

문형 사전을 위한 문형 빈도 조사*

Studying the frequencies of sentence pattern for a sentence patterns dictionary

김 유 미**
(Yu-Mi Kim)

요약 이 논문은 한국어 교육에서 문형 전자 사전을 바탕으로 하는 자동 문형 검사기를 설계하기 위해 문형의 출현 빈도와 사용 빈도 조사를 목적으로 하였다. 먼저 한국어 교육에서의 문형의 개념을 정의하고 그 유형을 구문 문형과 표현 문형으로 나누어 분류하였다. 서술어 중심의 구문 문형과 의존명사, 어미, 조사가 중심인 표현 문형이 학습자 코퍼스에서 어떻게 나타나는지 분석하였다. 학습자 코퍼스는 학습자들이 꼭 배워야 하는 것으로 표준 코퍼스와 학습자들의 생산물인 오류 코퍼스로 나누어 구축하였다. 한국어 교재로 구성된 표준 코퍼스에서의 문형 출현 빈도와 학습자들이 직접 작성한 글을 모은 오류 코퍼스에서 어떻게 문형이 사용되고 있는지 사용 빈도를 조사하였다. 학습자들의 문형 사용 빈도순은 문형 전자 사전에 기술되고, 이것은 문형 검색 속도를 최적화할 것이다.

주제어 한국어 문형, 문형 검사, 학습자 코퍼스, 문형 빈도, 격틀, 문형 전자 사전

Abstract The purpose of this paper is to examine the frequency and usage of sentence patterns appearing in electronic dictionaries used in Korean language education in order to design an automatic sentence patterns checking. First, the concept of sentence patterns is defined and it is classified into sentence structure patterns and sentential expression patterns. Sentence structure patterns and sentential expression patterns are analyzed how they are expressed in the Korean Learner's Corpus. Learner's Corpus is built into the Standard Corpus, which all Korean Learners must learn, and the Errors Corpus made by learners. From these research, we will find out how frequently the Sentential Patterns are being used in the Standard Corpus which has been made of Korean Texts and how the Sentential Pattern are being used in the Errors Corpus which were constructed from Korean Learner's writings. Finally, having described the Sentential Patterns on the Sentential Electric Dictionary, we determine the optimum speed in the search for the Sentential Pattern.

Keywords Korean language sentence pattern, sentence pattern checker, learners' corpus, sentence pattern frequency, case frame, electronic dictionary of sentence patterns

* 이 논문은 2003년 한국학술진흥재단의 지원에 의해 연구되었음(KRE-2003-908-A00038).

** 경희대학교 외국어대학 한국어학과, 연구세부분야: 한국어 교육, 언어 정보, 사전학
경기도 용인시 기흥읍 서천1리, TEL : 016-245-5985, E-mail : wangdon@hanmail.net

서 론

영어에 기본 문형이 있는 것처럼 한국어에 기본 문형이 정형화되어 있다면 자연 언어처리 기술이 지금보다는 더 급진적으로 발전할 수 있을 것이다. 한국어는 어순이 자유로워 다른 언어보다 언어 분석시 기본 문형을 추출하는데 어려움이 있다. 그러나 문형 정보는 기계 번역, 정보 검색, 문법 검사기와 같은 자연언어 처리 분야의 응용 시스템의 성능 향상을 위한 중요한 정보로 사용될 수 있다.

본 논문에서는 언어 형식에 대한 범주화(categorization)에 관심을 갖고, 한국어 교육에서의 문형의 개념을 정의하고, 이 문형의 공통 속성(attributes)들을 밝히어 한국어 교육에서의 문형 범주를 명시화하려고 한다. 범주에 대한 원형(prototype) 모형설은 사람들이 자신의 환경 내에서 범주화하는 방식에 관한 경험적 증거에 대한 반응으로 1970년대에 출현하였다¹⁾. 여기서 '문형 범주'는 어휘적 범주와 통사적 범주를 포괄하는 용어로 사용된다. 즉 의미적 속성과 통사적 속성에 의해 한국어 교육에서의 문형 범주화이다.

기존의 문형이라고 하면 서술어와 문장 성분을 중심으로 설정하고 있다. 다른 부류와 변별되는 요소가 서술어가 중심이 되어 그에 의존적인 요소들이 결합되는 것이다. 즉 서술어 중심의 문형은 서술어에 해당하는 어휘들의 격틀(case frame) 정보를 말하는 것이다. 또 다른 정의로 문장성분 중심의 문형은 '주어+서술어', '주어+보어+서술어', '주어+목적어+

서술어' 형식을 가리킨다. 그러나 엄밀히 말하면 둘다 서술어가 중심이 되는 것이고 표현 방식에 있어서 차이가 난다고 할 수 있다.

본 연구에서는 문형을 코퍼스에 바탕을 두고 한국어 교육 관점에서 문형을 정의하고 범주화 하였다. 한국어 교육에서의 문형은 학습자 코퍼스를 대상으로 그 출현 빈도와 사용 빈도를 조사하였다. 문형 사용 빈도를 가지고 한국어 교육용 문형 검사기를 위한 문형 사전에 설계하는데 이용하고자 한다. 문형 사전에 문형 사용 빈도에 따라 문형의 유형을 고, 중, 저빈도의 순으로 구성하여 문형 검사시 그 검색 속도를 최적화 하는데 목적이 있다.

문형 사전은 또 종이 사전으로 가공하여 한국어 교육용 학습사전으로 활용 가능하며, 사전 뿐 아니라 다른 교재 편찬에도 응용할 수 있다. 학습자들의 문형 사용 빈도는 한국어 문형의 등급(난이도) 설정의 기초 자료가 될 수 있다. 또 문형 검사기는 웹기반 학습 시스템에 연결시켜 한국어를 학습하는데 있어 학습 정보(문법, 문형)로 도움을 줄 뿐만 아니라, 한국어 교육 문법 검사기에도 응용할 수 있다.

다음 장에서는 이 문형 정보를 이용한 관련 연구들을 살펴보고, 3장에서는 문형 추출의 대상이 되는 학습자 코퍼스를 표준 코퍼스와 오류 코퍼스를 나누어 구축한다. 4장에서는 문형을 정의하고 그 유형을 분류하여 문형 범주를 설정한다. 5장에서 학습자 코퍼스에서 문형의 출현 빈도와 사용 빈도를 알아본다. 마지막으로 문형 검사기 시스템과 문형 사전에 대한 논의로 마무리를 하겠다.

1) 이러한 연구에서 중요한 지표는 가정용 용기의 범주화에 대한 Berlin과 Kay의 발견 및 Rosch와 그녀의 동료들의 연구였다.

선행 연구

문형 정보를 이용한 연구로는, 먼저 기본 문형 설정이 있다. ‘주어+서술어’, ‘주어+보어+서술어’, ‘주어+목적어+서술어’를 기본으로 하여 이에 수식어의 첨가와 서술어의 종류(체언+이다, 동사+형용사)에 따라 여러 형식으로 나누었다. 대표적으로 [13]에서는 일반문형을 구조 문형에 따라 기초 문형, 보태문형, 상황 문형으로 나누고 있다. [9]는 SV, SCV, SOV, SOOV, SOCV로 나누고 있다. 북한 학자인 [2]는 일반문형, 특수문형, 결합문형으로 크게 나누고, 일반문형은 다시 구조와 기능문형으로 나누고 있다. 구조문형이란 문장의 구조적 바탕을 이루며 우리말 문장들의 구조적 특성들을 전형적으로 재현하고 있는 문형을 말하고, 기능문형은 문장의 표현적 기능, 말하는 사람의 표현적 의도나 목적에 따르는 문장을 말한다. [1]은 기본 문형에서 변형의 형태로 확대되는 방식에 따라 문형을 분류했는데, 동사술어 기본 문형 10, 타동사술어 기본 문형 9, 양면동사술어 기본 문형 9, 형용사술어 기본 문형 8, 명사술어 기본 문형 5형으로 구분하고 있다. 이 외 개별 용언 중심으로 논의되고 있는데, 동사·형용사 전체적인 특성에서부터 개별 어휘들의 통어, 형태, 의미적 특성에 관한 연구들이다.

한국어 교육에서 어휘 빈도에 관한 연구는 많이 있지만 문형 빈도에 관한 연구는 전무하다. 유일하게 [8]의 연구가 있는데 총 문장수가 3,767개로 문어가 2,488개, 구어가 1,279개를 대상을 문형의 빈도를 조사하였다. 그러나 문형의 중심이 되는 술어 어휘수가 얼마인지 제시되어 있지 않고, 문형 유형의 빈도만 나

와 있다. 그리고 어떤 방법으로 빈도 조사하였는지 언급이 되어 있지 않다. 문형의 유형이 된 자료도 구체적으로 어떤 자료인지 제시되어 있지 않고, 신문 기사, 소설, 수필, 교과서, 만화, 드라마 대본, 대화 전사로 자료의 유형만 나와 있어 대표성이 있는 자료였는지 알 수 없어 신뢰성이 떨어진다고 할 수 있다.

연구 대상과 방법

연구 대상

이 연구는 실자료인 코퍼스를 대상으로 이루어졌다. 학습자 코퍼스는 특정한 분야나 목적에 맞도록 고안된 특수 코퍼스로 특정 언어를 학습하는 학습자들이 생산한 언어 자료를 수집한 것으로 학습자들의 쓰기, 말하기 등의 자료를 수집하여 구축하였다. 이렇게 구축된 학습자 코퍼스는 학습자들의 오류가 포함되어 있어 학습자 오류 코퍼스(learner's error corpus)라고도 한다. 학습자 오류 코퍼스는 특별한 분석 방법을 사용해 오류를 분석하고 그 결과를 가지고 언어간 대조연구, 학습자들의 학습 상태, 학습 변인, 중간 언어(interlanguage), 오류 원인 등을 연구해 왔다.

여기서는 학습자들의 문형을 학습자 코퍼스에서 추출하기 위해 먼저 학습자 코퍼스를 표준 코퍼스(standard corpus)와 오류 코퍼스(error corpus)로 구분하였다.

표준 코퍼스는 한국어를 학습하는 학습자들이 꼭 배워야 하는 것으로, 대표적인 예가 한국어 교재를 들 수 있다. 연구자가 구축한 표준 코퍼스는 국내 한국어 교육 기관에서 사용

하고 있는 한국어 교재를 대상으로 삼았다. 한국어 교재 중에 통합 교재의 본문, 즉 중심 대화 부분을 대상으로 하였다. 그리고 학습자 코퍼스 구축에 문어와 구어의 자료를 모두 포함시키기 위해 한국어 교재 중에 ‘한국어 회화’, ‘한국어 듣기’, ‘한국어 읽기’ 교재도 포함시켰다. 구어라고 하면 ‘말하기’와 ‘듣기’ 자료, 문어는 ‘쓰기’, ‘읽기’인데 실제 구분되어 출판되는 것은 그다지 많지 않았다. 통합교재 안에 대부분 본문이 말하기 즉 대화 형식으로 나와 있고, 읽기 텍스트도 따로 섹션으로 나와 있다. 그래서 통합교재의 본문 부분과 읽기 부분도 모두 포함시켰다. 표준 코퍼스에 선정된 교재는 다음과 같다.

(1) 한국어 학습자 코퍼스 - 표준 코퍼스

ㄱ. 통합교재

경희대학교 한국어 초급,중급,고급- I/II(총6권)

고려대학교 한국어 1~6.(총6권)

서강대학교 한국어 1~4.(총4권)

서울대학교 한국어 1~4.(총4권)

선문대학교 한국어 초급,중급,고급- I/II(총6권)

연세대학교 한국어 1~6(총6권)

이화여대 말이 트이는 한국어- I/II/III/IV.(총4권)

ㄴ. 기능별 교재

고려대학교 한국어 회화 1~6.(총6권)

이화여대 exiting 한국어 듣기-초급(총1권)

연세대학교 한국어 읽기 1~5.(총5권)

총 43권의 교재를 바탕으로 표준 코퍼스를 구축하였다²⁾. 표준 코퍼스의 크기는 약 120,052

2) 여기에 포함된 교재는 시중에 판매되는 교재를 대상으로 한다. 포함되지 않은 교재는 국외에서 출판된 교재, 국내 한국어 교육 기관에서 사용되

어절이다.

오류 코퍼스는 한국어 학습자들의 결과물을 말하는데 문어로는 쓰기, 구어로는 말한 내용을 녹취한 것이다. 이것은 학습자들의 오류가 그대로 반영된 자료로, 많은 언어 교육 연구자들이 사용하고 있다. 이 연구에 사용된 오류 코퍼스는 국립국어원에서 구축한 자료로 2002년 봄 학기 연세대학교 한국어를 수강하는 학습자가 주이고, 연세대 외에 이화여대, 경희대, 한국외국어대, 서강대, 고려대의 자료도 일부 포함되었다. 학습자들은 모두 1급에서 6급까지 분포하고, 성별은 여자 395명, 남자 250명이고 언어권은 영어권 22.2%, 일어권 41.6%, 중국어권 18.1%, 러시아권 13.8% 그 외 4.3%를 차지한다. 교포가 23.6%, 외국인이 76.4%를 차지하고 있다. 오류 코퍼스의 원시 코퍼스(raw corpus)와 주석 코퍼스(tagged corpus)로 구분되며 통계적 검증을 위해 각 변인별 조건에 맞는 자료를 추출한 결과 최종 통계적 검증에 이용된 자료는 총 645명이고 약 103,771 어절이다.

연구 방법

기존의 언어 교육분야에서 어휘 빈도 연구는 많이 있다. 어휘 빈도는 언어 교육에서 ‘학습자 필수 어휘 선정’, ‘어휘 등급 설정’, ‘어휘의 의미 빈도’가 있고, 이것은 교재, 단어집, 학습 사전에 이용되고 있다. 그러나 한국어 문형 빈도에 관한 연구는 거의 없다. 기본 문형 설정에 관한 연구와 개별 동사의 통사, 의미적 특징 연구는 많으나 동사의 문형에 대한

는 교재 중 내부용으로 사용되는 교재와 사설 학원용 교재들은 포함시키지 않았다.

연구는 코퍼스 구축 후에 조금씩 발표되고 있다. 그러나 그것도 서술어 중심의 용언에 대한 연구가 대부분이다. 한국어 교육에선 문형의 난이도 선정이 무엇보다 필요하다. 그에 기초 연구로 한국어 교육에서 어떤 유형의 문형이 있는지를 조사할 필요가 있다. 이에 본 논문에서는 학습자 코퍼스를 구축하고 어떤 유형의 문형이 나타나는지 조사하였다.

먼저 문형 유형을 뽑기 전에 한국어 교육에서 문형의 개념을 설정하였다. 기존의 문형이란 서술어 중심의 문형만을 지칭하는데 이 논문에서는 서술어 중심의 문형뿐만 아니라 의존명사, 조사, 어미 중심의 문형도 문형의 범주에 포함하였다.

한국어 교육용 문형을 추출하여 그 빈도를 조사하기 위해 먼저 한국어 코퍼스에 해당하는 자료(1)를 모아 전자 파일로 전환하였다. 여기서 해당 텍스트를 스캔하거나 직접 입력하는 방법이 있다. 여기서는 직접 입력하여 코퍼스를 구축하였다. 표준 코퍼스에서 구문 문형에 해당하는 동사·형용사와 표현 문형에 해당하는 의존명사, 조사 결합형, 어미 결합형의 어휘를 모두 추출하고 그 어휘가 들어가는 모든 용례를 ‘글잡이Ⅱ(직접)’³⁾라는 용례 추출 프로그램을 이용하여 뽑았다⁴⁾. ‘외국인을 위한 한국어 학습 사전’을 기준으로 해당 어휘의 용례들의 문형을 수작업으로 분류하였다.

표준 코퍼스에서 추출한 문형의 목록을 오류 코퍼스에서 학습자들이 어떻게 사용하고 있는지 사용 빈도를 조사하였다. 이것도 ‘글잡

이Ⅱ(직접)’를 이용하여 빈도를 계산하였다. 엑셀을 이용하여 학습자들이 사용하고 있는 문형을 고빈도에서 저빈도 순으로 정렬하여 문형 전자 사전에 기술하였다.

학습자 코퍼스에서 문형 빈도

한국어 교육에서 문형 정의

문장에서 술어는 술어를 중심으로 그와 관련된 명사구들과 함께 이루어져 문장 내에서 실현되는데, 동사나 형용사의 통사·의미적 특성에 따라 그에 수반되는 명사구는 일정한 격표지를 가지게 된다. 동사에 따라 일정한 격표지가 실현되는 명사구들과 동사와의 관계를 격틀(case frame)⁵⁾이라고 한다. 동사가 어떠한 명사구를 필수적으로 요구하는지 수의적인지 하는 것에 따라 동사의 격틀(문형)이 달라지고 이 격틀의 유형에 따라 기본 문형이 확정된다고 하였다.

한국어 교육에서의 문형은 서술어가 중심이

5) 격틀(case frame)이란 서술어의 종류에 따라 그 서술 용언에 이끌리어, 반드시 나타나야만 문장이 될 수 있는 최소한의 문장 성분으로 이루어진 구조이다. 동사 ‘가다’의 격틀로는 아래와 같이 제시한다.

㉠ I ①이 ②에로 가다 ①사람명사 ②장소명사 ㉡①이 ②를 -’로도 쓰임

㉢ II ①이 ②를 가다 ①사람명사 ②구경·일·휴가...

㉣ III ①이 가다 ①시간을 나타내는 명사
..... 용언은, 어떤 격틀을 가지고 어떤 논항과 함께 쓰이는지를 알아야, 그것의 정확한 용법과 의미를 파악할 수 있다. <연세 한국어 사전: 문법 용어 풀이>

3) 2000년도에 세종계획 결과물로 만들어진 글잡이 2(색인 탐색 - 용례 추출 프로그램)이다. 색인된 말뭉치를 대상으로 검색을 수행한다.

4) 구문 문형과 표현 문형에 대한 설명은 4.2.를 참조

되어 명사구들과 관계를 가지는 격틀 정보만을 지칭한다. 그러나 한국어 교재에 기술된 것을 살펴보면 다음 (2)에 해당하는 것들을 ‘문법, 문형, 표현, 어휘’란에 설명되고 있다. (2)에 기술한 것은 기존 문형의 개념에는 포함되지 않는 것들이다. 그러나 문형의 개념에서 서술어에 해당되지 않을 뿐 의존명사, 어미, 조사가 중심어가 되어 그것에 의존하는 요소들이 항상 제한적으로 나타나는 패턴들이다. 그래서 본 논문에서는 한국어 교육에서의 문형 범주에 (2)와 같은 것도 포함시키고자 한다. 즉 한국어 교육에서의 문형은 서술어 중심의 문형뿐만 아니라 의존명사, 조사, 어미 중심의 다음과 같은 것들도 문형에 포함되어야 한다.

(2) 한국어 교육에 포함되어야 할 문형

- ㄱ. -(으)ㄴ 편이다, -ㄴ 뿐이다, -는 셈이다
- ㄴ. -(으)ㄴ 수 있다/없다, -는/ㄴ/던/ㄴ 것 같다
- ㄷ. -기 때문에, -는 바람에, -ㄴ/는/ㄴ 때
- ㄹ. -밖에, -만큼

영어에서는 기본 문형 외에 Phrasal verbs, collocations, idiom, phraseology, pattern grammar, grammar pattern 등으로 불려지고 있다. 한국어에서도 동사구, 연어, 관용어, 관용구 등으로 책에 기술되고 있다. 이 논문에서는 이들을 문형이라고 기술하는데, 그 이유는 이들에게서 발견되는 공통적인 속성을 가지고 있다. 즉 한 단어를 중심으로 그 단어에 동반되는 어휘가 있다는 것이다. 문형과 어휘는 상호 도움을 받고 있는데, 그 상호 관계에서 문형은 어휘적 요소의 제한된 틀에서 발생하고, 또 역으로 어휘적 요소도 제한된 문형의 틀에

서 발생한다는 것이다. 게다가 문형은 의미와도 밀접하게 결합되어 있는데, 많은 경우 단어의 의미 차이로 문형의 의미가 다르게 나타난다.

교수법적인 측면에서 이런 것들을 개별 요소들의 결합으로 가르친다면 인지적인 면에서 효율성이 떨어질 것이다. 이것을 설명할 때 전체적인 구조에서 개별 요소들을 하나씩 분리하여 설명해야 하는데, 그렇게 되면 개별 요소들의 의미까지 설명해야 한다. 결합된 조사와 어미가 무슨 의미를 가지는지 이것이 결합된 패턴은 또 어떤 의미를 가지는지 가르쳐야 하는데 그렇게 되면 이 패턴이 가지는 의미를 유추할 수 없을 뿐더러 학습자들이 습득하는데도 방해요소만 될 뿐이다. 심지어 더 많은 오류를 발생하게 될 것이다. 하나의 문법 요소에 여러 가지 의미들이 들어 있는 것으로, 표현할 때 방해 요소가 되는 것이다. 하지만 이것을 별개의 요소가 아닌 하나의 덩어리, 문형으로 가르치면 구조와 의미 설명이 용이할 뿐 아니라 구조적인 분석이 아닌 하나의 패턴으로 인지될 것이다.

다음은 한국어 학습자 코퍼스에서 ‘가다’와 ‘-는 바람에’의 용례 일부이다. ㄱ의 ‘가다’의 용례를 보면 “-이 -에/로 가다”, “-이 -를 가다”가 가장 많이 나타나는 문형으로 한 곳에서 다른 곳으로 움직여서 이동함을 의미한다. (3) ㄴ은 명사 ‘바람’이 꼭 앞에 어미 ‘-는’과 뒤에 조사 ‘-에’를 동반한 형태로 나타나, 예상하지 못한 어떤 일이 일어남을 의미한다. 이 논문에서는 (3)의 ㄱ은 구문문형, ㄴ은 표현문형이라 부른다.

(3) ‘가다’와 ‘-는 바람에’ 용례(concordance)

7. 기숙사에서 사는 친구와 같이 신촌에
 [가서] 한잔했습니
 자주 [가는] 술집에 갔습니다.
 저녁을 먹은 후에 노래방에 [가서] 노래를
 불렀습니다.
 날씨가 아주 좋아서 서울에서 관광을 [가
 고] 싶었는데
 산꼭대기로 올라 [가고] 싶으면 다섯 시간
 정도 걸려요
 타고 싶어서 나는 친구들에게 여의도에 [가
 서] 배를 탔
 그래서 쉬는 날에는 가끔 영화관에 [가서]
 영화를 봅니
 해가 빛나서 우리는 경희궁에 [가기로] 해
 결했어요.
 일합니다만 저기에 우리는 늦게 [가서] 저
 는 친구와 같
 바다에도 사원이 있는데 이 사원 [가고] 싶
 어면 오후
 그리고 환타지아 랜드에 [가서] 즐거운 시
 간을 보내고
 나. 1995년 하반기까지 완전 철거한다고 [밖
 히는 바람에] 박물관은
 감출 장과 바람 풍으로 구성된 말인데 [생
 기는 바람에] 흩어지기
 시간이 별러 [없는 바람에] 다시 시모노새
 키에 가야 했다.
 시모노새키에 배를 탔고 싶는데도 태풍을
 [나는 바람에] 이틀도
 정보화 사회에서 컴퓨터 기수 큰 발전을
 [받는 바람에] 정보화
 [유학하는 바람에] 여행도 항상 해요.
 큰 돈을 [쓰는 바람에] 내일부터 자기가 사
 고싶은 것을 못 살지

그리고 학교나 학원 경쟁 [강하는 바람에]
 아이들은 친한 관계가
 그런데 하고 싶지 않을 때 거절할 수 [없는
 바람에] 좋은 일이 생
 역시 책 너무 [읽는 바람에] 공부못해서 시
 험도 떨어졌어요.

한국어 교육에서 문형은 단어들의 결합이
 비교적 빈번하게 발생하고 그것이 특별한 단
 어 선택에 의존하며 의미가 명확한 것을 문형
 범주화 한다.

(4) 한국어 교육에서 문형의 조건

- ㄱ. 단어의 결합이 빈번하게 출현
- ㄴ. 단어가 특별한 단어 선택에 의존
- ㄷ. 단어들의 결합에 분명한 의미가 존재

한국어 교육에서 문형은 서술어 중심의 문
 형과 의존명사, 조사와 어미 중심의 문형으로
 나눌 수 있다. 서술어 중심의 문형을 구문 문
 형(sentential structure pattern:SSP)이라고 하고,
 의존명사, 조사와 어미 중심의 문형을 표현
 문형(sentential expression pattern:ESP)이라 한다.

문형의 유형

구문 문형

서술어 중심의 구문 문형(sentential structure
 pattern:SSP)은 서술어를 이루고 있는 동사, 형
 용사의 의미에 따라 일정 명사항(NP)을 요구
 하고, 명사항은 일정한 격표지를 수반하는데
 각 용언마다 격틀이 주어질 수 있다.

구문 문형의 유형은 한국어 학습 사전의 격
 틀 정보를 참고했다. 각 용언의 격틀 중에 어

<표 1> 한국어 학습 사전의 격들(구문 격들)⁶⁾

[100] 1이 V	1이 <으>려고 V
1이 2의 V	[500] 1이 2를 V
1에 V	1에서 2가 V
[101] 1이 -어서 V	1이 2에 대해 V
1이 -어도 V	[510] 1이 [2를/-는 것을] V
1이 -느라고 V	[511] 1이 [2를/-기를] V
1이 <으니 V	[512] 1이 [2를/-음을] V
1이 <으면 V	[513] 1이 [2를/-는지를] V
1이 -도록 V	[514] 1이 [2를/-기로] V
[111] [1이/-는 것이] V	[515] 1이 2를/-다고 V
[112] [1이/-음을] V	[516] 1이 -다고 2가 V
[113] [1이/-기가] V	[517] 1이 -다니 V
[120] 1이 2가 V	[520] 1이 2와 3을 V
[121] 1이 [2가/-는 것이] V	(1과 2가 3을 V)
[121a] 1이 [2가/-는자가] V	[521] 1이 2를 3과 V
[121b] 1이 -을까 V	(1이 2와 3을 V)
[122] 1이 [2가/-기가] V	[530] 1이 [2를/2가] V
[123] 1이 2 V	[540] 1이 2를 3을 V
[123a] -은 지 1 V	[550] 1이 2를 3을 V
[130] (1에) 2가 V	1이 2로 3을 V
[131] (1에서) 2가 V	1이 -기로 2를 V
[140] 1이 2와 V(1과 2가 V)	1이 -다고 2를 V
[140a] 1이 2와[부/-게/3같이] V	[560] 1이 <-다고> 2를 V
[141] 1이 2와 V	1이 [2를/-는 것을]<다고> V
[141a] 1이 2 V	[600] 1이 2에 3을 V
[150] 1이 2보다 V	(1이 2를 3에 V)
[160] 1이 [부/-게/3같이] V	1이 2에 대해 3을 V
[170] 1이 2에 3이 V	[601] 1이 [2에/2로] 3을 V
1이 2와 3이 V	[602] 1이 2을 [3에/3과] V
[200] 1이 2에 V	[603] 1이 [2에/-는 것에] 2를 V
[201] 1이 2N까지 V	[610] 1이 2에 [3을/-다고] V
[202] 1이 2와 3에 V(1과 2가 3에 V)	[611] 1이 2에 3으로 V
[210] 1이 [2에/2로] V	[620] 1이 2에서 3을 V
[220] 1이 [2에/-는 것에] V	[700] 1이 2에게 3을 V
[221] 1이 [2에/-는 데에] V	[701] 1이 [2에게/2와] 3을 V
[222] 1이 [2에/-기에] V	[701] 1이 2에게 [3을/3이] V
[223] 1이 [2에/-음에] V	[710] 1이 2에게 [3을/-는 것을] V
[230] 1이 2에서 V	[711] 1이 2에게 [3을/-기를] V
1이 2로부터 V	[712] 1이 2에게 [3을/-음을] V
[231] 1이 [2에/2에서] V	[713] 1이 2에게 [3을/-는지를] V
[232] 1이 [2에서/2로부터] V	[714] 1이 2에게 [3을/-다고/라고] V
[300] 1이 2에게 V	[715] 1이 2에게 [3을/-게] V
	1이 2에게 [3을/-도록] V

[301] 1이 2에게 [부/3처럼/-게] V	[716] 1이 [2를/2가] -게 V
[310] 1이 2에게서 V	1이 [2를/2가] -도록 V
[400] 1이 2로 V	1이 [2를/2가] -이라고 V
(1에게) 2가 3으로 V	[720] 1이 [2에게/2에게서] 3을 V
(1에) 2가 3으로 V	[721] 1이 (2에게/2에게서)[3을/-다고] V
1이 (2에서) 3으로 V	[730] 1이 2에게 (3으로) V
1이 2와 3으로 V	[800] 1이 2를 3으로 V
[410] 1이 2로/이라고 V	[801] 1이 2를 [3으로/3에] V
[420] 1이 2로 V/ -다고 V	[802] 1이 2를 [3으로/3이라고] V
[421] 1이 2로 V/ -는 것으로 V	[803] 1이 2를 [3으로/부/-게] V
[430] 1이 [부/-게] V	[804] 1이 2를 3에서 4로 V
1이 -처럼 V	[810] 1이 2를 3으로 V/이 -는 것으로 V
[431] 1이 -기 위해 V	[820] 1이 2를 -는 것으로 V

편 것이 표준 코퍼스와 오류 코퍼스에서 빈번하게 사용되었는지 분석하였다. 한국어 학습 사전의 격들은 표 1과 같다.

표현 문형

표현 문형(sentential expression pattern:ESP)은 어휘 의미를 가지고 자립적으로 기능을 하던 것이 문법적인 기능도 하고 있는데, 문법적인 범주(의존명사, 조사와 어미)를 명확하게 분류하기 어려운 항목들로 문법 형태소를 포함하고 있는 덩어리(chunk)들이다⁷⁾. 표현 문형은 하나의 단위로 구성되어 있는데 이것은 학습

6) 외국인을 위한 한국어 학습 사전, p761.

7) 문법화 과정에 있는 요소들이라 할 수 있다. 즉 어휘적 기능을 하던 것에서 문법적 기능을 하는 것으로, 또 덜 문법적인 기능을 하던 것이 더 문법적인 기능을 하는 것으로 바뀌는 것을 문법화(文法化의, grammaticization)라 한다. Hopper & Traugott(1993)에서는 ‘문법화’에 대해 “We define grammaticalization as the process whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions, and once grammaticalized, continue to develop new grammatical function.” 으로 정의 내리고 있다.

자들로 하여금 자유롭게 문장을 만들거나 말하는데 있어서 어렵게 하는 요소들이다.

표현 문형은 표준 코퍼스의 자료가 되는 한국어 교재에서 제시하고 있는 것들을 모두 목록화하였다. 그러나 한국어 교재만으로 부족하여 ‘한국어 문형 표현 100’, ‘한국어 문법 사전’, ‘한국어 학습용 조사·어미 사전’⁸⁾에 표제어도 포함시켜 추출하였다.

문법적인 특성에 따라 표현 문형을 분류해 보면 의존명사 결합형, 조사 결합형, 어미 결합형으로 분류된다. 어미 결합형은 다시 전성 어미 결합형, 연결어미 결합형, 종결어미 결합형으로 나누었다. 그 유형은 표 2와 같다.

<표 2> 표현 문형의 유형

의존명사 결합형	것	-는(으)ㄴ/(으)ㄹ 것이다
	격	-는 격이다
	겸	-(으)ㄴ 겸
	길	-는 길에
	김	-는(으)ㄴ 김에
	나름	-기 나름이다
	나위	-(으)ㄴ 나위가 없다
	노릇	-(으)ㄴ 노릇이다
	대로	-는(으)ㄴ/(으)ㄹ 대로
	데	-ㄴ 데(에)
	동안	-는 동안(에)
	등	-는(으)ㄴ 등 -는(으)ㄴ 등
	듯	-는/은/을 듯(이)
	따름	-ㄴ 따름이다
	리	-(으)ㄴ 리가 없다/있다
	마련	-기/게 마련이다
	마당	-는 마당에
만	-ㄴ 만 하다	

	만큼	-ㄴ 만큼	
	망정	-기에 망정이다	
	바	-(으)ㄴ 바에야/-(으)ㄴ 바에는	
	바람	-는 바람에	
	법	-는(으)ㄴ 법이다	
	뻔	-(으)ㄴ 뻔하다	
	뿐	-ㄴ 뿐이다	
	성	-ㄴ 성 싶다	
	셈	-는 셈이다	
	수	-(으)ㄴ 수 있다/없다	
	심상	-기 심상이다	
	양	-ㄴ 양하다	
	일	-ㄴ 일이 있다/없다	
	일쑤	-기가 일쑤이다	
	적	-(으)ㄴ 적이 있다/없다	
	줄	-ㄴ/ㄹ 줄 알다/모르다	
	중	-는 중(에)	
	지	-(으)ㄴ 지	
	지경	-(으)ㄴ 지경(이다)	
	짜	-기가 짝이 없다	
참	-(으)ㄴ 참이다		
채	-(으)ㄴ 채로		
척	-는(으)ㄴ 척하다		
체	-는(으)ㄴ 체하다		
탓	-(으)ㄴ/(으)ㄹ 탓		
터	-(으)ㄴ 터		
통	-는 통에		
판	-ㄴ 판에		
편	-ㄴ 편이다		
조사 결합형	-이/가	-이가 아니다	
	-을/를	-을/를 비롯해서/비롯하여	
	-에	-에 관하여, -에 대하여	
어미 결합형	-(으)로	-(으)로 인해서	
	-과/와	-과/와 비교하다	
전성 어미	전성 어미	-기	-기 전에
	연결 어미	-아/어/여서	-어/아/여서 그런지
		-다(가)	-다(가) 보면
		-았/었/였(다)	-(았)다가 -(었)다가 하다
		-(으)면	-(으)면 (으)르수록
		-거나	-거나 하면
종결 어미	-ㄴ까요?		

8) 백봉자(1999), 외국어로서의 위한 한국어 문법 사전, 연세대출판부.

이윤진(2004), 한국어 문형 표현 100(한국어 학습자와 한국어 교사를 위한), 건국대학교 출판부.

이희자·이종희(2001), 한국어 학습용 조사·어미 사전, 한국문화사.

문형과 빈도 분석

표준 코퍼스에서 문형과 빈도

한국어 교재로 구성된 표준 코퍼스에서 문형과 그 출현 빈도를 조사 하였다. 표준 코퍼스에서 추출한 문형과 여기에 (5)에서 제시한 문형을 첨가하였다. 교재와 사전류에 중복되는 것은 하나로 통일시켰다. 이것은 교재뿐만 아니라 사전류를 포함하여 한국어 교육에서 제시하고 있는 문형을 알아보는 데 의의가 있다.

- (5) ㄱ. 한국어 문형 표현 100
- ㄴ. 한국어 문법 사전
- ㄷ. 한국어 학습용 조사·어미 사전

한국어 문형에서 구문 문형에 포함되는 용언들의 어휘는 총 3,058개이다. 교재와 사전류에 1번 이상은 나타난 단어들이다. 빈도 2이상의 어휘들은 1,318개로 전체 43.1%를 차지한다. 교재에 가장 많이 출현한 어휘는 ‘맞다’로 39번 나타났다. ‘맞다’가 나온 교재는 서/서강/선/선(주사)/연워/이들/경(주사)/고/교회/서/서(조건)/서강/선/선(적성/주사)/선/연/연(입)/경(비/생일/야단/적성)/서(주사)/서강/서강(성격/야단)/선/선(간/비/야단/주사/침)/연워/이/서강(딱/음식간/도둑)이다⁹⁾. 빈도 11이상 나온 어휘들로 72개로 표 3과 같다.

9) 서-서울대, 서강-서강대, 선-선문대, 고-고려대, 연-연세대, 이-이화여대, 경-경희대, 연워-연대워기, 교회-고대회화, 이들-이대듣기를 가리킨다. ‘서’란 서울대에 ‘맞다’의 대표적인 의미, 사전에 뜻풀이 1번으로 제시된 의미로 나타난 것이고, ‘선(주사)’란 논항 정보로 ‘주사를 맞다’의 의미로 교재에 제시되었다는 뜻이다.

<표 3> 한국어 구문 문형에 포함된 용언 어휘 빈도순

어 휘	빈도	어 휘	빈도
맞다	39	났다	12
들다	30	끓이다	12
보다	29	끓다	12
쓰다	28	고르다	12
나다	26	확인하다	11
넣다	18	편하다	11
내다	18	크다	11
나오다	16	치다	11
타다	14	초대하다	11
짓다	14	착하다	11
일어나다	14	지내다	11
쉬다	14	자르다	11
생기다	14	입다	11
보내다	14	울다	11
떠나다	14	운전하다	11
나타나다	14	오르다	11
찾다	13	어렵다	11
부탁하다	13	아프다	11
도착하다	13	싸우다	11
깨끗하다	13	심심하다	11
걸리다	13	신다	11
걸다	13	시키다	11
짜다	12	슬프다	11
전하다	12	세우다	11
있다	12	복잡하다	11
이용하다	12	배우다	11
유명하다	12	멀다	11
아름답다	12	떨어지다	11
씻다	12	돌아가다	11
시원하다	12	덥다	11
시끄럽다	12	달다	11
부르다	12	닭다	11
바꾸다	12	나가다	11
맷다	12	고생하다	11
닭다	12	결혼하다	11
내리다	12	가깝다	11

국립국어원에서 제시한 ‘외국인을 위한 학습 사전(A)’ 용언 어휘 2,303개와 ‘한국어 학습용(B)’의 용언 어휘 1,721개, 연구자가 추출한 ‘구문 문형(C)’ 용언 어휘 3,058개를 비교해 보겠다.

<표 4> 구문 문형 비교

항목	A	B	C
어휘수	2,303	1,721	3,058

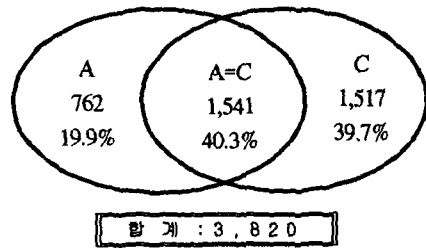
- * A :외국인을 위한 학습 사전 용언 어휘
- B :한국어 학습용 용언 어휘
- C :한국어 교육 구문 문형 어휘

◎ Case 1 : A, C 비교

‘외국인을 위한 학습 사전(A)’과 ‘구문 문형(C)’ 용언 어휘들을 비교해 보면, A에 있는데 C에 없는 용언 어휘, A>C 어휘는 762개이다. 구문 문형은 교재만을 대상으로 만든 코퍼스에서 추출한 것이고, 외국인을 위한 학습 사전은 대량의 코퍼스에서 한국어 어휘를 조사한 것이라 수치 차이가 있다. 역으로 C에는 있는데 A에 없는 용언 어휘들, A<C로 1,517개로 나타났다. 학습 사전의 어휘들은 「기초 어휘 의미 빈도 사전」 개발을 위한 기초단계로 나온 결과물이다. 교빈도의 어휘들을 대상으로 사전을 만들었을텐데 구문 문형에 나오는 많은 단어들어 빠진 것을 알 수 있다. 이것은 사전을 만드는데 기반이 된 코퍼스의 자료 선정에 문제가 있다. 학습자들이 배우는 대표적인 교재들의 단어들은 모두 포함되어야 한다. 그리고 실제 한국어 학습자들이 사용하는 글과 말에서 사용하는 어휘들을 조사 연구하여 그 결과도 학습자 사전(learner's dictionary)에 반영되어야 한다. A와 C에 중복되는 어휘들은

A=C 1,541개로 나타났다.

A		C	
2,303		3,058	
A>C	A=C	A<C	합계
762	1,541	1,517	3,820
19.9 %	40.3 %	39.7 %	100 %



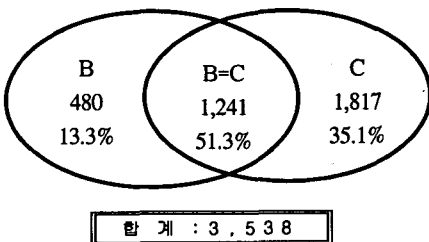
◎ Case 2 : B, C 비교

‘한국어 학습용 용언 어휘(B)¹⁰⁾’와 ‘구문 문형(C)’의 어휘들을 비교해 보면, B는 한국어 학습용 어휘들 중에 용언 부분만 뽑은 것인데 총 1,721개이다. ‘현대 국어 사용 빈도 조사’를 위해 국어 기초 자료로 코퍼스에서 상위빈도를 추출한 것으로 한국어 학습용 어휘를 선정하였다. Case 1의 학습 사전과 같은 코퍼스 자료에서 어휘들을 뽑은 것이다. ‘학습용 용언 어휘(B)’에는 있는데 ‘구문 문형(C)’에는 없는 어휘수(B>C)가 480개이고, 역으로 ‘구문 문형’에는 있는데 ‘학습용 용언 어휘’에 없는 어휘가(B<C) 1,817개이다. B는 학습자들이 알아야 할 필수 어휘들인데 C와 비교할 때 이런 차이가 나타나는 것은 한국어 교재를 포함한

10) 국립국어연구원에서 학습자들이 익혀야 할 한국어 필수 어휘를 선정하였다. 그 중 용언의 어휘만 뽑으면 1, 721개이다.

코퍼스에 최근에 출판된 교재들은 빠져 있다. ‘현대 국어 사용 빈도 조사’의 대상이 된 코퍼스는 1998년~2000년 구축한 자료들인데 이 시기의 한국어 교육 교재만을 대상으로 하였다. 이 시기 출판된 교재와 일본에서 출판된 교재 일부가 포함되어 있는데, 88년 서울 올림픽 이후 한국의 위상이 올라가면서 한국어 교육도 활기를 띠게 되고, 2000년대 이후 한국의 영향으로 많은 외국인들이 한국어에 관심을 가지게 되고 그로인해 많은 한국어 교재들이 출판되었다. 그러나 2000년 이후 출판되고 기존에 출판된 교재들도 수정판들이 나오게 되었는데 이 자료들은 반영되어 있지 않다. 또 2000년 이후 출판된 교재들 중, 대학 기관에서 사용하는 교재 중 시리즈로 출판되는데 대부분 중/고급 교재들이 출간되었는데 이 자료들이 빠져 있다. 그러므로 중/고급의 한국어 교육용 어휘들이 빠져 있고, 그 대신 국어 기초 자료 국어 자료들에서의 고빈도의 단어들이 포함되어 있다.

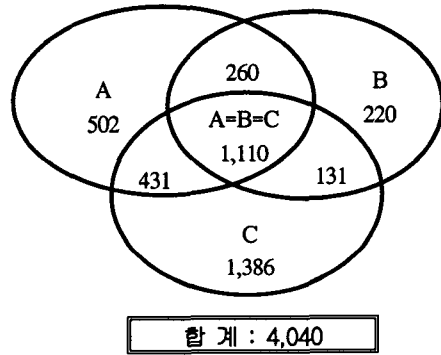
B		C	
1,721		3,058	
B>C	B=C	B<C	B+C
480	1,241	1,817	3,538
13.6 %	35.1 %	51.4 %	100 %



◎ Case 3 : A, B, C 비교.

마지막으로 ‘외국인을 위한 학습 사전(A)’과 ‘한국어 학습용 용언 어휘(B)’, ‘구문 문형(C)’ 용언 어휘에 모두 중복되어 나타나는 어휘의 수는 1,110개로 나타났다. 전체 어휘는 총 4,040개의 어휘수에서 27.5%의 비율을 차지한다.

A	B	C	A=B=C	A=B	B=C	A=C	합계
520	220	1,386	1,110	260	131	431	4,040
12.4%	5.4%	34.3%	27.5%	6.4%	3.2%	10.7%	100%



표준 코퍼스에서 표현 문형은 총 1,268개의 문형으로 가장 많이 출현한 문형은 ‘-아 보다’로 17회 출현했다. 회중, 선(1|3_2), 서강(1_2), 서강(1_2), 교회(1)/서(1), 경(1)/고(1_2), 이윤, 연(1_2), 이튼(1)로 초급 교재에 자주 나오는 문형이다¹¹⁾. 빈도 2이상 나온 문형이 모두 413개이고, 빈도 10이상의 문형은 22개의 문형으로 나타났다.

A, B, C에서 서로 중복되지 않는 어휘는 모두 4,040개이다. 이것은 “한국어 교육에서 필

11) 회중·이희자·이종희(어미·조사 사전), 이윤·이윤진(한국어 문형 표현), 선(1|3_2)란 선문대 1(초급)에 1번, 3(고급)에 2번 나왔다라는 표시.

<표 5> 한국어 표현 문형 고빈도순(표준 코퍼스)

문 형	품사	빈도
-아 보다	P	17
-아 주다	P	14
-고 있다	P	14
-리 수 없다/있다	N	13
-려고 하다	P	12
-리 줄 모르다/알다	N	12
-리 때	P	12
-기로 하다	P	12
-게 되다	P	12
-아야 하다	P	11
-아 놓다	P	11
-로 해서	J	11
-러 가다/오다	P	11
-리 만하다	N	11
-는 바람에	N	11
-니 것 같다	N	11
-고 싶다	P	11
-아도 되다	P	10
-아 보이다	P	10
-니 적이 없다/있다	N	10
-니 김에	N	10
-기 전에	P	10

수 용언 어휘”라고 할 수 있다.

표현 문형의 유형에 따른 빈도를 살펴보면 가장 많이 출현한 빈도의 문형으로 의존 명사 결합형 ‘-리 수 없다/있다’가 13회, 조사 결합형 ‘-로 해서’가 11회, 어미 결합형 ‘-아 보다’가 17회로 나왔다. 빈도 10이상의 표현 문형들을 추출한 것이 표 5이다. 각 항목별로 고빈도

에 속하는 것 중에 빈도 10을 기준으로 했을 때 어미 결합형 문형이 가장 많은 수를 차지하고, 다음이 의존명사 결합형, 조사 결합형 순으로 나타났다. 교재 편찬 시 한 유형의 문형이 편중되지 않게 문형의 수와 난이도를 잘 선정하여 교재에 제시해야겠다.

오류 코퍼스에서 문형과 빈도

표준 코퍼스에 나타난 문형들이 학습자들의 실제 데이터인 오류 코퍼스에서는 어떻게 나오는지 조사하였다. 오류 코퍼스란 실제 학습자들의 글로 구성된 자료이기 때문에 학습자들의 실제 문형 사용 빈도를 알 수 있다. 표준 코퍼스에서 추출한 문형들은 한국어 교재에 나타나는 문형들, 그 빈도라 하면 교재에 많이 나타난 것이다. 교재에 많이 나온 것은 그만큼 중요하고 학습자들이 꼭 배워야 하는 것으로 반복적으로 제시한 것으로 판단할 수 있다. 그러면 이런 것들을 학습자들이 많이 정확하게 사용해야 하는데 정말 그런지를 알아보기 위해 학습자들이 직접 쓴 자료인 오류 코퍼스에서 살펴 보고자 하는 것이다.

먼저 구문 문형이 오류 코퍼스에서 어떤 양상으로 나오는지 검토해 보면, 학습자들이 어떤 구문 문형을 많이 사용하는지 알 수 있다. 구문 문형의 경우 <표 1>의 격틀 유형을 기준으로 분석하였다. 구문 문형의 용언 어휘 3,058개의 용례를 모두 추출하여 수작업으로 문형을 구분하였다. 용례를 추출하는 작업은 용례 추출 프로그램인 ‘글잡이 II(직접)’를 이용하여 정규표현으로 용례를 추출하였다.

오류 코퍼스에서 학습자들이 가장 많이 사용해던 문형은 ‘가다1’의 1번째 의미로 문형 ‘210/500’이다. ‘가다1’의 1번째 의미는 ‘(한 곳

<표 6> 구문 문형 사용 고빈도순(오류 코퍼스)

구문 문형	빈도	의미	품사	문형	논항1	논항2
가다 ¹²⁾	1039	1	V	210/500	유정	장소
위하다	805		V	500		
재미있다	787		A	120	사람	일/표정
아름답다	713	1	A	100		
중요하다	588		A	111		
건강하다	489		V	100	사람,몸	
쉽다	474	1	A	200		
가다 ¹	432	2	V	100		학교/교회/회사
행복하다	428		A	100	사람	
시작하다	424	2	V	500	사람	
결혼하다	409	2	V	140	사람	사람
힘들다	376		V	111/101	사람	일/몸
노력하다	374		V	431	사람	
필요하다	368		A	120		
갖다	366		V	500		
마시다	355	1	V	500	사람	물/술/차
예쁘다	351	1	A	100		
가지다	348	2	V	500	사람	인생관/기대/동 정심
일하다	337	1	V	260	사람	회사/은행
같다	335	1	A	140		
나쁘다	306	1	A	100		

에서 다른 곳으로)움직여서 이동하다'로 문형 '210/500'은 <표 1>에 나와 있듯이 ' [210] 1 이 [2예/2로] V', '[500] 1이 2를 V(1에서 2가 V)'의 문형으로 1,039회 사용되었다¹³⁾.

오류 코퍼스에서 표현 문형의 사용 빈도를 살펴보자. 학습자들은 구문 문형의 용언 어휘들을 사용하여 글을 작성하는 것 보다, 표현 문형을 사용하여 작문을 하는 것이 학습자들

12) 동형어: 형태는 같으나 품사가 서로 다르거나 의미가 서로 다른 낱말들.
가다¹-동사(집에 가다)
가다²-보조동사(먹어 가다)

13) 아직 어휘 145개를 분석하고 있는 중이다, 이 어휘들의 데이터가 빠진 상태에서 빈도 분석한 결과이다. 분석이 완료되면 목록이 달라진다.

에겐 더 어렵다. 하나의 패턴(pattern)처럼 문법적 요소인 조사와 어미, 의존명사가 중심어가 되고 거기에 또 다른 조사와 어미가 결합된 형태로 나타나기 때문이다. 또 결합된 문법적 요소들에 따라서 의미의 차이도 나타나기 때문에 표현 문형이 어렵다. 개별적인 단어로 학습하는 것보다 한 덩어리(패턴)로 학습하는 것이 오류를 범하지 않고 인지하는데 더 효과적인 방법일 것이다¹⁴⁾.

표현 문형 1,268개 중 학습자들이 가장 많이 사용한 문형은 ‘-가 있다/없다’로 총 5,554회 출현한다. 이 문형은 표준 코퍼스의 출현 문형 빈도를 보면, 1회 나온 문형으로 저빈도에 속하는 문형인데, 학습자들의 글인 오류 코퍼스에서는 가장 많이 나온 문형으로 나타났다. 그 외에 고빈도의 표현 문형을 표 8로 제시한다.

의존 명사 결합형은 ‘-ㄴ 것 같다’가 1,230회, 조사 결합형 ‘-가 있다/없다’의 문형이 5,554회, 어미 결합형 ‘-고 있다’가 2,963회로 가장 많이 사용된 문형이다. 표준 코퍼스에서 표현 문형 빈도 표 5과 비교해 보면 어미 결합형 문형이 가장 많이 출현한 것을 볼 수 있는데 표 7에서는 고르게 사용하고 있다. 한국어 교재에는 어미 결합형 문형이 많이 제시되어 있는데 학습자들은 많이 제시되어 있는 어미 결합형보다 다른 표현 문형들도 고르게 사용하고 있다. 그리고 교재에 고빈도로 나타난 표현문형을 많이 사용하지 않고, 중/저빈도에 랭킹되어 있는 문형들을 실제 많이 사용하고 있는 것을 볼 수 있다. 왜 이런 현상이 나타

<표 7> 표현 문형 사용 고빈도순(오류 코퍼스)

표현 문형	품 사	빈 도
-가 있다/없다	J	5,554
-고 있다	P	2,963
-지 않다	P	2,542
-기 때문에	P	2,444
-고 싶다	P	2,334
-가 좋다	J	1,675
-ㄴ 때	P	1,489
-에 있다	J	1,357
-기 위해(서)	P	1,281
-ㄴ 것 같다	N	1,230
-ㄴ 것	N	1,204
-ㄴ 수 없다/있다	N	1,129
-는 것이	N	971
-에 대해서	J	914
-을 하다	J	901
-가 되다	J	886
-ㄴ 것이다	N	831
-에 가다	J	820
-게 되다	P	763
-고 하다	P	660
-에 대해(서)	J	641
-을 보다	J	600
-ㄴ 일이다	N	559
-기 전에	P	551
-에 대한	J	542
-과 같이	J	542
-가 아니다	J	524
-도 -지만	J	485
-을 먹다	J	440

14) 개별적 어휘 교육이 아닌 문형에 대한 교수, 학습 방법은 이 논문의 중심 주제에서 벗어나 더 이상 다루지 않는다.

나는지 보다 세밀한 연구가 필요하다. 교재에 제시된 문형의 난이도 문제인지, 학습자의 학습 인지 능력인지, 교수법의 문제인지 보다 심의 있는 조사가 필요할 것이다.

결론

한국어 교육에서 문형을 범주화하기 위해 먼저 학습자 코퍼스를 한국어 교재인 표준 코퍼스와 학습자들의 직접 쓴 자료인 오류 코퍼스로 구분하여 구축하였다. 코퍼스를 대상으로 한국어 교육에서의 문형 빈도를 살펴 보았다. 문형 범주화를 위해 문형의 개념을 코퍼스에 기반하여 한국어 교육의 관점에서 정의하였다.

한국어 교육에서의 문형의 공통된 속성은 첫째, 단어의 결합이 빈번하게 일어나야 하며 둘째, 단어가 특별한 단어에 의존해야 하며 셋째, 단어들의 결합에 분명한 의미가 있어야 문형의 범주에 포함되었다. 이에 해당되는 문형들을 유형화 했는데, 서술어 중심의 구문문형과 의존명사·조사·의미 중심의 표현문형으로 나누었다. 기존의 구문 문형 외에 표현 문형이 문형의 범주에 포함되어야 하는 이유는 무엇보다 가장 중요한 이유는 학습자들의 문형 인지하고 습득하는 과정에서 구조주의 관점에서의 분석보다는 하나의 패턴 문형으로 인지되고 설명되는 것이 유리하고 경제적이기 때문이다.

한국어 교육에서 문형의 유형은 구문 문형은 동사·형용사로 구분하고, 표현 문형은 의존명사 결합형·조사 결합형·어미 결합형으로 다시 분류했다.

(6) 한국어 교육에서의 문형 유형

- ㄱ. 구문 문형
 - 동사
 - 형용사
- ㄴ. 표현 문형
 - 의존명사 결합형
 - 조사 결합형
 - 어미 결합형
 - 전성어미
 - 연결어미
 - 종결어미

구문 문형과 표현 문형을 학습자 코퍼스에 서의 출현 빈도와 사용 빈도를 각각 추출하여 비교해 보았다. 여기서 조사한 문형은 출판된 한국어 교재를 대상으로 구축한 표준 코퍼스에서 나오는 문형들이다. 즉 교과서에 나오는 문형만은 모두 포함시켰다.

문형 정보는 기계번역(machine translation)의 한국어 생성, 정보 검색(information retrieval)시 주요 정보 추출과 구문 분석, 문법 검사기 등 자연언어 처리 분야에서 질 높은 결과를 만들기 위해 필요하다. 여기서 문형 빈도 자료는 한국어 교육에서의 문형 검사기 개발에 사용할 문형 사전 설계에 이용한다. 웹기반으로 설계될 한국어 문형 검사기는 학습자들의 한국어 학습시 문형 정보를 제공하고, 텍스트의 문형 검사를 함으로써 문형 학습에 도움을 주 고자 하였다.¹⁵⁾

참고문헌

강은국 (1993), 조선어 문형 연구, 박이정.
 고동혁 (1994), 조선어문형개론, 한국문화사.

15) 해당 문형의 목록과 빈도는 김유미(2005)를 참조.

- 고석주 외(2004), 한국어 학습자 말뭉치와 오류 분석, 연세국학총서 38, 한국문화사.
- 김유미 (2000), 학습자 말뭉치를 이용한 한국어 학습자 오류 분석 연구, 연세대학교 교육대학원 외국어로서의 한국어교육학과 석사 논문.
- _____ (2005), 한국어 교육에서 문형 검사기를 위한 문형 추출, 경희대학교 정보통신대학원 멀티미디어시스템공학과 박사 학위 논문.
- 남현숙, 권혁주, 김수남, 유승훈, 신민규, 권혁철 (1999), “웹에 기반한 국어 작문 학습 시스템 구현”, 제 11회 한글 및 한국어 정보처리 학술대회 논문집
- 남현숙, 김상훈, 김지원, 권현주, 정유진, 권혁철 (2000), “한국어 철자/문법 검사기와 웹 기반 언어 학습 시스템의 통합 환경 구축”, 한국인지과학회 춘계 학술대회.
- 노은희 (1999), 한국어 교재 개발을 위한 한국어의 문형 빈도 조사, 국어교육연구, 서울대학교 국어교육연구소
- 박영순 (1993), 현대 한국어 통사론, 집문당.
- 안주호 (1997), 한국어 명사의 문법화 현상 연구, 한국문화사.
- 이상섭 (1988), 옥스포드 영어 사전의 편찬 원칙과 형성 과정, 『사전편찬학연구』 제8집, 한국문화사.
- _____ (1995), 「말뭉치: 그 개념과 구현」, 『사전편찬학연구』 5·6집, 한국문화사.
- 이익섭·남기심 (1986), 국어 문법론 I, 방송통신대학 출판부.
- 이현형, 황이규, 이용석 (2000), “문형과 단문 분할을 이용한 한국어 구문 모호성 해결”, 한국인지과학회 학술대회지.
- 임지룡·김동환 (1998), 인지언어학 개론, 태학사.
- 서상규·한영균 (1999), 국어정보학 입문, 태학사.
- 송영빈, 채영숙, 최기선 (1999), “동사의 애매성 해소를 위한 구문의미사전 구축”, 한글 및 한국어 정보처리 학술대회, 전주, 1999년 10월 8-9일.
- 최용석, 이주호, 최기선 (1999), “격틀 자동구축과 격틀평가 방법에 관한 연구”, 제 11회 한글 및 한국어 정보처리 학술대회 논문집.
- 홍윤표 외 (2002), 한국어와 정보화, 태학사.
- Paul J. Hopper and Elizabeth Closs Traugott (1993), Grammaticalization, Cambridge University Press.
- Susan Hunston & Gill Francis (2000), Pattern Grammar, John Benjamins Publishing Company.
- 한국어 교과서
경희대학교 한국어 초급,중급,고급-I/II, 국제교육원 한국어 교육부, 경희대 출판부.
- 고려대학교 한국어 1~6, 한국어 문화 연구부, 고려대학교 민족문화연구소.
- 고려대학교 한국어 회화 1~6, 한국어 문화 연구부, 고려대학교 민족문화연구소.
- 서강대학교 한국어 1~4, 서강대학교 한국학센터, 하우.
- 서울대학교 한국어 1~4, 서울대학교 어학연구소, 한림출판사.
- 선문대학교 한국어 초급,중급,고급-I/II, 선문대학교 한국어교육원, 생각하는 백성.
- 연세대학교 한국어 1~6, 연세대학교 어학당, 연세대학교 출판부.

연세대학교 연세대학교 어학당, 연세대학교
출판부.
처음 배우는 한국어 읽기 1
재미있는 한국어 일기 2.
즐거운 한국어 읽기 3
생각하는 한국어 읽기 4
주제가 있는 한국어 읽기 5
이화여대 말이 트이는 한국어- I/II/III/IV, 이
화여자대학교 언어교육원, 이화여자대학
교 출판부.
이화여대 exiting 한국어 듣기-초급, 이화여자대
학교 출판부.

사전류
백봉자, 외국어로서의 위한 한국어 문법 사전,
연세대출판부, 1999.
이운진, 한국어 문형 표현 100(한국어 학습자
와 한국어 교사를 위한), 건국대학교 출
판부, 2004.
이희자·이종희, 한국어 학습용 조사·어미
사전, 한국문화사, 2001
외국인을 위한 한국어 학습 사전, 문화관광부
한국어세계화 재단, 2004.

1 차원고접수: 2005. 4. 14

2 차원고접수: 2005. 5. 25

최종게재승인: 2005. 6. 10