

DELE 코퍼스 구축 및 등급별 스페인어 기본 어휘 선정: B1, B2를 중심으로*

신 자 영

단독/연세대학교

Shin, Jayoung (2011), Compilación del corpus de DELE y la selección del vocabulario fundamental para el nivel B1 y B2.

Abstract El objetivo de este estudio consiste en presentar la metodología de seleccionar el vocabulario fundamental para el nivel B1 y B2. Para este objetivo, hemos compilado los modelos de exámenes del nivel B1 y B2 de DELE (Diploma de Español como Lengua Extranjera) desde el año 2005 hasta el 2009 y hemos creado un corpus en formato de textos electrónicos. En la sección 2 revisamos el concepto de vocabulario básico y el vocabulario fundamental y abordamos las medidas aplicadas para la selección del vocabulario fundamental y su cuantía de acuerdo al nivel del estudiante de la lengua extranjera. En la sección 3 proponemos el método técnico para crear un corpus que consta de guardar el corpus convertido en archivo de texto y luego, por medio del programa de concordancia, extraer la lista y la frecuencia de todas las palabras utilizadas en el nivel B de DELE. Posterior a ello, marcamos los lemas y las categorías gramaticales para seleccionar solamente las categorías léxicas como sustantivo, adjetivo, adverbio y verbo y así escogemos la lista y su frecuencia. La lista de vocabulario selecta en esta primera fase en base al corpus es compensada tras comparar con la lista de vocabulario de alta frecuencia de otros corpus de mayor dimensión como el Corpus del Español y CUMBRE y considerar el criterio subjetivo del autor antes del listado final del vocabulario fundamental del español de nivel medio. De acuerdo a los resultados de la presente investigación, el vocabulario básico de un estudiante de nivel B1 es de unas 4.200 palabras mientras que el vocabulario fundamental del nivel B es de unas 7.300 palabras. El presente estudio es significativo al proponer una teoría específica para seleccionar el vocabulario fundamental del español basado en estadística. Asimismo, los

* 본 논문은 2010-2011 스페인 문화부 문화 교류 협력 프로그램에 의하여 지원받았음.

resultados obtenidos a través del corpus pueden ser reutilizados para la edición de diccionarios y textos didácticos de español de nivel medio y en la enseñanza del léxico al mismo tiempo que pueden ser aprovechados para estudios lingüísticos y aportar en la educación del español.

Key words enseñanza del léxico en español, DELE nivel B, compilación del corpus, vocabulario fundamental, frecuencia léxica
 스페인어 어휘교육, 델레 중급, 코퍼스 구축, 기본어휘 선정, 어휘 빈도

I. 연구의 목적

본 연구의 목적은 스페인어 공인 자격 시험인 DELE(Diploma de Español como Lengua Extranjera) 중급(Nivel B1, B2) 기출 문제를 전자 텍스트 형식의 코퍼스로 구축하고, 이를 기반으로 중급 스페인어 학습자를 위한 기본 어휘 목록을 선정하고, 이를 가공하여 스페인어 교육 및 언어 연구에 활용하는 방법을 모색하는데 있다. 최근 스페인어에 대한 수요가 증가하면서 스페인어 중급, 고급 학습자의 수요도 꾸준히 증가하고 있으며, 따라서 DELE 중급, 고급 자격증에 대한 관심도 증가 추세에 있다. 그러나, 국내에서 DELE에 대한 체계적 연구는 매우 부족하며, 이를 준비할 수 있는 교재는 많지 않은 실정이다. 특히, 대부분의 연구는 초급 학습자를 위한 교재 개발이나 기초 어휘 등에 집중되었으며, 중급 및 고급 학습자를 위한 교재 개발이나 어휘 연구는 많지 않은 실정이다. Lewis(2000:8)에 의하면 외국어 교육의 중, 고급 단계의 가장 큰 특징으로 볼 수 있는 점은 급격한 어휘량의 증가이다. 스페인어 기본 문형 및 문법은 초급 단계에서 대부분 다루어지며, 일상적인 의사소통을 위하여 필요한 기초 어휘 교육 또한 초급 단계에서 다루어진다. 따라서, 중급 단계 이후의 외국어 교육은 각각의 단계에 맞는 적절한 어휘가 제시되어야 하고 학습자가 자유롭게 이해하고 표현할 수 있도록 다양한 교수법 및 교재가 개발되어야 할 것이다. 그렇다면, 중급 교육을 위하여 적절한 규모의 어휘를 선별하는 연구가 우선되어야 할 것이다.

본 연구에서는 중급 학습자를 위한 기본 어휘 선정을 위하여 텍스트로서 가장 규범성이 있다고 고려되는 스페인어 공인 자격 시험인 DELE 기출문제를 그 대상으로 삼았다.¹⁾ 우선 2005년부터 2009년까지의 DELE의 중급(B1, B2) 기출 문제 자료를 전자 텍스트 형식의 코퍼스로 구축하고, 이를 가공하여 계량학적 방법을 이용한 DELE B1, B2의 기본 어휘 목록을 1차적으로 선정하고, 이를 기존의 다른 대용량 코퍼스의 고빈도 어휘 목록과 대조하여 보완한 후, 스페인어 교육의 현장 경험이 있는 연구자의 주관적 판정을 거쳐 최종적으로 정제된 스페인어 중급 기본 어휘 목록을 구축하는 것을 목표로 한다. 본 연구는 스페인어 자료 구축 방법론의 구체적 사례 및 방법론을 제시하여 추후 다양한 스페인어 연구에서 자료 구축 방법의 기초를 제시하게 될 것이며, 기본 어휘 선정의 계량적, 객관적 방법론 및 메타 검증식 방법론의 실제 예를 통하여, 보다 객관적 어휘 선정의 방법론을 제시한다는 점에서 의의가 있을 것이다. 또한, 본 연구의 결과로 구축된 코퍼스 자료는 다양한 분야에서 재사용되고, 활용되어 순수 언어학적 연구 뿐 아니라 스페인어 교육의 자료 개발에 기초를 제공하게 될 것이다.

II. 어휘 선정 방법론

1. 기초 어휘(vocabulario básico)와 기본 어휘(vocabulario fundamental)

학습자를 위한 기본 어휘를 선정하기 위하여, 우선 기본 어휘의 개념에 대한 정리가 필요할 것이다. 학습자를 위한 어휘 선정 연구에서 ‘기초 어휘’ 혹은 ‘기본 어휘’와 같은 용어가 많이 쓰여 왔다. 기초어휘(vocabulario básico, vocabulario nuclear)와 기본 어휘(vocabulario fundamental, léxico

1) 언어교육을 위한 유럽공통 참조 지침(Marco Común europeo de referencia para las lenguas) 149장에 의하면 모든 언어의 자격 시험 및 교재는 등급별 어휘 선별의 원칙 하에 작성되어야 함을 명시하고 있다. 어휘 선정의 기준은 기초 어휘를 포함하며, 계량적 방법에 의한 선정, 구어와 문어의 실제 자료에 근거한 선정, 학생의 요구 반영 등을 제시하고 있다.

fundamental)에 대한 개념정의는 연구자에 따라 강조점이 다르게 나타나긴 하지만 대체적으로 기초 어휘는 일상 생활을 영위하고 사용상의 편의를 도모하기 위한 필수적이고 핵심적인 어휘라는 측면에서 의사소통의 기초가 되거나 어휘 생성의 기초가 되는 어휘로 정의 된다. 이에 비해 기본 어휘는 특정한 목적 지향적으로 선정된 어휘라는 개념으로 주로 논의되었다. 즉, 특정한 분야를 위해 선정된 어휘이기 때문에 중급 스페인어 학습자를 위한 기본 어휘라 함은 스페인어 학습자의 학습 목적에 맞는 어휘라 할 수 있다.

임지룡(1991)에서는 일본 국립 국어연구소의 자료를 바탕으로 다음 <표 1>에서와 같이 어휘수와 이해도의 관계를 제시하고 있다. 이 자료에 의하면, 스페인어의 경우 대략 1000개 정도의 어휘로 일상 생활의 81.0%를 이해할 수 있으며, 5000개 어휘로는 일상어의 92.5%를 이해할 수 있다고 하였다.

스페인어의 기초 어휘에 대한 기존의 연구에서도 유사한 어휘량을 제시하고 있는데, Benítez Pérez(1994)에서는 A. Juiland y E. Chand Rodríguez(1964)를 인용하면서, 스페인어 기초어휘 1000개가 일상 생활에서 사용하는 85.5%의 어휘를 포함한다고 하였다. Manuel Santiago(2001)에서도 한 언어의 고빈도 4000개 어휘가 언어 텍스트의 97.5%를 차지한다는 것을 밝힌 바 있다. 김경희(2001)에서는 기본적으로 의사소통에 필요한 스페인어 기초 어휘를

〈표 1〉 언어별 어휘수와 이해도 (임지룡:1991)

어휘수\언어	영어	프랑스어	스페인어	중국어	일본어
1-150				48.0	
1-300				59.7	45.3
1-500				67.1	51.5
1-1000	80.5	83.5	81.0	76.5	60.5
1-1500				79.0	
1-2000	86.6	89.4	86.6		70.0
1-3000	90.0	92.8	89.5		75.3
1-3500					75.3
1-4000	92.2	94.7	91.3		
1-5000	93.5	96.0	92.5		81.7

2000개 내외로 보았으며, 성인학습자를 대상으로 하는 대부분의 외국어 학습 교재는 1500-2000개의 어휘를 포함하고 있음을 밝혔다.²⁾ 이러한 사실을 근거로 초급 스페인어 어휘 교육은 이러한 2000개 내외의 기초 어휘를 중심으로 이루어져야 한다는 데에 대하여는 대부분의 학자들이 의견의 일치를 보이고 있으며, 스페인어 초급 어휘 목록을 제시한 다양한 연구 결과물들을 찾을 수 있다.

본 연구에서 선정하고자 하는 어휘의 목록은 중급 수준의 스페인어 학습이라는 명백한 목표를 가진다. 기초 어휘의 규모가 대략 2000개 내외인 반면에 본 연구에서 목표로 하는 중급 스페인어 학습 어휘는 기초 어휘에 비교하여 훨씬 많은 량의 어휘를 포함하게 될 것이다. 따라서 본 연구에서는 ‘기초 어휘’라는 용어 대신 중급 스페인어 학습자를 위한 ‘기본 어휘’라는 용어를 사용하도록 하겠다.

2. 등급별 학습 어휘 규모

중급 수준의 스페인어 학습자를 위한 어휘를 선정하기 위해서는 두 가지 중요한 방법론적 문제를 제기할 수 있다. 우선 중급 스페인어 학습자를 위한 적절한 어휘 선정 규모가 무엇인가? 하는 문제와, 이러한 어휘 선정을 위하여 어떠한 방법을 취할 것인가 하는 문제이다. 우선, 김광해(2003)가 제시한 외국어로서 한국어 교육에 필요한 등급별 어휘량에 대하여 살펴보면 <표 2>와 같다. 김광해(2003)에서는 모국어 학습자의 발달 단계에 따른 어휘규모를 바탕으로 외국어로서의 한국어 학습자에 맞는 어휘 규모를 제시하고 있다. <표 2>에서 제시한 바와 같이, 외국어로서 한국어 학습자의 중급 단계에서 필요한 어휘는 대략 모국어 화자의 1등급 어휘의 수와 일치하는 1845개이며, 중급 수준에서는 모국어 화자의 2등급 어휘 규모와 일치하는 대략 6090개의 어

2) 김경희(2001:6)에서는 Español Directo 1A, 1B에는 1295개의 어휘가 수록되어 있으며, Berlitz에서 개발한 “Basic Spanish Workbook”에서는 약 1125개의 어휘가 수록되어 있음을 밝혔다.

〈표 2〉 등급별 습득 어휘수 (김광해:2003)

어휘량	누계	국어교육용		한국어 교육용		
		등급	개념	4구분	6구분	개념
1845	1845	1	기초어휘	초급	1	자국인과 어휘량을 일치시키는 방향으로 조절함
4245	6090	2	정규교육이전	중급	2	
8538	14448	3	정규교육개시 -사춘기 이전	상급	3 4	
19377	33825	4	사춘기 이후	고급	5 6	
32946	66751	5	전문화된 지적 성장 단계			
45569	112340	6	저빈도어: 대학 이상, 전문어			
125670	238010	7	누락어: 분야별 전문어, 기존 계량 자료 누락어휘			

휘를 제시하고 있다.

스페인어의 중급 학습자를 위한 기본 어휘 규모의 적절성을 판정하기 결정하기 위해서는 한국어 등 다른 언어에서의 외국어 교육의 중급 어휘의 규모를 참고로 할 수 있을 것이다. 그러나, 본 연구에서는 어휘를 규모를 정한 후 연역적으로 빈도수, 분포 등과 같은 계량학적 방법으로 어휘를 선별하는 방법보다는, 중급 스페인어 학습자를 위한 텍스트에 출현한 어휘를 출발점으로 하여 귀납적으로 어휘를 선정하고 그 규모를 조절하는 방법을 취하기로 한다.

3. 기본 어휘 선정 방법

중급 스페인어 학습자를 위한 등급별 교육용 기본 어휘 선정 방법은 첫째, 말뭉치 등의 어휘 빈도 정보를 비롯한 각종 계량적 정보를 적극 활용하여 선정하는 객관적 방법이 있고³⁾ 둘째, 말뭉치의 어휘 빈도정보를 토대로 하여 한국어 교육 전문가의 주관적 평정을 활용하는 절충적 방법,⁴⁾ 그리고 기존 어휘 목록을 비교하면서 어휘 일부를 보충하여 선정하는 메타 계량 방법⁵⁾으

3) 서상규(2009)

4) 조남호(2002)

5) 김광해(2003)

로 정리해 볼 수 있다.⁶⁾

객관적 방법으로는 어휘 출현 빈도(frecuencia), 분산도(dispersión)등이 기준이 될 수 있으며, 장르, 사용역별 기초 어휘 선정 및 주관적 판정이 어휘 목록의 정제 작업에 보완적 역할을 할 수 있다. 본 연구에서는 일차적으로 DELE 기출 어휘 목록을 선정하여, 빈도 1이상의 모든 어휘를 기반으로 1차 목록을 작성하고, DELE 코퍼스 이외의 대규모 스페인어 균형 코퍼스의 어휘 빈도를 통하여 보완하고, 주관적 판정에 의하여 정제 작업을 하는 절충적 방식을 취할 것이다. 그러나 무엇보다도 DELE 코퍼스를 기반으로 한 어휘 목록 확보가 중급 어휘 선정의 출발점이 될 수 있을 것이다.

III. DELE 코퍼스 구축 및 어휘 선정 방법론

스페인어에 대한 자료 기반 연구를 위해서는 공신력 있는 기관에서 구축한 대용량 코퍼스에 의존하는 방법과 연구자의 연구 목적에 따라 자신이 연구하고자 하는 코퍼스를 직접 구축하는 두 가지 방법이 있을 수가 있다. 후자의 경우 스페인어의 가장 수요가 많고 활용도가 높은 자료를 탐색하는 것이 필요할 것이다. DELE는 스페인의 공인된 기관에서 실시되는 전 세계적으로 실시되는 공신력있는 시험이며, 텍스트로서 규범성과 활용성이 매우 크다. 특히 등급별 어휘의 통제가 체계적으로 되어 있어, 등급별 스페인어 어휘 선정에 객관적 자료로 쓰일 수 있다. 본 연구에서는 일차적으로는 스페인어 DELE 코퍼스를 구축하는 방법론을 제시할 것이며, 구축된 코퍼스를 기초로 하여 출현 어휘 목록을 선정하고 이를 가공하고, 분석한 후 중급 스페인어 어휘 최종 목록을 선정하게 된다. 코퍼스 구축 및 어휘 선정의 방법은 여러 단계의 절차를 거쳐 이루어지게 된다. 다음은 본 연구에서 실행한 작업의 단계를 개략적으로 보인 것이다.

6) 서상규 (2009)

- 1) DELE 시험지 목록 및 pdf 화일 확보
- 2) 화일 형식 변환 : pdf 파일을 text 파일로 전환
- 3) 자료 전처리 : 불필요한 space 조절, 기호 삭제. 인코딩 변환(UTF-8), 특수 기호 정리, 오타자 수정.
- 4) 용례색인 프로그램을 이용한 형태빈도 추출 및 어휘(형태소) 목록 확보
- 5) 레마화 작업을 통한 단어 사전 등재 원형 기술
- 6) 등급별 레마 빈도 추출
- 7) 대용량 코퍼스(Corpus de español, CUMBRE) 의 상위 빈도 목록과 의 대조 작업
- 8) 결과 자료의 해석 및 주관적 어휘 정제 작업
- 9) 최종 기본 어휘 목록 확정
- 10) 어휘 가공 및 어휘 교육 방안 마련

1. DELE 코퍼스 구축의 기초 자료 확보

DELE 시험의 주관 기관인 Instituto de Cervantes의 웹페이지(<http://diplomas.cervantes.es/candidatos/modelo.jsp>)에서는 2005년부터 2009년까지의 기출문제를 PDF 형식으로 공개하고 있다. 공개된 기출 문제를 다운받아 저장한다. 본 연구에서 구축할 DELE 시험지 목록과 및 시험의 구성은 <표 3>, <표 4>와 같다. 매회 시험의 모든 영역의 문제를 포함하며, 듣기 시험의 경우 별도로 제공되는 대본을 이용한다. 해답도 포함한다. 총 56개 기출 시험 문제에 대한 코퍼스 구축을 시도하게 된다.

〈표 3〉 년도별 DELE 자료 구성

	2005	2006	2007	2008	2009	총계
Inicial escolar B1		1	1	1	1	4
Inicial B1	3	4	5	4	2	18
Intermedio B2	2	4	5	4	2	17
총계	7	13	16	13	7	56

〈표 4〉 DELE 시험의 구성

Pruebas I y II: 독해 및 작문 (Interpretación y producción de textos escritos)
Pruebas III y IV: 듣기 및 말하기 (Interpretación de textos orales y Conciencia comunicativa)
듣기 대본(Transcripción de la prueba de Interpretación de textos orales)
해답 (Clave de respuestas)

2. DELE 코퍼스 자료의 화일 형식 변환 및 전처리 과정

DELE 기출문제는 모두 pdf 형태로 제공되므로 이를 모두 text 화일 형식으로 변환하여야 코퍼스 구축이 가능하다. 화일 변환 과정에서 스페인어 강세가 있는 모음이나 ñ, ÿ 와 같은 특수 기호 등이 깨지거나 하는 변형이 있을 수 있는데 이를 수작업으로 교정하는 것이 필요하다. 또한, 인코딩을 유니코드나 UTF-8로 변환하여 다시 저장할 필요가 있다. 또한 수작업으로 단순 철자 오류 및 띄어쓰기 오류를 수정하여야 하며, 엔터, 탭, 불필요한 스페이스 등의 기호를 제거하는 전처리 과정이 필요하다. 또한 변환된 텍스트는 원본 pdf 화일에 근거하여 검수하고 교정한다.

텍스트에 기본적인 정보를 표시하는 것을 텍스트 마크업이라고 한다. 본 연구에서는 제목부분과 본문(body) 부분을 구분하여 표시하고 DELE의 경우 등급(B1, B2)(Nivel) 및 영역(section, part) (Interpretación/producción de textos escritos/Comprensión auditiva /Gramática y vocabulario) 시험 날짜(date) 등의 표시를 하였다. 특히, 지시문과 본문의 구분을 하였는데, 지시문의 경우 매 항목마다 반복되는 경우가 많아 이를 본문과 구분하는 표지인 <instruction>를 두지 않으면, 후에 어휘 빈도 추출 시에 혼란의 여지가 있기 때문이다. 다른 코퍼스와의 호환을 염두에 두고 모든 구분자는 영어를 사용하였다. 국제적 표준인 SGML 방식의 인코딩을 한다. 정보 표시의 예를 들면 다음 <표 5>와 같다.

3. 어휘 목록 추출 및 레마화 및 어휘 범주 표시

이렇게 구축된 코퍼스의 규모는 다음 <표 6>과 같다. 코퍼스의 규모의 척도가 되는 것은 출현한 단어의 수인데, 일반적으로 출현형 혹은 출현 항목(token)과 낱말 유형(type)으로 구분한다. 출현형(token)은 텍스트를 구성하고 있는 총 단어로, 모든 유형이 구체적으로 실현되어 코퍼스 상에서 나타난 사례들을 뜻한다. 일반적으로 코퍼스의 규모를 나타낼 때에 사용되며, 코퍼스를 구성하는 총 ‘어절수’ 라고도 한다. 낱말 유형 혹은 형태소 유형(type)은

〈표 5〉 텍스트 마크업의 예

```

<text>
  <body>
    <level> Nivel B1 </nivel>
    <section> INTERPRETACION DE TEXTOS ESCRITOS </section>
    <part> Parte número 1</part>
    <date>12 de mayo de 2006 </date>
    <instruction> Instrucciones A continuación, encontrará usted un texto y tres
preguntas sobre él. Marque la opción correcta en la Hoja de Respuestas
Número 1.</instruction>
    <p> DICCIONARIO PANHISPANICO DE DUDAS
Centenares de hispanohablantes de todo el mundo acuden a diario a las
diferentes Academias de la Lengua Española -22 en total: 19 en países
hispanoamericanos, una en España, otra en Filipinas y la última en
Norteamérica- para exponer sus dudas sobre cuestiones ortográficas,
léxicas o gramaticales y para pedir aclaración sobre ellas. .... 후략 </p>
  </body>
</text>

```

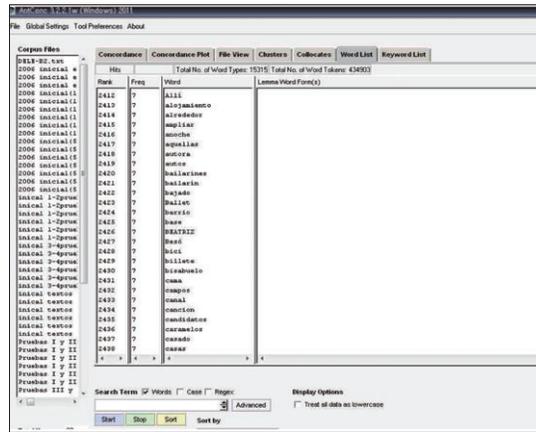
〈표 6〉 DELE B1, B2의 코퍼스 규모⁷⁾

등급	출현형 수(token)	낱말 유형 수(type)
B1	314,362	7,278
B2	120,541	11,453
B1+B2	434,903	15,153

텍스트를 구성하고 있는 단어 중에서 중복이 없이 순수하게 한 번씩만 집계했을 때 보이는 낱말의 종류 혹은 유형으로 단어의 형태가 다르면 다른 유형으로 간주된다. 즉, 한 단어의 다양한 변이 형태 come, comes, comiste 는 각각의 유형으로 취급된다.

코퍼스에서 항목 수(token)를 낱말 유형의 수(type)로 나눈 수치는 코퍼스 내에서 한 형태가 평균적으로 몇 번 반복적으로 쓰이는 가를 볼 수 있다. 즉, B1 등급에서는 총 출현하는 어절 수를 단어/형태의 종류로 나눈 수치는 B1에서는 43.19 로 동일 어휘가 평균 43.19회 반복 출현함을 알 수 있는 것이다. 반면에, B2에서는 10.52 로 어휘가 반복되는 횟수가 B1에 비하여 현저하게

7) 표에서 B1과 B2의 어절 수(token)의 산술적 총합은 B1+B2의 총합이 된다. 그러나, 단어 유형의 수(type)는 B1과 B2의 유형의 수의 총합(18731개)이 B1+B2의 총합 수(15153)가 될 수 없다. 이는 B1과 B2에 공통적으로 출현하는 단어 종류의 수인 3578 개를 뺀 수치이다.



〈그림 1〉 Ant Conc 을 이용한 DELE 어휘 목록 및 빈도 추출

낮은 것을 알 수 있다. 이는 B1에 비하여 B2에 훨씬 다양한 어휘 군이 사용되고 있음을 나타낸다.

구축된 DELE 코퍼스를 파일 별로 저장하고, 콩코던스 프로그램⁸⁾에 로딩한 후 총 출현 어휘목록 추출(word list) 기능을 이용하여, 구축한 DELE 코퍼스에 출현한 총 어휘 목록 및 어휘의 빈도를 추출할 수 있다. <그림 1>에 제시한 그림은 구축된 DELE B1, B2 시험지를 Ant Conc 프로그램에 로딩하여, 출현 어휘의 빈도를 구한 화면의 일부를 보인 것이다. 화면에서 나타난 항목은 빈도의 순위(rank), 출현 회수(frequency), 단어 형태(word)를 각각 나타내며, 7회 출현하면서 빈도의 순위가 2412위에서 2438위에 있는 어휘 목록을 보인 것이다. 어휘 목록 검색의 결과는 text파일로 저장이 가능하며 이를 엑셀로 변환하여 필요한 작업을 실행할 수 있다. 이 단어들은 사전 등재된 원형 혹은 레마(lemma)의 형태가 아닌 형태 통사적 변이형⁹⁾이므로, 이들의 원형을 모두 기술하고 원형(lemma)을 기술하는 후속 작업이 필요하다. 이 작업은 엑셀(ms excel)을 이용하여 수작업으로 진행하였다.

8) 용례색인 프로그램이라고도 한다. 본 연구에서는 Ant Conc 3.2.2.1.w (2011년 버전)을 사용하였다.

〈표 7〉 엑셀을 이용한 레마 및 품사 기술

등급	빈도	형태	레마	품사
B1	436	es	ser	v
B1	298	numero	número	n
B1	239	HOMBRE	hombre	n
B1	237	MUJER	mujer	n
B1	234	ejemplo	ejemplo	n
B1	223	dos	dos	a/n
B1	214	Respuestas	respuesta	n
B1	213	Hoja	hoja	n
B1	202	VOZ	voz	n
B1	199	Numero	número	n
B1	180	Parte	parte	n
B1	169	Texto	texto	n
B1	153	correcta	correcto	a
B1	150	texto	texto	n
B1	150	Dialogo	diálogo	n
B1	148	muy	muy	adv
B1	133	PRUEBA	prueba	n
B1	128	opcion	opción	n
B1	126	mas	más	a/adv

<표 7>은 엑셀 상에 실행한 수작업의 과정을 보여준다. 우선 모든 출현하는 단어의 형태의 레마형과 품사를 기술한다. fue, fui와 등과 같이 중의적 성격이 있는 어휘의 경우 코퍼스에서 각각의 용례를 찾아 'ir'와 'ser'의 정확한 분포를 기술한다. 또한, 대소문자와 같은 폰트를 통일하게 되며, 고유명사의 경우 나라 이름을 제외하고는 모두 명사의 범위에서 제외하였고, 이메일, 웹 주소, 알파벳 자모 등도 품사 기술 목록에서 제외하였다. 모든 단어 형태에 대한 레마 기술이 완료된 후에는, 레마순으로 정렬한 후 엑셀의 함수 기능을 이용하여 동일한 레마의 출현 빈도를 합산하여 B1, B2 등급별 총 출현 레마 목록 및 빈도를 구할 수 있게 된다.

9) 사전 등재형(lemma)은 한 형태의 모든 굴절 변이형을 포함하는 기본형으로, comer, comiste, comes 등의 레마형은 'comer'가 된다.

4. 어휘범주의 레마 빈도 추출

<표 6>에서 보인 어절수 통계는 명사, 형용사, 부사, 동사와 같은 어휘적 범주뿐 아니라, 접속사, 대명사, 전치사, 관사 등과 같은 기능범주를 포함한다. 일반적으로 기본 어휘 및 어휘 빈도의 연구는 주로 어휘범주를 대상으로 한다. 이는 기능 범주의 경우 언어의 교수 등급과는 무관하게 항상 고빈도로 출현하며, 어휘 교육의 대상이기 보다는 문법 교육의 영역과 중복되기 때문이다. 다음 <표 8>은 레마화, 품사 표시 작업이 끝난 후, 어휘 범주(명사, 형용사, 동사, 부사)만을 대상으로 한 B1, B2 등급 어휘에 대한 출현수 통계를 정리한 것이다.¹⁰⁾

위의 <표 6>에서 B1에서 출현한 총 단어의 수는 314,362어절이었다. 이 중 어휘 범주만을 대상으로 한 총 어절 수인 31,746를 제외하면 B1등급에서 출현한 기능범주의 수는 282,616가 된다. 마찬가지로 B2에서도 총 <표 6>에 근거하여 총 출현한 어절의 수 120,541에서 총 출현한 어휘 범주의 어절 수인 65,799를 빼면 B2등급에서 총 출현한 기능범주의 수(54,742)가 구해진다. 이러한 기능범주의 출현 횟수는 본 연구의 주제인 기본 어휘 선정에서는 큰 의미를 가지지 않으며, 오히려 이를 제외한 어휘 범주의 총출현수와 특히 레마의 총 유형 수가 본 연구의 과정을 통하여 구해진 가장 의미 있는 수치가 될 것이다.

즉, 본 연구에서 실행한 코퍼스 구축 및 레마화, 어휘 범주 표시 등의 작업을 통하여 일차적으로 구해진 DELE B1 등급에서 한 번 이상 출현한 어휘는 그 원형을 기준으로 2872개가 되며, DELE B1과 B2 등급 모두를 통틀어서 한

<표 8> DELE B1, B2의 어휘 범주(명사, 형용사, 동사, 부사) 분포

	출현형(token)	낱말 유형(type)	레마(lemma)
B1	31,746	6,498	2,872
B2	65,799	10,315	4,767
B1+B2	97,545	11,973	5,439

10) DELE B1, B2에 출현한 레마를 빈도순으로 정렬한 자료는 지면상의 제약으로 본 논문을 통하여 모두 제시하기는 어렵다. 그러나, 이중 고빈도 상위 200개 어휘 목록을 부록 1)에서 제시하도록 하겠다.

번 이상 출현한 어휘의 종류는 총 5439개가 된다.¹¹⁾ B2 등급의 총 레마 유형은 4767 개로 B등급 전체에 출현한 레마 유형과 큰 차이를 보이지 않는다. 또한, B2 단계에서 출현한 4767 개의 어휘 중 2200개는 B1에서 중복 출현하고 있음을 알 수 있는데, 따라서 B2 등급의 학습자는 B1 단계의 어휘를 기본으로 B2 단계의 어휘를 보완하는 것이 필요하다. 본 연구에서는 B1 단계 기본 어휘와 B단계 전체 어휘 목록을 중심으로 기본 어휘 목록의 보완 방법을 모색하도록 하겠다.

5. 목록 비교를 통한 어휘 보완

본 연구를 통하여 구축된 DELE 코퍼스와 어휘 빈도 목록은 중급 단계의 스페인어 수준을 위한 필수 어휘 목록에 포함되어야 한다. 그러나, 2005년부터 2009년까지의 DELE 시험지의 목록의 텍스트 규모를 고려할 때에 2차적 보완 작업이 반드시 필요하다. 이를 위하여, 본 연구에서는 기구축된 대용량 스페인어 코퍼스에 나타나는 어휘 빈도와 대조 및 보완 작업을 실행하도록 한다. 본 연구에서 기준으로 삼은 코퍼스는 1억어절 규모의 Corpus del español¹²⁾와 1650만 어절 규모의 CUMBRE¹³⁾ 코퍼스이다. 두 코퍼스의 어휘 빈도 목록을 확보하고, 이를 합산한 후 고빈도 상위 3000개 목록과 5000개 목록을 선정할 후 이를 1차 구축된 DELE B1 2872개 목록 및 DELE B1, B2 총 출현 5439목록과 대조한다. 구체적 대조 방법은 다음과 같다.

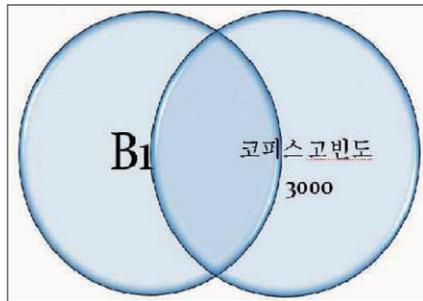
11) 연구의 목적 및 성격에 따라서는 DELE 코퍼스에 출현한 상위 고빈도 목록을 구할 수도 있으나, 본 연구에서는 DELE 기출문제에서 한 번 이상 출현한 모든 어휘를 대상으로 하였다. 이는 일상적인 언어의 텍스트를 대상으로 구축하는 코퍼스와는 달리 본 연구의 대상인 DELE 기출문제는 일상적 텍스트가 아니며, 시험이라는 텍스트의 성격상 이미 한번 정제된 특징을 가지며, 구축한 코퍼스의 규모가 그리 크지 않다는 특성 때문이다. 1회 출현하는 어휘라도 주관적 관정에 의하면 중급 DELE 기본어휘로써 충분한 가치가 있다고 판단되었다. B1등급에서 출현한 총 2872의 레마 중에서 1회만 출현한 단어의 수는 1015개 이며, 이들 중 일부의 예를 들면 다음과 같다. agudo, albino, alegre, alemán, apasionante, aplicable, asombroso, aterrizado, atractivo, atrevido, auricular, australiano, azulado, batato, beneficioso, benéfico, bienvenido, bohemio, boliviano, brillante, característico, carnívoro, céntrico, circense, climático, colectivo, colonial, comprensivo, comunitario, constipado, contentísimo, convencional, creativo, etc.

- 1) DELE B1에서 출현한 2872개 어휘 목록 중 Corpus del español(이하 CDE)와 CUMBRE의 상위빈도 3000개 목록과 대조하여, 고빈도 3000 어휘 목록 중에서 DELE B1 어휘 목록에 없는 어휘는 전문가의 판정 절차를 거쳐 최종 B1 기본어휘 목록에 추가한다.
- 2) DELE B1, B2 에서 출현한 5439개 어휘 목록 중 CDE와 CUMBRE의 상위 빈도 5000개 목록과 대조하여, 고빈도 5000 어휘 목록 중에서 DELE B등급 어휘 목록에 없는 어휘는 전문가의 판정 절차를 거쳐 최종 DELE B등급 기본 어휘 목록에 추가한다.

1) DELE B1 목록의 분석 및 목록 보완

본 연구의 결과로 산출한 DELE B1 최종 레마목록 2872개와 CDE와 CUMBRE 코퍼스 상위 3000개 목록 대조를 도식화하면 다음 <그림 2>와 같다.

우선 두 목록에서 공통적으로 포함하고 있는 어휘의 수, 즉, B1과 고빈도 3000의 교집합($B1 \cap$ 고빈도3000)에 해당하는 어휘의 수는 1701개였다. 두 목록에서 일치하지 않는 어휘의 목록 중 DELE 어휘에는 있으나 고빈도 3000에는 없는 어휘는 즉, DELE B1과 코퍼스 고빈도 3000 목록의 차집합에 해당하는 어휘는 총 1171 항목이었다. 이 1171개 항목은 이미 DELE B1 기본 어휘



<그림 2> B1 어휘 목록과 코퍼스 고빈도 3000어휘의 비교

- 12) <http://www.corpusdelespanol.org>. 13세기부터 20세기에 걸친 1억어절 규모의 스페인어 코퍼스이다. 크게 학술(académico), 뉴스(noticia), 소설(ficción), 구어(oral)의 네 장르를 포함하며, 검색에 편리한 tool이 장착되어 있다.
- 13) CUMBRE(Corpus lingüístico del español contemporáneo) 코퍼스는 SGEL의 후원으로 Gran Diccionario de Uso del Español Actual를 제작을 위하여 Murcia대학을 통하여 구축한 2000만 규모의 현대 스페인어 코퍼스이다.

〈표 9〉 B1-고빈도3000 목록 (1171개 중 일부)

puzle, gelatinado, gelatinar, multivitamínico, regordeto, preinca, fotomulta, pub(s), podcast, net, mail, espresso, fitness, gourmet, puzle, raíl, ring, rock, surf, suspense, taichi, vals, folklore, aeronáutico, albino, arqueológico, audiovisual, auricular, conmemorativo, constipado, cultoso, desinfectante, fluorescente, folclórico, herbívoro, homogéneo, inmobiliario, interpersonal, multicolor, radiofónico, sociocultural

목록에 포함된 어휘이면서, 전체 스페인어 고빈도 3000개 순위 이내에는 들지 않는 어휘 집합이다. 이 어휘는 DELE 텍스트의 특성을 파악하는데 좋은 단서가 된다. 즉, DELE 시험의 경향이 어떤 유형의 텍스트 혹은 어휘의 특성을 포함하고 있는가를 파악하는 단서가 될 수 있다. <표 9>에서 이러한 총 1171개 어휘 중 일부의 예를 보이도록 하겠다.

위의 어휘들의 특징을 살펴보면, *Real Academia Española*(이하 RAE) 사전에 등재되지 않은 신조어, 외래어 등이 다수 포함되어 있음을 알 수 있다. 또한, 시사적이며 실용적 생활 어휘가 많이 포함되어 있다. *gelatinado, gelatinar, regordeto, multivitamínico, preinca, fotomulta*와 같은 단어는 RAE 미등재어이며, *pub(s) podcast net mail espresso fitness, gourmet, puzle, raíl, ring, rock, surf, suspense, taichi, vals, bestseller*와 같은 외래어도 다수 포함되어 있다.

다음으로는 코퍼스 고빈도 3000 목록에는 포함되어 있으나, DELE B1 2872개 목록에는 포함되지 않은 어휘 목록에 대하여 언급하고자 한다. 이 목록들은 DELE B1 텍스트에는 출현하지 않았으나, 스페인어 상위 3000 목록에 포함되는 어휘이므로 B1 기본 어휘 목록에 보완되어야 할 목록이 될 것이다. 코퍼스 고빈도 3000과 DELE B1 목록의 차집합, 즉 (코퍼스 고빈도 3000 - DELE B1)에는 약 1299개의 어휘 목록이 포함된다. 이들 어휘는 아직까지 DELE B1 시험에 출제되지 않았으나, 고빈도 어휘의 성격상 중급 B1 수준의 학습자들에게 보완되어야 할 어휘 목록이 될 것이다. 이 어휘들을 자세히 살펴보면 두 가지 유형으로 분석할 수 있다. 우선 <표 10>에서 제시한 어휘와 같이 초급 수준의 기초 어휘로¹⁴⁾ B1 기출 시험에서는 출현하지 않았으나, 복

〈표 10〉 B1에 포함되지 않은 고빈도 어휘 중 A등급 어휘 (B1에서 복습할 어휘) (일부)

abogado, abrazar, absolutamente, absoluto, acercar, acostar, actor, aguantar, alegría, análisis, felicidad, frecuente, fumar, gas, gato, gobierno, guerra, hielo, hierba, hierro, hijo, hombro, independiente, insistir, labio, laboratorio, lágrima, lamentar, lección, marcha, muchacho, océano, ocultar, ocupación, odiar, odio, patio, patria, paz, pecho, pedazo, pelear, pelota, pintado, pintor, primavera, puente, puro, quitar, quizás, rápidamente, rey, ruso, sentir, silla, similar, temer, tía, toro, traducción

〈표 11〉 B1에 포함되지 않은 고빈도 3000어휘 중 예습할 필요가 있는 어휘 (일부)

abordar, abstraer, acceder, acelerar, acumular, acusación, acusar, agitar, agrario, agregar, albergar, aliado, alianza, coincidir, creciente, criterio, dotar, eficacia, eje, electrón, fiscal, forestal, fortalecer, fracasar, fracaso, fusión, huelga, impacto, integración, integrante, inversión, invertir, involucrar, jurídico, legislación, legislador, masivo, meridional, molécula, monarca, monarquía, monetario, montañoso, obstar, penal, privilegio, proseguir, proteína, provenir, radiación, ratificar, relieve, reto, sindical, transcurrir, tropa

〈표 12〉 스페인어 B1 기본 어휘 (최종)

B1 코퍼스 어휘 목록 (2872 항목) + 고빈도 보완 어휘 목록 (1299 항목)
= 스페인어 B1 기본 어휘 최종 목록 (4171 항목)

습이 필요한 어휘 목록으로 볼 수 있을 것이다.

1299 목록 중에는 DELE A1, A2의 기본어휘에 포함시키기에는 난이도가 높은 어휘들이 포함되어 있는데, DELE B1에서 아직 한 번도 출현하지 않았으나, 앞으로 출제 가능성이 높은 고빈도 어휘 목록으로 DELE B1의 보완 목록 중 예습 목록으로 포함시킬 수 있을 것이다. <표 11>은 어휘 목록 중 일부를 보인 것이다.

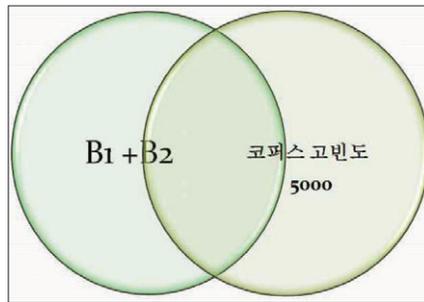
이미 본 연구의 1장과 2장에서 밝힌 바와 같이, 본 연구의 목적은 스페인어 DELE 코퍼스 구축을 통하여 중급 스페인어 학습을 위한 기본 어휘의 규모 및 어휘 목록을 확정하는 데 있다. 그렇다면, 1차적으로 DELE B1 코퍼스의 출현 어휘 2872개 항목에 고빈도 3000 개 목록 중 B1어휘 목록과 중복되지 않는 1299 어휘를 보완한다면, B1 등급을 위한 기본 어휘 목록을 선정할 수 있게 된다. <표 12>에서 보이는 것처럼 대략 4200개 항목의 어휘를 B1 단계의 기본 어휘로 선정할 수 있다.

14) DELE A1, A2 어휘 목록에 포함되는 어휘이다.

2) DELE B 등급 전체 어휘의 분석 및 목록 보완

다음으로는 DELE B 단계 전체 목록에 대한 분석 및 보완 방법에 대하여 살펴보도록 한다. 본 연구의 결과로 산출한 DELE B1, B2 최종 레마목록 5439 항목과 CDE와 CUMBRE 코퍼스 상위 5000개 목록 대조를 벤다이어그램으로 도식화하면 다음 <그림 3>과 같다.

우선 두 목록에서 공통적으로 포함하고 있는 어휘의 수, 즉, B1+B2 코퍼스 어휘 5439 항목과 고빈도 5000 항목의 교집합(B1+B2) ∩ 고빈도5000)에 해당하는 어휘의 수는 3190개였다. 또한, 코퍼스 고빈도 목록 5000 항목과 (B1+B2) 기본 어휘의 차집합, 다시 말하면 스페인어 상위 고빈도 5000개 목록 중에서 B1+B2 기본 어휘와 중복되지 않은 어휘의 수는 1810개 어휘가 된다. 이 어휘들은 앞서 B1 기본 어휘의 선정에서와 마찬가지로 스페인어 B등급 어휘의 최종 목록에서 보완되어야 할 목록이 될 것이다. 따라서, 스페인어 B등급 최종 어휘의 규모는 다음과 같이 산정할 수 있다.



<그림 3> B1, B2 총어휘 목록과 코퍼스 고빈도 5000어휘의 비교

<표 13> 스페인어 B 등급 기본 어휘

B1, B2 코퍼스 어휘 목록(5439 항목) + 고빈도 보완 어휘 목록(1810 항목) = 스페인어 B 등급 기본 어휘 최종 목록 (7249 항목)
--

15) Bartól Hernandez(2006)에서는 DELE 중급 기본 어휘의 규모를 능동적 어휘(vocabulario activo) 5817개, 수동적 어휘(vocabulario pasivo) 3363으로 총 9180개 규모의 어휘를 제시하고 있다.

즉, 스페인어 중급 학습자가 습득하여야 할 어휘의 규모는 대략 7300 항목 규모라는 결론을 내릴 수 있다.¹⁵⁾

DELE B등급 코퍼스의 5439 어휘에 보완되어야 할 고빈도 1810개 어휘는, DELE B1, B2 텍스트에는 출현하지 않았으나, 스페인어 상위 5000 목록에 포함되는 어휘이므로 중급 단계에서 학습할 가치가 높은 어휘항목이며, 따라서 B 등급 기본 어휘 목록에 보완되어야 할 목록이 된다. 이들 1810개 어휘는 아직까지 DELE B 단계 시험에 출제되지 않았으나, 고빈도 어휘의 성격상 중급 수준의 학습자들에게 반드시 보완되어야 할 어휘 목록이 될 것이다. B1 등급 어휘 분석에서와 마찬가지로 이 어휘들을 분석한 결과 다음과 같은 두 유형으로 분류할 수 있는데, 우선 <표 14>에서는 제시한 어휘 목록은 DELE B 등급 학습자를 위한 복습 어휘 목록이 되겠다. 즉, B 등급의 기출 어휘는 아니나, 고빈도 5000에 포함되고, DELE A 등급 출현 어휘로 B 단계의 학습자라면 반드시 이미 알고 있어야 하는 필수 어휘 목록이다.

다음으로 <표 15>에서는 DELE A 등급, B 등급의 기출어휘는 아니나, 고빈도 상위 5000개 어휘에 포함되어 있고 추후 B등급 시험에 출제 가능성이 있어 학습할 가치가 있는 예습 어휘 목록을 제시하도록 하겠다.

<표 14> B1, B2 목록에 없으나, 고빈도 상위 5000개 목록에 있는 A등급 복습 어휘(일부)

abandonado, abandono, abogado, acaso, asustado, atacar, bañar, basura, blando, bomba, callar, despertar, doblar, doña, enormemente, espontáneo, esposo, estallar, herir, huir, húmedo, inocente, lágrima, licencia, líder, mezcla, militar, cupación, odio, particularmente, pecho, pelea, perdonar, pescador, pintor, pionero, presidente, prestar, prima, primavera, primitivo, rechazo, situado, sobrino, oplar, sordo, sospecha, subterráneo, echo, tía, vacaciones, vino

<표 15> DELE B 등급 예습 어휘 목록 (일부)

abstraer, asalto, asentamiento, asesor, asomar, atmosférico, atómico, azufre, bachillerato, balanza, barril, élico, bíblico, bordar, bóveda, bulto, buque, cadáver, canceller, canoso, clausurar, combustible, demibar, derrotar, derumbar, desembocar, desfile, designar, deslizar, despojar, destrozar, deteriorar, disolver, dispersar, dramaturgo, eclesiástico, englobar, espantar, estirar, erroviario, fraude, frotar, hemisferio, hereditario, hormona, ilícito, iluminación, implantar, impregnar, impunidad, índole, jurídico, jurisdicción, maldecir, mandíbula, meridional, obstar, óptico, órbita, penumbra, plática, platicar, prevalecer, quirúrgico, recitar, sacudir, sonda, tallo, tragar, tregua, voltear

본 연구에서 구축한 스페인어 DELE 중급 기본 어휘 목록은 스페인어 교육 현장의 경험이 있는 연구자의 주관적 판정과 보완 작업이 후속 작업으로 이루어지게 되어, 새로운 어휘 목록이 지속적으로 보완, 수정되어야 할 것이다. 또한, 새로운 DELE 기출문제가 공개 될 때마다 코퍼스로 구축되어 기존의 코퍼스의 규모를 점차적으로 확장하여, 좀 더 객관적으로 신뢰성이 높은 자료로 구축되는 것이 필요할 것이다.

IV. 결론 및 활용 방안

본 연구에서는 스페인어 중급 학습자를 위한 기본 어휘를 B1단계에서는 대략 4000개 내외로, B 전체 단계에서는 7000개 내외로 산정하였다. 이를 위하여, 코퍼스 구축 방법론 및 어휘 선정의 객관적, 주관적, 절충적 방법론을 제시하였다. 본 연구에서 구축한 코퍼스 및 선정 어휘 목록을 활용할 수 있는 방안은 다양하다. 우선적으로 스페인어 중급 교재 편찬의 기본 어휘를 제공할 수 있으며, 스페인어 학습 사전의 표제어 선정 및 사전의 예시문 선정의 기초 자료가 될 수 있다. 사전의 규모에 따라서 소사전, 중사전의 경우 가장 빈도가 높고, 활용도가 높은 기본 어휘를 선정하는 것이 필요한데, 이를 위하여 DELE 기본 어휘 목록이 어휘 선정의 기준으로 사용될 수 있을 것이다. 또한, 중급 교재의 편찬에 있어서 필수 어휘 목록을 제공할 수 있을 것이며, 교재 집필 시 어휘 통제의 기준을 마련할 수 있을 것이다. 또한, DELE B등급 시험 준비를 위한 교재 편찬에도 매우 유용하게 쓰일 수 있는데, 특히 B등급 부터는 어휘량이 급격하게 늘어나는 시기이므로, 중급 DELE 자격증의 준비를 위한 수험서는 어휘 중심의 교재가 적당할 것이다.

그러나, 선정된 어휘를 어떻게 효과적으로 교수할 것인가하는 문제가 여전히 남게 되는데, 제시 방법으로는 품사별, 알파벳 순, 의미적 관련어 순 등의 다양한 방법이 있을 수 있다. 효과적 어휘 학습 방법에 대하여는 추후 연구에서 더욱 체계적 방안이 제시되어야 하겠으나, 그 중 코퍼스의 용례색인

〈표 16〉 medio를 중심으로 자주 출현하는 클러스터

medio ambiente, la protección del medio ambiente, medio de transporte, medio de comunicación, medio de visitantes, durante medio año, cuál es el medio~, de medio siglo, el medio local, nacional~, el medio más rápido, el medio que usted prefiere~, en medio de, en ningún medio (televisión, Internet)~, evidentemente el medio más~, hay un medio, hay un medio sencillo, medio sencillo y eficaz, coste (alto, medio~, ¿Qué medio~, en nivel medio, medio kilo, identificaba en medio de~, la musica como medio, medio de la multitud, alto, medio o bajo, medio y superior, situada en medio de

프로그램의 클러스터(cluster) 추출 기능을 활용하는 방법을 제안하면 다음과 같다. 클러스터 혹은 연어 검색 기능은 특정 어휘를 중심으로 좌, 우로 가장 자주 결합하는 어휘를 보이는 기능이다. 이 기능을 통하여 실제로 어휘가 어떤 결합형 혹은 숙어형으로 자주 쓰이는 가를 볼 수 있으며, 어휘 교육에서 가장 중요한 부분이 연어 교육의 기초 자료를 제공해 준다. 아래 <표 16>는 medio를 중심으로 좌, 우 3개의 단어 범주 내에서 가장 자주 결합을 보이는 클러스터만을 추출하여 정리한 것이다.¹⁶⁾ 예에서 보는 바와 같이 매우 유용한 숙어 어휘 결합이 나타남을 볼 수 있다.

그 밖에도 관련어(유의어, 반의어, 파생어 등)를 통한 제시 방법이 있을 수 있는데, 특히 ‘aclaración, aclarar, claro’, ‘acomodar, cómodo, incómodo’ 과 같이 품사에 상관없이 굴절과 파생 변이형을 모두 포함함 단어군(word family)을 함께 제시한다면, 학습자들이 어휘 간의 연상 작용을 도와 어휘 학습의 효율성을 높여줄 수 있을 것이다. 그 밖에도 ‘의, 식, 주’와 같은 의미적 영역에 따른 어휘의 제시와 같은 방법이 있을 수 있으나, 이는 추후 연구의 주제로 남겨 놓도록 하겠다.

본 연구에서 사용된 코퍼스 구축 및 가공 방법은 스페인어 연구를 위해 직접 코퍼스를 구축하려는 연구자들에게 자료를 전산화 하고 정보를 입력하는 방법론을 제시함으로써 이를 토대로 다른 코퍼스 구축의 확장을 가능하게 할 수 있다는 점에서 그 의의를 둘 수 있을 것이다. 또한, 스페인어 공인 자격 시험인 DELE에 대한 실증적, 계량적 연구를 실행함으로써 DELE에 대한 객

16) 용례색인 프로그램의 좌우 단어의 span을 설정하는 기능을 활용한다.

관적인 분석을 할 수 있었으며, DELE를 준비할 수 있는 교재 개발에 기초 자료를 제공할 수 있을 것이다. 선정된 DELE 기본 어휘 목록은 특정 교재의 어휘 난이도 분석 및 어휘 분포 등을 점검하는 기준 목록이 될 수 있을 것이다. 또한, 구축된 코퍼스는 지속적으로 확장, 보완되어 스페인어의 순수 언어학 연구 및 응용 언어학 연구에서 재사용될 수 있다는 점에서 매우 활용도가 높다고 볼 수 있다.

참고문헌

- 김경희(2001), 「스페인어 기초 어휘 선정과 교수법」, 서어서문연구, 19권, pp. 3-18.
- 김광해(2003), 「국어교육용 어휘와 한국어 교육용 어휘」, 국어교육, 11, pp. 255-291.
- 김광해(1993), 『국어 어휘론 개설』, 집문당.
- 서상규(2009), 『교육용 기본 어휘 선정을 위한 기초 연구』, 국립국어원.
- 서상규, 강현화, 유현경(2000), 『한국어 기초 어휘 의미 빈도 사전의 개발 사업 결과 보고서』, 문화관광부.
- 서상규, 한영균(1999), 『국어정보학 입문』, 인문학과 컴퓨터, 태학사.
- 신동광(2011), 「기본 어휘의 선정 기준 -영어 어휘를 중심으로」, 국어교육학연구, 40권, pp.217-243.
- 임지룡(1991), 『한국어의 기초어휘에 대한 연구』, 1990년도 교육부 지원 한국학술진흥재단의 자유공모과제 학술연구.
- 임철성(2002), 「초급 한국어 교육용 어휘 선정 연구」, 국어교육학연구 14, 355-385.
- 조남호(2002), 『현대 국어 사용 빈도 조사-한국어 학습용 어휘 선정을 위한 기초 조사』, 국립국어 연구원.
- 최재철(2007), 「언어 교육을 위한 유럽 공통 기본지침에 관한 연구」, 스페인어 문학, 43권, pp. 233-250.
- Almela, R., Cantos, P., Sánchez, A., Sarmiento, R., Almela, M.(2005), *Frecuencias del español. Diccionarios y estudios léxicos y morfológicos*,

- Madrid: Ed. Universitas.
- Bartol Hernández, José A.(2006), “La disponibilidad léxca,” *Revista española de lingüística* 36, pp. 379-396.
- Davies, M.(2006), *Corpus del español* [On line], <http://www.corpusdelespanol.org>.
- Davies, M. & Face T.L.(2006), “Vocabulary Coverage in Spanish textbooks: How representative is it?,” In Sagarra and A. J. Toribio (eds.) *Selected proceedings of the 9th Hispanic Linguistics symposium*, Somerville: Cascadilla, pp. 132-143.
- Davies, M(2006), *A Frequency Dictionary of Spanish: Core Vocabulary for Learners*, New York: Routledge.
- Fajardo Aguirre, Alejandro(1991), “Los vocabularios básicos,” *Actas de III Congreso Nacional de ASELE*.
- Hamamatsu, Noriko(2005), “Propuesta de elaboración de un diccionario básico bilingüe español-japonés, japonés-español para principiantes,” *Actas del II Congreso Internacional de Lexicografía Hispánica*, pp. 428-435
- Lewis, Michael, ed.(2000), *Teaching collocation: Further Developments in the Lexical Approach*, Hove: Language Teaching Publications.
- Manuel Santiago, José(2001), “La selección del vocabulario didáctico: un nuevo enfoque,” *Revista de Lexicografía*, 7, pp. 155-172.

신 자 영

서울 서대문구 신촌동 134, 연세대학교 언어정보연구원
 jayoungshin@yonsei.ac.kr

논문투고일: 2011년 9월 14일

심사완료일: 2011년 10월 10일

게재확정일: 2011년 10월 17일

부 록

〈표 1〉 DELE B1, B2 출현 고빈도 어휘 목록 (상위 200 선정)

순위	레마	품사	빈도	순위	레마	품사	빈도
1	ser	n	2879	101	adaptar	v	168
2	número	n	1511	102	gramática	n	164
3	haber	v	1157	103	niño/a	n	162
4	texto	n	1094	104	espacio	n	155
5	tener	v	1052	105	podíamos	v	155
6	estar	v	886	106	así	adv	154
7	nivel	n	828	107	noticia	n	154
8	segundo	a	820	108	semana	n	150
9	ir	n	744	109	comprar	v	150
10	prueba	n	729	110	rellenar	v	150
11	intermedio	a	671	111	ciudad	n	149
12	opción	n	666	112	poner	v	149
13	todo	a	664	113	necesario	a	148
14	más	a/adv	664	114	país	n	145
15	ejemplo	n	643	115	leer	v	145
16	hacer	v	624	116	curso	n	144
17	hoja	n	618	117	contestar	v	144
18	instrucción	n	568	118	carta	n	141
19	respuesta	n	550	119	lugar	n	140
20	dos	a/n	549	120	antes	adv	139
21	pregunta	n	548	121	tarde	n	139
22	poder	v	489	122	utilizar	v	139
23	marcar	v	479	123	caso	n	135
24	muy	adv	458	124	mundo	n	135
25	escribir	v	451	125	obra	n	135
26	año	n	436	126	siempre	adv	133
27	parte	n	424	127	premio	n	133
28	mucho	a	421	128	concurso	n	131
29	decir	v	395	129	nada	n	130
30	deber	n	365	130	inicial	a	129
31	cada	a	361	131	viaje	n	128
32	continuación	n	361	132	vida	n	128
33	correspondiente	a	341	133	parecer	v	128
34	hombre	n	332	134	venir	v	128

〈표 1〉 계속

순위	레마	품사	빈도	순위	레마	품사	빈도
35	bueno	a	326	135	final	n	127
36	mujer	n	313	136	creer	v	127
37	vez	n	302	137	trabajar	v	127
38	diálogo	n	289	138	chico	n	126
39	día	n	279	139	tipo	n	126
40	dar	v	277	140	actividad	n	124
41	querer	v	276	141	España	n	124
42	otro	a	274	142	tercero	a	123
43	caja	n	274	143	marca	n	123
44	comprensión	n	273	144	mañana	n/adv	122
45	siguiente	a	270	145	cuarto	a	120
46	escuchar	v	269	146	mes	n	120
47	uno	a/n	266	147	además	adv	119
48	ver	v	265	148	diferente	a	116
49	persona	n	262	149	chica	n	116
50	candidato	n	257	150	presentar	v	116
51	saber	v	257	151	amigo/a	n	115
52	después	adv	251	152	público	a	113
53	español	a/n	248	153	tomar	v	113
54	pasar	v	244	154	auditivo	a	112
55	tiempo	n	243	155	palabra	n	111
56	entrevistador	n	240	156	salir	v	111
57	correcto	a	239	157	medio	a	110
58	lectura	n	239	158	ahora	adv	109
59	casa	n	232	159	importante	a	108
60	inscripción	n	228	160	tal	a	108
61	gustar	v	228	161	dejar	v	108
62	hablar	v	223	162	tanto	a	107
63	columna	n	222	163	hoy	adv	107
64	tres	a/n	220	164	tan	adv	105
65	voz	n	220	165	festival	n	105
66	cuatro	n	218	166	empezar	v	104
67	vertical	a	217	167	favor	n	103
68	audición	n	213	168	interpretación	n	103
69	blanco	a	212	169	empresa	n	102
70	oir	v	208	170	padre	n	102

〈표 1〉 계속

순위	레마	품사	빈도	순위	레마	품사	빈도
71	sólo	adv	207	171	pensar	v	102
72	también	adv	206	172	volver	v	102
73	encontrar	v	202	173	esperar	v	100
74	pues	adv	201	174	pedir	v	99
75	verdadero	a	199	175	mayor	a	98
76	quedar	v	199	176	considerar	v	97
77	libro	n	198	177	participar	v	97
78	falso	a	197	178	cualquier	a	96
79	nuevo	a	197	179	internacional	a	95
80	hora	n	194	180	música	n	95
81	llamar	v	193	181	noche	n	95
82	elegir	v	189	182	vivir	v	95
83	grande	a	185	183	cinco	n	94
84	mismo	a	184	184	verdad	n	94
85	trabajo	n	182	185	contar	v	94
86	fixar	v	182	186	grabación	n	93
87	poco	a	181	187	radio	n	93
88	ya	adv	179	188	último	a	92
89	llevar	v	176	189	autor	n	91
90	gracia	n	175	190	mostrar	v	91
91	conocer	v	175	191	programa	n	90
92	bien	adv	174	192	clase	n	89
93	mejor	a	173	193	diploma	n	89
94	vocabulario	n	172	194	narrador	n	89
95	llegar	v	172	195	escritor	n	88
96	transcripción	n	170	196	cuadro	n	87
97	aparecer	v	170	197	buscar	v	87
98	realizar	v	170	198	preferir	v	87
99	terminar	v	170	199	recibir	v	87
100	ofrecer	v	169	200	solamente	adv	86