 모든 말뭉치에서 빈도수가 높았지만 동사가 아닌 다른 의미로도 쓰이기 때문에 여기에서는 제외하기로 한다.

표 5. 구축된 말뭉치에 나타난 본동사의 순위

순위	NS	NNS1	NNS2
1	work	buy	think
2	make	think	make
3	study	make	learn
4	learn	want	work
5	give	work	buy
6	need	learn	get
7	go	get	want
8	believe	study	go
9	get	choose	know
10	think	go	spend
11	say	take	take
12	find	give	choose

이와 관련하여, think, know, want, feel 등의 私的 동사(private verbs)가 비격식적인 참여적 영어 산출(Involved Production)에서 자주 사용된다는 Biber(1995)의 보고는 주목할 만한 것이다.

원어민의 get의 사용경향을 파악하기 위해 Brown 말뭉치에서 본동사 부분만을 살펴보았다. 가장 주목할 만한 특징은 정보적 산문(A-J)에 비해 상상적 산문(K-R)에서 get의 사용이 상당히 높다는 것이다. 후자에서는 said 다음으로 get이 많이 사용되었지만(410회, 0.14%), 전자에서는 make, work, use, said와 같은 동사들보다 빈도수가 낮다(332회, 0.04%). 이것은 get이 비격식체 글이나 구어체에서 많이 쓰이는 동사임을 보여준다고 추측할 수 있겠다.<sup>18)</sup>

이와 같은 사실에 근거하여, 우리나라 학습자들의 think, want, get의 과도사용 역시 그들의 비격식적이고 평상체에 가까운 글쓰기 경향을 보여주는 것으로 해석할 수 있겠다.

우리나라 학습자들의 이러한 본동사 사용은 역시 유럽의 영어학습자들의 경향과 아주 유사하다는 사실에 주목할 필요가 있겠다. ICLE를 바탕으로 한 Ringbom(1998)에 따르면, 모국어와 상관없이

<sup>18)</sup>여기에서 LOB와 ACE(Australian Corpus of English)를 기반으로 get-수동태가 영국과 호주 영어의 문어보다는 구어에서 훨씬 더 많이 사용된다고 하는 Collins(1996)의 보고도 주목할 만하다.

유럽 7개국의 영어학습자들은, 본동사 중 think를 가장 많이 사용하는데 그 과도사용의 정도가 가장 심하며, 그 다음으로 get을 과도사용 한다는 것이다.<sup>19)</sup> Granger와 Rayson(1998)의 보고도 이와 유사하다. 프랑스 영어학습자들은 논쟁적 글쓰기에서 state, support, believe, argue 등의 다양하고 구체적인 어휘들을 과도사용 하는 반면, 명확하지 않고 범위가 넓은 think를 지나치게 과도사용 하여 어휘의 다양성 면에서 한계를 드러내고 있다는 것이다.<sup>20)</sup>

우리나라와 유럽의 영어학습자들이 이와 같이 비슷한 경향을 보임은, 이러한 경향이 학습자 영어의 전반적인 특징이라는 Ringbom (1988) 등의 주장을 더욱 공고히 한다고 할 수 밖에 없을 것이다.

Ringbom은, 특히, think의 과도사용이 I think 구의 과도사용과 밀접하게 관련된다고 주장하고 있다. 그에 따르면, 모든 학습자들이 I think 구를 과도사용하며, 스웨덴 학습자의 경우 원어민에 비해 5배 정도로, think의 모든 사용 중에서 50%이상을 차지한다는 것이다.

NS와 NNS2에서 I think의 사용을 보면, NS에서는 17개의 용례 중 I think 구는 9개로 53%를 차지하는데, NNS2에서는 52개의 용례 중 I think 구가 32개로 62%에 이른다. TOEFL 에세이의 성격상 NS에서도 다른 원어민의 글에 비해 I think 구의 사용이 높은 편이어서, think의 빈도수 대비 I think 빈도수의 비율의 측면에서는 NS와 NNS2가 별 차이가 없어 보이는 것이다. 그러나, 두 말뭉치의 I think 구 자체의 빈도수를 비교할 때에는, 우리나라 학습자들이 I think 구도 매우 과도사용하는 것이 명백하고, 이는 think의 과도사용의 한 원인이 되고 있다고 이해할 수 있겠다.

실제로 Brown 말뭉치에서는 전체에서 think가 432회로 0.04%밖에 차지하지 않으며, 그 용례를 구체적으로 살펴보면 I think 구로 쓰인 표현은 약 28%밖에 되지 않는다. 정보적 산문과 상상적 산문

<sup>19)</sup>그는 영어학습자들이 get를 과도사용하는 주원인으로, 문맥에 적절한 다른 동사를 찾지 못하기 때문이라고 설명한다.

<sup>20)</sup>우리나라 학습자들의 think의 과도사용 경향도 동어나 주장 등의 TOEFL 에세이 성격에 맞는 구체적이고 정확한 동사를 사용하는 능력의 부족을 보여 준다고 볼 수 있다. 이는 진술한 Hasselgren(1994)의 테디 곰 인형 원리와의도 관계된다고 추측하며 학습자 언어의 발달단계에서 나타나는 공통적인 특징으로 볼 수 있겠다.

을 비교하면, 전자에서 약 29%, 후자에서 약 28%로 거의 똑같이 사용되었다. 이것은 원어민이 학습자들에 비해 think를 상당히 적게 사용하며, 그들이 I think 구에 그다지 의존적이지 않음을 여실히 보여 주는 것이라 하겠다.

### 3.2.5. 축약

Biber(1995)에 의하면, 축약형은 그 의미가 모호할 수 있기 때문에 맥락 속에서 추론되어야 한다. 따라서 대면 대화나 사적인 편지 등 비격식적이고 담화자 참여적인 영어에서 자주 나타나며, 학문적 글이나, 연설 등 정보적 영어에는 자주 쓰이지 않는다고 한다.

NS와 NNS1·NNS2에 나타난 축약형의 빈도를 살펴보자. 그 결과는 표 6에 기술되어 있다.

표 6. 축약형의 빈도수

Word	NS	NNS1	NNS2	ICE-GB
don't	6(0.04%)	102(0.11%)	14(0.10%)	0
doesn't	1	25(0.03%)	2(0.01%)	0
didn't	0	16(0.02%)	4(0.03%)	5(0.02%)
can't	0	46(0.05%)	13(0.09%)	2
won't	0	2	0	0
I'd	0	13(0.01%)	4(0.03%)	0
I'll	0	16(0.02%)	0	0
I'm	0	18(0.02%)	5(0.03%)	0

표 6에서 NS와 ICE-GB의 결과를 살펴보면, 원어민은 격식체 글쓰기에서 축약형을 거의 사용하지 않고 있다. 하지만 NNS1과 NNS2를 보면 우리나라 학습자들이 거의 모든 형태의 축약형을 TOEFL 에세이라는 격식체 글쓰기에서 사용하고 있으며, 특히 don't와 can't를 가장 많이 사용하고 있음을 알 수 있다.

Brown 말뭉치 예에서도 원어민이 정보적 산문(A-J)보다는 상상적 산문(K-R)에서 축약형을 훨씬 많이 사용함을 보여준다. 우리나라 학습자들이 가장 많이 사용한 don't의 경우, 정보적 산문에서 0.02%, 상상적 산문에서 0.11%로 상당히 많은 차이가 난다. didn't의 경우

도, 정보적 산문에서 73회로 0.01%에도 이르지 못하지만, 상상적 산문에서 324회로 0.11%에 이르고 있다. 다른 형태의 축약형도 전자보다는 후자에서 골고루 많이 사용되고 있다.

따라서 우리나라 학습자들의 축약형의 과도사용도 그들의 평상체에 더 가까운 글쓰기 경향을 보여준다고 할 수 밖에 없겠다.

## 4. 요약, 결론 및 교육적 시사점

### 4.1. 요약 및 결론

이 연구는 Granger(1998)의 연구방법론에 근거하여, TOEFL 에세이라는 격식체 글쓰기에 가까운 말뭉치를 통해, 어휘 빈도수와 용례를 분석하여, 우리나라 중, 상급 영어 학습자 영어의 전반적인 특징을 밝히기 위한 것이었다.

이 연구에서 밝혀진 우리나라 학습자들의 경향은 다음과 같다. 첫째, 부정관사 a를 과도사용하고 정관사 the를 과소사용한다. 둘째, 다소간 전치사 of, in을 과소사용하는 경향이 있다. 셋째, 인칭대명사에서 I, you를 과도사용하고 he를 과소사용한다. 넷째, 본동사 think, get, want 등을 과도사용이다. 다섯째, 원어민들이 격식체 글에서 거의 사용하지 않는 축약형을 상당히 과도사용한다.

우리는, 원어민 말뭉치 분석과 사적 동사(private verbs), 축약, 1-2인칭의 사용 등이 “담화참여자 개입적인”(involved) 글의 특징이 된다는 Biber(1995, p.142) 등의 선행 연구에 비추어, 위 사실을 우리나라 학습자들이 격식체 글쓰기에서 평상체, 구어체 글쓰기 경향을 지님을 보여주는 것으로 해석하였다.

이는 Ringbom(1998)과 Granger & Rayson(1988)에서 보고된 유럽학습자들의 경향과 유사하다는 점에서 아주 흥미롭다고 하겠다. 즉, 외국어 습득 단계적인 특징의 하나로서, 모국어와 무관하게 영어학습자들은 논쟁적이고 설명적인 글에서조차 비격식-평상체 글쓰기에서 격식체 글쓰기로 발전해 간다는 것이다.

### 4.2. 교육적 시사점

이 연구를 통해 우리는 우리나라 중·상급 학습자영어의 비격식체 글쓰기 경향을 파악할 수 있었다. 이러한 특징은 영어교육에 대하여 시사하는 바가 크다고 생각한다.

첫째, 글쓰기 교육에서 글의 형식에 적합한 어휘나 문장을 선별,



구성하고 글의 짜임새까지 소화하도록 지도하는 것이 절실히 필요하다 하겠다. 이 연구에서 밝혀진 바와 같이, 우리나라 학습자들은 학습자 개인의 연습과 전략에도 불구하고, 격식체 글쓰기 장르에 적합한 문체와 어휘를 구사하는데 많은 한계를 드러낸다. 좀더 능숙한 영어사용자를 길러내기 위해, 영어학습자들에게 가급적 넓은 범위의 어휘에 접할 수 있게 해 주고 장르별 글쓰기에 대한 좀더 체계적인 훈련을 제공할 필요가 있다고 하겠다.

둘째, 다양한 영역에서 한국인 학습자 말뭉치를 구축하여, 학습자 영어의 경향과 특징, 우리나라 학습자들의 특징이나 오류를 파악할 필요가 있다고 하겠다. 이를 교육에 활용한다면 영어학습에 훨씬 도움이 될 것이다. 이 연구를 통해 TOEFL 에세이라는 격식체에 가까운 장르에서 학습자의 경향을 다소 파악하였으며, 격식체 글쓰기에서 지녀야 할 문체와 어휘 선택에서 유의해야 할 점을 얻게 되었다고 본다.

#### 참고문헌

- Ahn, Sung-Ho and Eun-yeon Lee. 2003. A corpus-based analysis of Korean learners' e-mail English. Presented at the 2003 KASELL International Conference.
- Biber, D. 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, D. 1995. *Dimensions of Register Variation: A Cross-linguistic Comparison*. Cambridge: Cambridge University Press.
- Chafe, W. E. 1982. Integration and involvement in speaking, writing, and oral literature. In D. Tannen, ed., *Spoken and Written Language: Exploring Orality and Literacy*, 35-53. Norwood: Ablex.
- Chafe, W. L. and J. Danielewics. 1987. Properties of spoken and written language. In R. Horowitz and S. J. Samuels, eds., *Comprehending Oral and Written Language*, 83-113. San Diego: Academic Press.
- Collins, P. C. 1996. *Get-passives in English*. *World Englishes* 15, 43-56.
- Collins COBUILD English Dictionary. 1987. Collins Cobuild Publishers.
- Curme, G. 1947. *English Grammar*. New York: Harper and Row.
- Hasselgren, A. 1994. Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics* 4, 237-58.
- Granger, S. 1998. *Learner English on Computer*. Edinburgh: Longman.
- Granger, S. 1998b. *Learner English on Computer*. Edinburgh: Longman.
- Granger, S. and P. Rayson. 1998. Automatic profiling of learner texts. In S. Granger, *Learner English on Computer*. Edinburgh: Longman.

- Jespersen, O. 1909-49. *A Modern English Grammar on Historical Principles*. Copenhagen: Munksgaard.
- Johansson, S. 1985. Word frequency and text type: Some observations based on the LOB corpus of British texts. *Computers and the Humanities* 19, 23-36.
- Meyer, C. F. 2002. *English Corpus Linguistics*. Cambridge University Press.
- Poole, M. and T. Field. 1976. A comparison of oral and written code elaboration. *Language and Speech* 19, 305-12.
- Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik. 1972. *A Grammar of Contemporary English*. London: Longman.
- Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Ringbom, H. 1998. Vocabulary frequencies in advanced learner English: A cross-linguistic approach. In S. Granger, *Learner English on Computer*. Edinburgh: Longman.
- Simpson, J. and E. Weiner. 1989. *The Oxford English Dictionary*, 2nd. Oxford U. P.

안성호

서울특별시 성동구 행당동 17  
 한양대학교 영어교육과  
 우편번호: 133-791  
 전화번호: 02) 2290-1145  
 E-mail: shahn@hanyang.ac.kr

이영미

서울특별시 성동구 행당동 17  
 한양대학교 영어교육과  
 우편번호: 133-791  
 전화번호: 02) 2290-1145  
 E-mail: simple425@hanmir.com

접수일자: 2004. 1. 7.

게재결정: 2004. 2. 15.

**부록 1: NNS의 13개의 과제<sup>21)</sup>****1. 동의여부(20)**

- 1) Employees of businesses should be hired for their whole lives when they get jobs. Do you think that policy is justified, or not?
- 2) Parents make the best teachers. Do you agree or disagree with that statement?
- 3) Teachers' pay should depend on how successfully their students learn. Do you agree with that view, or not?

**2. 선호서술(10)**

- 1) Imagine that someone has given you some money as a gift. With the money, you could either buy an attractive piece of jewelry or go to a concert. Which do you think is a better use of the money?
- 2) Imagine that a friend of yours has been given a sum of money and plans to spend it all either to buy an automobile, or to take a vacation. If your friend asked you for advice, what would you say? Compare these two options and explain which one you think is better for your friend.

**3. 설명(8)**

- 1) The twentieth century saw great change. In your opinion, what is one change that should be remembered about this century?
- 2) On the job, we work with various kinds of people. These people are our co-workers. What do you think are good qualities for a co-worker to have?

**4. 주장(7)**

- 1) Imagine that you have received some land to use as you wish. How would you use this land?
- 2) Imagine you had enough money to buy either a business or a home. Which would you prefer to buy?

---

<sup>21)</sup>부록에서는 각 유형별로 2-3개의 과제만 제시한다. 괄호 안의 수치는 유형별 텍스트의 수의 합계를 나타낸다.

## 부록 2: NS의 22개의 과제

## 1. 동의여부(22)

- 1) Do you agree or disagree with the following statement?  
Parents are the best teachers.
- 2) Do you agree or disagree with the following statement?  
Television has destroyed communication among friends and family.
- 3) Do you agree or disagree with the following statement?  
Games are as important for adults as they are for children.

## 2. 선호서술(4)

- 1) It has been said, "Not everything that is learned is contained in books." Compare and contrast knowledge gained from experience with knowledge gained from books. In your opinion, which source is more important? Why?
- 2) A company has announced that it wishes to build a large factory near your community. Discuss the advantages and disadvantages of this new influence on your community. Do you support or oppose the factory? Explain your position.

## 3. 설명(5)

- 1) People attend college or university for many different reasons (for example, new experiences, career preparation, increased knowledge). Why do you think people attend college or university?
- 2) Nowadays, food has become easier to prepare. Has this change improved the way people live?

## 4. 주장(3)

- 1) If you could change one important thing about your hometown, what would you change?
- 2) Imagine that you have received some land to use as you wish. How would you use this land?