

분과

1

디지털 환경에서의 한국어 평가와 교사의 역할

사회 : 김은호(중앙대학교)

## 음성 인식 기술(Speech Recognition)을 기반으로 한 한국어 학습자의 발음 진단·평가 시스템 구축 방안 연구

이미향(영남대학교) · 안미애(경북대학교)

### 1. 들어가며

발음은 모국어의 간섭을 가장 많이 받는 영역이다. 한국어 학습자의 발음은 초기에는 숙달도에 영향을 받으나 오류의 화석화로 나중에는 숙달도에 비례해 발전하는 영역이 아니다. 이러한 관점에서 현재 한국어 교육 분야의 한국어 발음 연구 결과는 300편이 넘는다. 그러나 이들 연구의 대부분이 한국어 발음 현상에 대한 이론적 접근을 기반으로 한 것이다. 300편 중에서 교육학으로 분류된 것이 30여 편, 공학 논문이 약 40편인 데 비해, 인문학 분야가 230여 편인 연구 현실은 발음 교육 연구 분야가 이론적 성격이 강하다는 점을 보여준다. 김상수·송향근(2006)에서도 1986년~2006년까지 101편의 발음 교육 연구는 대체로 구체적인 교수법보다 일반적인 현상 나열이 많다고 하였다. 이러한 연구 현황은 2006년 이후, 최근 10년간도 크게 다르지 않다.<sup>1)</sup>

한국어 발음 교육에 관한 연구 중에서도 실험음성학적 또는 대조언어학적 관점을 활용한 연구는 발음 오류의 원인을 파악하고 교육 방안을 제안하며 주관적인 발음의 영역에 객관성을 높였다는 데 의의가 있다. 그런데 이러한 연구에 한계가 있다. 첫째, 발음은 모국어 간섭이 큰 영역이라 언어권별로 살피는 경향이 있는데, 이는 일반화에 한계가 있다. 둘째, 실험음성학적 또는 대조언어학적 관점의 연구 방법이 언어권별로 교수 방안을 제언하는 데 그치는 아쉬움이 있다. 셋째, 언어권별 연구 외에 범언어적으로 적용 가능한 초분절음소 연구에 주목하지 못한 면이 있다. 발음의 정확성을 높이는 데 음소 교육이 중요하지만, 발음의 유창성에 영향을 미치는 초분절음소에 대한 교육 내용이 소홀히 다루어져서는 안 되기 때문이다. 넷째, 실험 음성학이라는 연구 방법을 접목한 연구들이 소수의 피험자를 대상으로 한 개별 표본 연구가 대부분이라는 면이 있다. 이제는 개별적인 연구들의 결과물을 종합하여 객관적으로 수집되고 정리된 양적 음성 데이터베이스에 기반한, 종합적인 관점에서 실용성 높은 연구가 필요한 시점이다.

또한 한국어 학습자의 학습 기간이나 거주 기간이 발음 숙달도나 정확도에 완전히 비례하지 않는다는 점도 고려해야 한다. 안미애(2014)는 고급 수준의 한국어 학습자라고 해서 원어민의 수준에 이르지 못하며, 이해정(2015)은 거주 기간이 발음 숙달도에 결정적 영향을 주지 않음을 증명했다. 이와 같은 맥락에서 학습자의 발음 향상이 발음 연습 횟수나 기간보다는 발음 교수 방법에 영향을 받는다는 윤은경(2013)의 연구는 발음 숙달도가 시간보다는 의도적 훈련에 의해 향상됨

1) 발음 오류 분석(김현주 2008, 허용 2012), 대조 분석(김선정, 육효창 2010), 발화 속도 연구(김미란, 강현주, 노주현 2014) 등 이론적 접근이 대부분이다. 한국어 교육적 연구도 발음 교육 요소(권경근 2016), 발음 교재 분석(박정은, 이주희 2008), 발음 습득 과정 연구(김선정 2014, 백경미, 김태경 2014), 한국어 음운이론의 교육적 접목(박기영 2010, 송향근 2011) 등이며, 좀 더 실제성 높은 연구로 발음 규칙 적용 방안(정명숙 2008, 김형복 2009, 김서형·장향실·차재은·전나영 2016)과 학습자의 발음 인식(안미애·Wei Qun·이미향 2016), 한국어 교사교육(권성미 2014, 백소영 2016, 이미향·안미애 2018)이 있다.

을 지지하는 연구 결과이다. 해당 연구에서는 발음 훈련 전의 ‘음성적 차이에 대한 지각 훈련’이 말하기 능력을 미약하게나마 향상시킨다고 보고한 바 있다. 따라서 좀 더 객관적이고 체계적인 한국어 발음 교수와 연구를 위해, 학습자 음성 코퍼스 구축이 무엇보다도 필요한 시점이다.

그러나 한국어 학습자의 음성 코퍼스는 텍스트 코퍼스에 비해 구축하기가 어렵다. 우선 음성 자료를 속달도별로 수집해야 하며, 수집한 음성 자료를 다시 텍스트로 전사해 데이터베이스화하는 단계가 수반되기 때문이다. 기존에 구축된 학습자 음성 코퍼스가 전무한 것은 아니다. 대용량 음성 코퍼스로 이석재 외(2004) 등이 구축한 ‘외국어로서의 한국어 음성 코퍼스(L2KSC)’가 있으나 이 음성 코퍼스는 외국어로서의 한국어 학습자의 발음, 음성 산출 오류 등의 기초 연구 자료로 쓰인다. 교육 현장의 교사나 학습자의 손에서 활용되는 단계, 이른바 ‘실용적 목적’을 달성하는 직접적인 도구는 아니다. 본 연구는 이러한 상황을 직시하고, 연구 분야와 방법을 확산하여 학제간 연구를 통해 학습자의 한국어 발음 오류를 음성 코퍼스로 구축할 방법론을 모색해 보고자 한다. 그리고 이를 바탕으로 자모를 통한 언어권별 접근과 함께, 억양에 중점을 둔, 범언어권별 연구를 확산할 계기를 마련하고자 한다.

## 2. 연구의 필요성

4차 산업시대에 교육학 분야에는 생체 신호, 알고리즘, AI 등 공학적 접근이 모색되고 있다. 그러나 학습자 발음 진단과 평가 시스템 개발은 아직 시대적 요구에 부응하지 못하고 있다. 그 까닭은 개인이 대량의 음성 코퍼스를 구축하거나 분석하기에 어려울 뿐만 아니라 그 결과를 상용화하여 성과를 얻으려면 공학적 접근이 필요한데, 그간 한국어 발음 교육가 학제 간 연구로 이루어지지 않은 데 있다. 공학계에서는 현실적인 수요와 요구를 알아 이를 프로그램화하려 노력하고 있으나, 언어학 및 교육적 접근에서 섬세하지 않은 실정이다. 인문사회계 전공자가 대부분인 한국어교육학계는 이를 개발할 필요성을 제안하는 데 머물러 있고, 개발된 소수의 프로그램에는 사회적 요구가 충분히 반영되어 있지 못하다.

지금까지 교재를 통해 제공된 발음 학습 내용은 주로 교사의 눈으로 관찰된 입이나 혀 모양 정보를 이용한 것이다. 이러한 방법으로는 개인 경험의 누적은 객관성을 확보할 수 없다. 눈으로 관찰되지 않는 부분이 있고, 발음이 사람마다 생리적 특성에 따라 다른 것은 자명하기 때문이다.

이를 극복하기 위해 한국어 발음 연습을 위한 다양한 프로그램이 출시되어 있다. 그러나 여기에는 언어적 접근상의 어려움, 발음 진단과 개선 정보 미제공 등 여러 문제가 있다. 예를 들어 국립국어원의 ‘바른 소리’는 올바른 발음에 대한 모범만 제시하고 있고, 설명 언어도 영어와 일본어로 한정되어 있다. 세종학당의 ‘Poppopping Korean’은 PC와 스마트폰으로 학습자들이 접근할 수 있으며, 학습자가 녹음한 자신의 발음을 확인하는 단계에 나아갔다. 그러나 흥미로운 연습 과정임에도 불구하고 학습자의 발음을 진단해 주거나 개선 정보를 제공해 주지 않는다.

또 다른 프로그램인 ‘Korean Listening Practice’는 영어 기반 한국어 듣기 연습 앱이다. 해당 언어에 능숙하지 않은 학습자가 접근하기가 어려울 뿐만 아니라, 발음을 진단하거나 학습 방법을 제공하는 기능이 없다. 가장 최근에 개발된 ‘세종학당 AI 선생님’(2020)에는 발음 진단 기능이 있으나, 속달도가 높은 학습자가 느리게 발화할 때 오히려 발음 일치도가 낮은 것으로 나오는 문제가 확인된다.

특정 대상자에게 외국어 발음을 정확하게 학습하게 하려면 대상 학습자의 발음 정확도부터 파악해야 하는데, 이를 위해 해당 학습자의 속달도별 오류를 수치화하는 것은 필수적이다. 이에 본 연구는 한국어 학습자 음성 코퍼스 구축과 이에서 도출된 발음 오류 데이터를 바탕으로 한 발음 진단 및 평가 시스템 개발의 필요성을 제기한다. 또한 실제로 한국어 학습자의 발음을 진단하고 평가할 한국어 발음 평가 시스템을 개발하기 위한 학제 간 연구 방법론을 모색해 보고자 한

다. 한국어 학습자 대상 음성 정보 구축과 누적된 교육 경험만으로는 현재 다양한 학습자의 요구를 감당하기에 한계가 있다. 학습 공간을 확장하고 자기 주도적 학습에 용이하여 정규 학습 과정의 부족한 교육 내용을 보완해 줄 모바일 러닝이 다양한 분야에서 접목되고 있을 뿐만 아니라, 학습자에 의한 요구가 오히려 보고되고 있는 단계다. 이제는 교육 현장에서 학습자가 개별적으로 참여하여 자가 진단할 한국어 발음 교육 프로그램이 일상화되어야 할 단계다.

### 3. 한국어 학습자의 발음 진단 및 평가 시스템 구축 방안

#### 3.1. 발음 오류 음성 코퍼스의 구축 대상과 규모

발음 진단 및 평가 시스템을 개발하기 위해서는 한국어 학습자 음성 코퍼스 구축 및 이 과정에서 도출된 발음 오류 데이터가 우선적으로 필요하다. 먼저 한국어 발음 오류 코퍼스를 구축하기 위한 방법이 마련되어야 한다. 본 연구팀이 2017년에 개발한 <한국어 발음 학습 모바일 애플리케이션>을 이용하여 한국어 학습자의 발음을 수집하는 것이 한 방안이 된다. 음성 코퍼스는 학습자의 모국어에 따라 차이가 있기에 언어권별로 구축되어야 한다. 또한, 범언어권 학습자의 수요에 부응할 역량 오류의 수치화도 필요하다. 이렇게 측정된 자료를 한국인의 발음과 비교해 봄으로써, 학습자의 발음에서 나타나는 오류 실현율을 수치화할 수 있다.

본 연구팀은 안미애·Wei Qun·이미향(2016)에서 중국인 학습자 40명을 대상으로 낭독 발화 총 5,680개 문장, 자유 발화 160,000음절을 구축함으로써 대상 학습자들의 발음 특징을 살펴 중국어권 학습자의 발음 특성과 오류 경향, 요구 사항을 파악한 바 있다. 연구 결과는 중국어권 학습자의 발음을 진단하고 평가하는 앱 개발의 기초 자료가 되었다. 이러한 연구 경험을 바탕으로, 본 연구는 ‘범언어권 한국어 학습자’가 ‘공통적으로 보이는 발음 오류’를 연구 대상으로 삼아, 발음 진단 및 평가 프로그램 개발의 활용도를 높이는 것이 가능하다는 점을 강조하고자 한다.

이러한 연구를 위해서는 우선 범언어권의 범위라 할 정도의 대상 학습자의 언어권과 연구 대상자의 수를 정해야 한다. 기존에 구축된 대용량 한국어 학습자 음성 코퍼스(이석재 외(2004))<sup>2)</sup>는 이러한 점을 고려해 구축된 음성 코퍼스이나 본 연구팀이 목표로 하는 실용적이고 실수요자를 고려한 한국어 발음 진단 및 평가 시스템을 구축하는 데는 부족하다. 이에 본 연구팀은 이석재 외(2004)의 음성 코퍼스와 별도로 현재(2021년) 국내외 한국어 학습자 수로 상위권에 있는 베트남인 한국어 학습자와 그외 러시아권의 우즈베키스탄 한국어 학습자를 고려하여 자체적으로 소규모 한국어 학습자 음성 코퍼스를 구축하고자 한다. 본 연구팀은 이석재 외(2004)가 구축한 음성 코퍼스를 고려하여, 한국어 발음의 진단 및 평가에 필요한 필수적인 요소를 추출하고, 이를 바탕으로 한국어 발음 진단 및 평가 시스템을 개발하기 위한 ‘필수 음성 코퍼스’를 구축할 것이다. 본 연구팀이 구축하고자 하는 한국어 학습자 발음 진단 및 평가 시스템 개발을 위한 필수 음성 코퍼스는 4개 언어권 각 30명을 대상으로 총 120명이 발음한 음운, 단어, 문장 수준으로 단음절(음운) 4,800개, 2음절 단어 4,200개, 3어절 기준 문장 1,200개 3,600어절, 1,069음절로 구성된 문장 총 3,000개 수준으로 문장 수 기준 8,400개와 4,800개의 규모로 구성하고자 한다.<sup>3)</sup> 이는 기존의 텍스트 기반 한국어 학습자 코퍼스나 개별 연구자가 소수의 피험자로

2) 이석재 외(2004)는 영어권 외국인 화자 58명, 일본어권 41명, 중국어권 43명, 기타 25개 언어에서 가용 인원별 1-4인(2인 평균)으로 100명, 한국인 화자 40명, 국외 영어권 한국어 학습자 26명, 국외 중국어권 한국어 학습자 38명의 총 356명을 대상으로 음성과 음운을 고려한 한국어 학습자 음성 말뭉치를 구축하였다. 이는 중규모의 말뭉치로, 다양한 모국어권 한국어 학습자의 음성 코퍼스를 살필 수 있다는 장점은 있으나 영어, 일본어, 중국어권을 제외한 다른 지역 출신의 한국어 학습자 수가 적다는 단점이 있다.

3) 이석재 외(2004)는 기초 연구 자료로서 한국어 학습자 음성 코퍼스를 구축하였다는 장점을 가지고 있으나, 실용적 목적의 한국어 발음 진단 및 평가를 위한 목적으로 설계된 음성 코퍼스는 아니다. 이에 본 연구는 발화 오류 연구를 위해 필요한 기초 음성 코퍼

구축한 음성 코퍼스 연구 성과보다 유의미한 규모로, 한국어 학습자의 발음 오류를 대상으로 한 대량 정보의 디지털 데이터화로 의의가 있는 연구가 될 것으로 기대된다.

이 연구의 일차적인 목표는 ‘언어권별 한국어 학습자 발음 오류 코퍼스의 구축’이다. 그러나 궁극적인 목표는 이를 바탕으로 한 ‘범언어권 한국어 학습자를 위한 한국어 발음 진단 및 평가 시스템의 구축’으로 한국어 발음교육의 실용화를 확산시키는 것이다. 코퍼스 구축의 대상은 다양한 언어권을 고려하여 선정하여야 하나 이 연구는 언어학적 대조 연구가 아닌바, 한국어교육의 실수요자를 고려하여 우선 학습자를 선정하면서 순차적으로 진행해 가야 한다고 본다. 이것은 또한 연구의 가능 범위를 고려하여 연구의 현실화에 기여하는 주요한 조건이다.

이에 한국에 현재 유학 중인 외국인들의 국적의 분포와, 이 연구에서 중점적으로 살펴볼 ‘역양’ 변수에 영향을 미칠 수 있는 성조 언어의 여부를 고려하여 코퍼스 구축 대상의 범주와 수를 선정하고자 한다. 행정 안전부의 국내 거주 외국인 통계(2019. 11. 1. 기준)를 기반으로, 국내 거주 중인 외국인의 주요 국적은 중국(42.6%), 베트남(11.1%), 태국(4.4%), 미국(4.4%)의 순이다. 또한 통계청에서 발간한 한국의 사회 동향 보고서에 따르면, 외국인 유학생의 출신 국가는 아시아권에 집중되어 있고, 학부 기준 출신 국가의 비율은 아시아권의 경우, 중국, 베트남, 몽골과 우즈베키스탄 등이 상위권이며 미주권의 경우, 미국, 캐나다가 외국인 유학생의 출신 국가 중 높은 비중을 차지하고 있다. 이러한 한국어 학습자의 분포는 연구의 우선 순위를 정하는 데 중요한 정보로 쓰인다.

또한 이와 같은 연구에서 코퍼스 수집 시에는 한국어 학습자의 사회 언어학적 변수를 고려해야 한다. 사회언어학적 변수란 연령, 성별, 한국어 학습 기간 및 숙달도, 학습 유형(지역, 기관) 등을 이른다. 모바일 앱을 통한 발음 진단 시 성별이라는 변인은 목소리로 실현되는 발음에서 성공률을 확인할 수 있게 하는 중요한 변수이다. 그리고 연구 대상의 규모는 학습자의 숙달도를 토픽(TOPIK)을 기준으로 할 수 있는데, 토픽 3~4급 중심의 유학생이 다수를 차지하는 현실에서 한국어 학습자의 목적에 맞는 발음 진단과 훈련 가능성을 감안할 때 교육적 의의뿐만 아니라 학습자에 대한 실용성이 높아질 것이다. 피험자는 국내 학습자의 주요 연령대를 고려해 20대로 제한하고자 한다. 학습자의 한국어 수준을 객관화할 자료를 바탕으로 숙달도를 살피고, 언어권 및 모국어의 성조 여부, 숙달도, 성별을 고려해 선정한다.

한편, 발음 정보를 구축할 때 필요한 또 하나의 주요 요인이 있다. 성조 언어권과 비성조 언어권으로 나누어 피험자를 선정하는 것이다. 중급 학습자는 자모의 발음만으로 발음의 유창성을 확보할 수 없다. 이미향·안미애(2018)에서는 한국어 학습자의 휴지와 억양을 포함한 초분절적 요소가 유창성에 더 큰 영향을 미치고 있음을 청자 반응을 통해 확인한 바 있다.

이와 같이 발음 데이터 구축을 위해서는 양적 규모의 적절성, 주요 언어권 선정의 기준, 고려할 사회학적 변수 등이 포함된다. 연구 의의와 한계, 실용적인 측면과 우선 반영 기준 등을 종합적으로 고려할 때, 통계적으로 유의미한 수로 확보할 수 있는 한국어 학습자로서, 남녀 비율(각 50%)과 한국어 숙달도를 고려한 피험자는 중국 30명, 베트남 30명, 우즈베키스탄 30명, 미국 등의 영어 원어 사용 학습자 30명의 총 120명 정도가 가능한 범위로 보인다.

---

스로 이석재 외(2004)의 음성 코퍼스를 활용하되, 현실적으로 앱 개발을 위해 필요한 ‘필수 음성 코퍼스’ 구축에 집중하고자 한다. 본 연구에서는 한국어 자모음 40개와 한국어 음운 현상이 포함된 문장 35개, 억양을 살필 수 있는 서법별 문장 10개(3어절 기준), 낭독 시 의미를 고려한 억양 실현 정도를 살필 수 있는 낭독문 25개를 실험 자료로 구성할 것이다. 자유 발화 코퍼스는 이석재 외(2004)에서 중국인과 영어권 한국어 학습자를 대상으로 일부 확보하고 있으므로, 본 연구에서는 베트남인과 우즈베키스탄 한국어 학습자를 대상으로 자유 발화를 추가 수집할 수 있다. 이에 본 연구팀이 구축할 앱 개발용 한국어 학습자 ‘필수 음성 코퍼스’의 규모는 기본 발음 성취도를 살피기 위한 자료 기준으로 한국어 자모 4,800개, 한국어 음운 현상을 살펴보기 위한 단어가 포함된 문장 4,200개, 서법별 문장 1,200개, 낭독 문장 3,000개로 8,400개의 문장과 4,800개의 음운이 구축될 것으로 예상된다.

### 3.2. 음성 코퍼스 및 발음 진단평가 시스템의 구축 방안

음성 코퍼스를 통한 발음 진단 및 평가 시스템 구축 연구는 실험 음성학, 한국어 교육학, 의공학의 학제간 연구로 이루어진다. 첫 번째, 실험 음성학적 연구 분야는 기존의 음성 코퍼스를 분석하고, 사회언어학적 변수를 고려한 한국어 학습자 발음 코퍼스의 구축을 가능하게 한다. 두 번째, 한국어 교육 분야는 선행 연구를 기반으로 발음 코퍼스 구축을 위한 이론을 정립하고, 구축된 발음 코퍼스에 대한 발음 오류 유형을 해석하며, 이를 교육적으로 적용하기 위한 수정 방안을 연구한다. 세 번째, 의공학 분야에서는 발음 코퍼스를 구축하기 위한 모바일 애플리케이션 개발과 구축한 발음 코퍼스에 대한 자동 레이블링, 한국어 교육 분야에서 도출한 발음 오류 유형과 수정 방안을 바탕으로 발음 오류 진단 알고리즘을 개발하는 것이다. 이러한 연구 단계에서 각 단계에서 수행할 연구 내용과 방법을 설계하여 보이면 다음 [그림 1]과 같다.



[그림] 연구 분야와 연구 단계

첫째, 실험 음성학 분야의 목표는 범언어권 한국어 학습자의 발음 오류 수집과 분석 및 완성도 높은 발음 오류 코퍼스의 완성이다. 한국어 학습자의 발음 오류 코퍼스 구축을 위해 우선 기존에 수집된 한국어 학습자 음성 코퍼스의 결과물을 분석하고, 이를 바탕으로, ‘필수 음성 코퍼스’ 구축을 위한 기준을 마련함으로써 피험자를 확보하여 음성 코퍼스 구축을 진행한다. 필수 음성 코퍼스 구축의 1단계는 의공학 분야에서 개발한 코퍼스 수집용 모바일 애플리케이션을 이용하여 소음이 차단된 공간에서 학습자가 자신의 발음을 애플리케이션으로 입력하여 서버로 바로 저장하여 수집하는 것이다. 학

습자가 발음한 어휘와 표현의 정확도를 음성 인식 기술을 통해 인식하게 하고 그 결과를 수치화하는 작업을 병행해야 한다. 한국어 학습자의 음성과 조음 생체 신호를 수집할 시, 비교군으로 표준 한국어 구사자의 음성과 생체 신호를 1단계에서 함께 수집하여, 3단계에서 발음 오류 진단에 대비할 수 있다. 이때 이석재 외(2004)에서 구축한 한국인 병렬 음성 코퍼스도 활용 가능하다.

다음으로 발음 정보와 철자 정보까지 음성 데이터를 레이블링한 범언어권 한국어 학습자 발음 오류 필수 음성 코퍼스를 구축한다. 코퍼스 구축의 모델은 미국 오하이오 주립대에서 구축한 벅 아이 코퍼스(The Buckeye Speech Corpus)와 이석재 외(2001)에서 구축한 한국어 음성 코퍼스(Written-Spoken Korean Corpus), 윤원희 외(2015)의 한국어 자연 발화 음성 코퍼스(서울 코퍼스)를 참고한다. 벅 아이 코퍼스는 사회 언어학적 변수를 고려한 균형 코퍼스로 오하이오주를 고향으로 한 40명의 구어체 대화를 40시간 정도 녹음한 30만 단어 정도의 음성 코퍼스이며, 이석재 외(2001)의 한국어 코퍼스는 한국인의 영어 발화 음성 데이터 베이스 구축을 위해 한국어 낭독 발화 음성 코퍼스로 국내 8개 지역에서 연령과 성별을 고려한 342명의 단어, 문장, 담화 음성 코퍼스이다. 윤원희 외(2015)는 서울 출신의 10대에서 40대까지의 성별과 연령을 고려한(세대별 남녀 5명씩) 자연 발화 코퍼스로 음성학적 정보와 음운론적 정보를 레이블링하여 연구용으로 사용이 가능한 자연 발화 음성 코퍼스이다.

수집된 음성 데이터는 목적별로 분절하여 음성 레이블링 작업 후 코퍼스로 구축한다. 음성 분절 시 데이터의 음질을 고려하여, 노이즈가 있을 경우에는 노이즈의 종류에 따라 적절히 가공하여 잡음을 감소시킨다. 음성 레이블링은 음절, 단어, 문장 단위로 철자와 발음, 억양 수준에서 진행된다. 철자와 발음, 억양 정보가 담긴 음성 코퍼스는 Praat 프로그램을 통해 분석되도록 가공될 것이다. 이렇게 구축된 코퍼스는 언어학적 연구뿐만 아니라 음성 인식 및 합성과 같은 음성 공학적 연구에서도 유용하다.

둘째, 한국어 교육 분야의 역할은 진단 및 평가 시스템을 개발하는 데 주요한 정보를 제공하는 것이다. 이를 위해 한국어 학습자의 요구 조사와 현장 교수법 사례를 수집하고 구축한 음성 코퍼스를 바탕으로 범언어권 한국어 학습자의 발음 오류를 유형별로 분류하여 분석한다. 연구의 진행에 앞서, 한국어 발음 진단 및 평가 시스템의 실질적 수요자인 한국어 학습자를 연구의 목적에 맞게 확보하여 유형별로 적절하게 참여시키는 것은 매우 중요하다. 특히 학습자의 다양한 상황을 아는 현장 교육자를 통해 학습자가 연구에 참여해야 하는 의의를 잘 설명함으로써 진정성 있는 자료를 구축하는 바탕을 마련해야 한다.

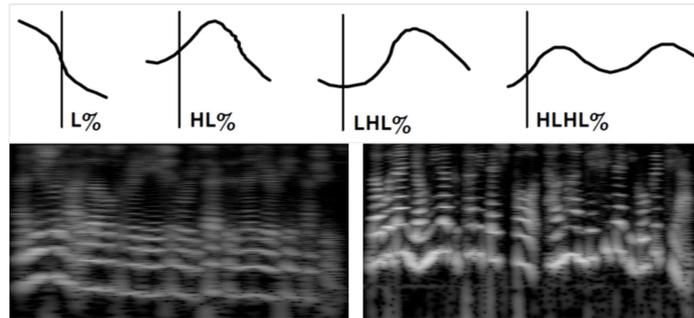
또한 발음 오류 개선 교수 방안을 실제 현장에 적용하는 연구와 실험 수업을 진행하며 의공학 분야의 연구진에게 피드백을 제공한다. 무엇보다도 한국어 학습자를 대상으로 한 발음 교육 목표와 내용에 따라 그간의 현장에서 꾸준히 드러난 교육적 난제를 해결하는 연구자의 시각이 포함된다. 연구의 전 과정에서 한국어교육 분야에서는 구축한 발음 오류의 언어권별, 숙달도 간, 남녀 간 동질성을 검증하고, 오류 유형의 원인과 수정 방안을 기 연구 성과와 실험음성학적 연구 결과를 기반으로 도출한다.

셋째, 의공학 분야에서는 음성 인식 및 조음 생체 신호를 활용한 하이브리드형 한국어 학습자의 발음 오류 원인 자동 분석 알고리즘을 개발하고 이를 바탕으로 범언어적 한국어 학습자를 대상으로 한, 발음 진단 및 평가 시스템 구현을 위한 알고리즘을 개발할 것이다. 이를 위해 본 연구는 Wei Qun, 이미향, 박희준, 안미애(2018)의 연구 선례를 활용한다. 먼저 음성 인식 기술을 활용하여, 한국어 학습자 발음 오류 코퍼스 구축을 위한 모바일 애플리케이션과 수집한 음성 데이터베이스의 대량 분석 알고리즘이 필요하다. 기존 음성 인식 기술에 기반한 딥 러닝 음성 인식 시스템은 부정확한 발음까지 예측하고 인식하는 데 한계가 있어 정확한 발음을 위한 ‘학습’과 ‘훈련’ 목적에 바로 활용하기 어렵다.

이러한 점을 극복하기 위해서 본 연구는 두 가지 알고리즘이 필요할 것으로 본다. 먼저 발음의 정확도는 하이브리드형

인공지능 음성 인식 알고리즘을 개발하여 평가하고자 한다. 하이브리드형 알고리즘은 학습자가 부정확한 발음을 할 경우, ‘얼마나 부정확한가’에 대해 진단하고, 이에 대한 평가를 할 수 있다는 장점이 있다. 또한 조음 생체 신호 데이터는 학습자의 조음 기관의 움직임의 정도를 평가할 수 있어 좀 더 정확한 발음 진단 및 평가가 가능하다.

이어 억양의 패턴화 및 원어민 대비 한국어 학습자의 억양 패턴 실현을 평가는 아래 그림과 같이 F0 곡선을 추출하여 발음 에너지 분석 전용 신경 회로망(Neural Networks)을 활용한, 억양 분석 알고리즘을 개발해 활용할 수 있다.

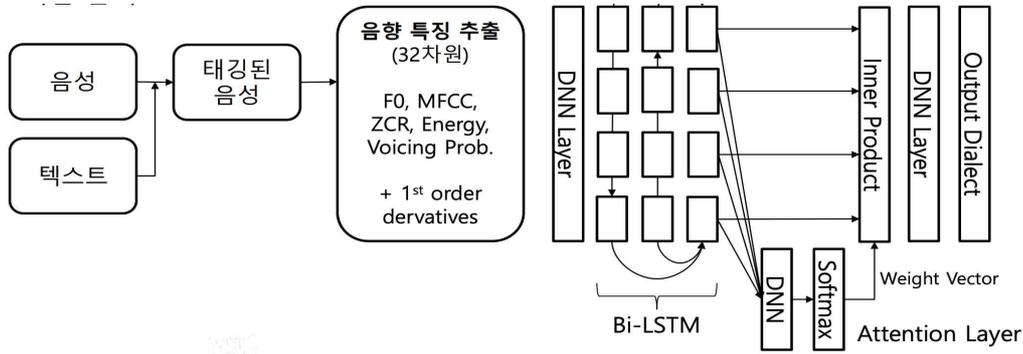


[그림 2] 추출된 억양 패턴 예시

진단 및 평가를 진행하기 위해 본 연구팀은 억양 분석을 위한 알고리즘을 두 단계로 나누어 개발한다. 정확한 억양을 찾기 위한 훈련 단계와 실제 분석 및 평가를 위한 대화 단계가 그것이다. 이 두 단계 모두 MFSC를 통해 F0값을 추출할 것이나, 억양 훈련 부분은 발음 분석 전용 신경 회로망을 활용해 데이터를 분류하고, 억양 훈련을 진행하고자 한다. 1차 년도에 의공학 분야에서는 한국인 원어민을 대상으로 조음 생체 신호를 수집한다. 한국어 학습자의 억양 오류를 대비할 한국인 화자의 조음 생체 신호의 수집은 목 근육에 Ag-AgCL 전극 네 개를 부착해 근전도 검사를 함께 시행할 예정이다. 음성 신호 수집은 SM 7B(슈어사)를 이용하고, 근전도 검사는 생체 신호 수집 전용 장비인 MP 160(Biopac, USA)를 활용할 것이다.

이를 바탕으로 의공학 분야에서는 한국어 학습자의 발음 진단과 평가 시스템을 구현하기 위한 알고리즘을 개발한다. 특히 발음 오류 원인을 정확히 분석하였는지를 발음 기관 관련 복합 생체 신호(음성, 근육의 움직임 등)를 활용하여 확인해볼 것이다. 그리고 발음 진단 및 평가 알고리즘을 도출해 이를 실제 시스템으로 구현하되, 구현 시 인공지능을 기반으로 자동 학습이 가능한 딥 러닝(SVM, NN, CNN, RNN) 기반 음성 인식 알고리즘을 활용하고자 한다.

의공학 분야의 최종 목표는 한국어 학습자 발음 오류 진단 및 평가 알고리즘을 모바일 애플리케이션으로 개발하는 것이다. 학습자와 교사가 학습 환경에서 주로 활용할 매체가 스마트폰으로 예측되므로 본 연구팀에서도 스마트폰의 하드웨어적 한계를 고려하여, 기존의 억양 분석 및 평가 알고리즘을 최적화된 결과물을 안드로이드와 애플의 IOS 스마트폰에 이 발음 진단 및 평가 애플리케이션을 탑재할 계획을 가지고 있다. 본 연구팀에서 최종적으로 개발하고자 하는 애플리케이션의 구동 과정은 [그림 3]과 같다.



[그림 3] 한국어 발음 및 억양 진단 평가 시스템의 구현도

또한 스마트폰을 활용한 애플리케이션이므로 음성 자료의 수집과 학습, 훈련의 용이성을 높이기 위해 구글 클라우드 컴퓨팅 플랫폼(Google Cloud Computing Platform)이나 아마존 웹 서비스(AWS)에 본 연구팀에서 개발할 모바일 애플리케이션을 탑재하여 애플리케이션의 원활한 배포와 구동을 위한 컴퓨팅 플랫폼을 구축할 계획이다.

이상의 학제 간 연구를 통해 도출할 연구 성과물은 실험음성학적 연구 방법과 한국어 교육적 관점의 분석과 해석, 그리고 의공학의 기술적 지원을 바탕으로 한 융합적 연구 성과가 될 수 있다.

#### 4. 결론 및 논의

지금까지 범언어권 한국어 학습자를 대상으로 발음 오류 말뭉치를 구축하고, 이를 통해 한국어 학습자의 발음을 진단하고 평가할 한국어 발음 평가 시스템을 개발하기 위한 학제 간 연구 방안을 모색해 보았다. 본 연구를 통해 본 연구팀은 한국어 학습자를 대상으로 한 실제 발음 자료를 음성 인식 기술(Speech Recognition)을 기반으로 수집하고, 실험음성학과 의공학적 연구 방법론을 통해 구축된 코퍼스를 분석할 계획을 세우고 있다. 이러한 연구 과정을 통해 그동안 주관적 영역으로 여겨진 발음 평가에 객관성과 보편 타당성을 높이고자 하는 것이 본 연구의 최종 목적이다. 이러한 연구 전개는 언어 간 대조를 통해 수행되어야 하는 발음교육 연구를 음소뿐만 아니라 초분절음소의 영역으로 확장하여, 범언어적 발음 지도에 두루 활용할 수 있는 방안을 제공하는 데 기여할 새로운 방향이 되었다고 본다.

그러나 본 연구에서 구축할 음성 코퍼스의 규모와 양 그리고 질적인 수준은 짧은 기간 동안 개인 또는 소규모 연구팀이 수행하기에는 무리가 있다는 한계가 있다. 이 연구에서 구축하고자 하는 음성 데이터는 연구 자료를 수집 및 가공하고 레이블링하는 시간을 추산하면 거의 1년 이상이 걸릴 것으로 보이기 때문이다. 이는 40명을 대상으로 한 윤원희 외(2015)의 코퍼스가 3년을 소모한 것을 고려해 볼 때 적절한 수준의 예측이다. 또한 대규모 다년 연구로 이석재 외(2004)에서 구축한 한국어 학습자 음성 코퍼스가 완전 레이블링 코퍼스<sup>4)</sup>가 아니라는 점도 음성 코퍼스 구축이 쉽지 않다는 것을 보여준다. 특히 이 연구에서 구축할 음성 코퍼스 중 자연 발화 부분은 윤원희 외(2015)보다는 양이 적으나 낭독 발화와 자연 발화를 망라하므로 단년 과제로는 마무리하기 어려운 작업 양이다.

또한 음성 코퍼스 구축에 이어 진행할 음성 데이터의 발음 오류 분석과 발음 진단 및 평가 시스템 개발에 필요한 시간의 확보가 필요하다는 점도 이 연구가 계획한 연구 규모와 방법의 한계가 될 수 있다. 음성 데이터 수집과 발음 오류 분

4) 이석재 외(2004)에서는 정밀 레이블링을 당초의 연구 목표로 한 것이다. 그러나 최종 결과물에서는 연구 목표를 수정하여 음소 차원의 몇 가지 기준에 대한 오류 분석과 자모음, 억양을 대상으로 한 주요 언어권별 음성 음향적인 연구 결과를 제시하고 있다.

석을 하는 시기는 발음 평가 시스템을 개발하는 것과 겹칠 수 없다. 전자에 의존하여 후자를 설계하기 때문이다. 이러한 측면에서 시간이 많이 걸리고 어려움도 큰 학제 간 연구를 전폭적으로 지지할 사회적 합의도 필요해 보인다.

한국어 발음교육에서 학습자의 발음 진단과 평가 시스템 구축은 학습 결과의 객관화를 높일 수 있을 뿐만 아니라 연구의 결과물을 국내외 한국어 교육 현장에서 활용 가능성 높은 스마트러닝 교육 도구로써 활용할 수 있다. 한국어 학습자의 양적·질적 수요 변화를 충족하는 자가 학습용 도구로써 자가 학습의 기회 확장에도 기여할 수 있을 것이다. 모바일 러닝은 학습 공간을 확장하고 자기 주도적 학습에 용이하여 정규 학습 과정의 부족한 교육 내용을 보완해 줄 수 있기 때문이다. 무엇보다도 범언어적 한국어 학습자라는 연구 대상 또한 다양한 언어권의 이주민이 유입된 한국의 사회적인 상황에 적절한 연구 대상이다. 이러한 점을 고려할 때, ‘범언어적 한국어 학습자’를 대상으로 한 대량의 발음 오류 음성 코퍼스 구축과 이를 바탕으로 한 발음 진단 및 평가 시스템 애플리케이션의 개발은 반드시 수행되어야 할 연구이다.

#### ■ 참고문헌 ■

- 권경근(2016), 한국어 발음 교육의 요소와 범위, 『한국민족문화』 58, 부산대학교 한국민족문화연구소, 115-133.  
 권성미(2014), 발음 교육에 대한 한국어 교사의 인식 연구, 『국어교육』 146, 한국어교육학회, 505-529.  
 김미란, 강현주, 노주현(2014), 학습자의 발화 속도 변이 연구, 『한국어학』 63, 한국어학회, 103-132.

- 김상수, 송향근(2006), 한국어 발음 교육 연구 동향 분석. 『한국어학』 33, 155-183.
- 김서형, 장향실, 차재은, 전나영(2016), 외국인 한국어 학습자를 위한 단모음의 발음 설명, 『韓民族語文學』 72, 한민족어문학회, 93-122.
- 김선정(2014), 한국어 발음 습득과 한국어교육, 『사단법인 한국언어학회 2014년 여름학술대회 발표논문집』, 한국언어학회, 9-18.
- 김선정, 육효창(2010), 다문화 가정 한국어 학습자를 위한 한국어 발음 교육- 베트남 여성결혼이민자를 중심으로. 『나라사랑』 119, 131-160.
- 김현주(2008), 발음 오류에 대한 원인 분석과 교육 방법 모색- 한국인 프랑스어학습자를 대상으로, 『프랑스어문교육』 28, 한국프랑스어문교육학회, 71-93.
- 김형복(2009), 한국어 Phonics를 통한 발음 규칙 교수-학습 방안, 『코기토』 66, 부산대학교 인문학 연구소, 277-298.
- 박기영(2010), 한국어 음운론과 한국어 발음 교육의 상관성에 대한 일고찰- 자모 교육과 음운 변동 교육을 중심으로, 『語文論集』 43, 중앙어문학회, 7-30.
- 박정은, 이주희(2008), 외국인을 위한 한국어 발음 교재의 분석과 개선방향 연구. 『국어국문학』 150, 국어국문학회, 565-602.
- 박창원(2005), 『한국어 학습자용 말뭉치의 구축과 활용』 국립국어원.
- 백경미, 김태경(2014), 제2언어로서의 한국어 습득 과정에 나타난 자음 탈락 현상 연구. 국『제어문』 63, 국제어문학회, 257-281.
- 백소영(2016), 한국어 교사들의 발음 교육에 대한 인식 연구, 『한국언어문학』 99, 한국언어문학회, 451-466.
- 송향근(2011), 음운 이론과 한국어 발음 교육. 『한국어학』 50, 한국어학회, 1-27.
- 안미애(2014), 인도네시아인 고급 한국어 학습자의 한국어 단모음의 실험음성학적 대조 분석, 『언어와 문화』 10-1, 한국언어문화교육학회, 177-205.
- 안미애(2015), 한국어 낭독체의 운율구 형성 및 음조 실현 양상 연구:인도네시아인 한국어 학습자를 중심으로, 『어문론총』 63, 한국문학언어학회, 9-40.
- 안미애, Wei Qun, 이미향(2016), 중국인 한국어 학습자의 한국어 발음 인식 양상 연구, 『어문론총』 70, 한국문학언어학회, 33-64.
- 안미애, Wei Qun, 이미향(2018), 한국어 발음 학습과 자가 평가를 위한 스마트폰 앱(App) 개발 및 교육적 적용 방안 연구:중국인 한국어 학습자를 대상으로, 『이중언어학』 70, 이중언어학회, 167-195.
- 안미애, 이미향(2019), 한국어 학습자의 텍스트와 담화 간 끊어 말하기에 대한 인식과 실현 양상 고찰, 『국어교육연구』 71, 국어교육학회, 135-164.
- 윤원희 외(2012~2014), 한국어 자연 발화 음성 코퍼스 구축을 위한 기초 연구, 한국연구재단.
- 윤은경(2013), The effects of perceptual training on speech production - focusing on Korean vowels, 『외국어교육연구』 27 권 2호, 한국외국어대학교 외국어교육연구소, 1-28.
- 이석재 외(2001~2003), 한국어 음성 코퍼스, 한국 연구재단 사업.
- 이석재 외(2004), 『외국어로서의 한국어 음성 코퍼스 구축과 이를 통한 외국인의 한국어 음성 음운 체계 습득 양상 연구 사업 결과 보고서』, 한국 연구 재단 사업.
- 이석재 외(2006), 외국어로서의 한국어 음성 코퍼스의 설계, 구축 및 응용 방안 연구, 『한국어(조선어)교육연구』 4, 중국 한국(조선)어 교육연구학회, 267-282.
- 이혜정(2015), 숙달도와 거주 기간에 따른 베트남인 학습자의 발음 습득 양상-자음과 음운현상을 중심으로, 『이화어문논집』 35, 이화어문학회, 143-172.
- 정명숙(2008), 한국어 학습자를 위한 전략적 발음 교육- 중국인 학습자를 중심으로, 『한국어학』 38, 345-369.
- 허용(2012), 외국인 학습자의 한국어 발음 오류에 대한 음운론적 분석-음운현상을 중심으로. 『한국학논집』 46, 계명대학교 한국학연구원, 201-232.
- Wei Qun, 이미향, 박희준, 안미애(2018), Using Formant Comparison Method to Design a Korean Pronunciation Self-Study Mobile Application System for Chinese, *Journal of Advanced Research in Dynamical and Control Systems*, Vol 10, 11 Special issue, pp.1245-1249.

## 〈토론〉 “음성 인식 기술(Speech Recognition)을 기반으로 한 한국어 학습자의 발음 진단·평가 시스템 구축 방안 연구”에 대한 토론문

권성미(부경대학교)

이미향-안미애 선생님의 연구는 한국어 학습자의 발음에 대한 진단·평가 시스템 구축 방안을 모색하신 거의 최초의 시도로 볼 수 있는 연구인 것 같습니다. 외국어 교수·학습의 여러 영역 중 발음 영역은 이론 및 가설을 통해 학습자 발음의 실재를 예측하기에 가장 용이한 분야이기도 합니다. 하지만, 원고에서 언급하셨듯 학습자에 따른 효과적인 발음 교수 내용과 방안을 개발하기 위해서는 언어권별로 일반화되는 가설적인 정보들에 대한 확인과 실증적 검증이 필요한 것 같습니다. 이러한 점에서 두 분의 연구는 이론적 측면에서 언어권별 한국어 학습자 발음 연구에 관한 선행 연구들의 이론과 가설을 검증하는 동시에, 실제적인 측면에서 효용성 있는 교수 방안을 설계할 수 있는 토대가 되는 연구로, 한국어 발음 교육에 큰 족적을 남길 것으로 기대되는 연구라고 생각합니다. 다만 연구 내용이 음성 처리 관련 기술적 측면에서 상당히 전문성을 띤 부분을 포함하고 있어, 원고에 대한 저의 이해가 완전하지 못한 상황에서 질의를 드리게 되어 송구한 마음입니다. 그저 토론자로서 몇 가지 질문을 드림으로써 선생님들의 연구에 대한 제 이해를 확인하고, 연구에 대한 선생님들의 고견을 더 듣는 기회를 갖고자 합니다.

첫 번째는 연구 범위에 대한 질문을 드리고자 합니다. 원고에 따르면, 학습자 발음 오류를 음성 말뭉치로 구축한다고 기술하고 있으신데요. 단순한 표현의 문제일 수도 있겠습니다만, ‘오류 음성 코퍼스’라는 표현이 저는 조금 어색한 것 같습니다. ‘오류’를 논하기 위해서는 정확도 차원에서 다층의 정도성을 가진 학습자 언어 자료에 대한 비교 관찰이 필요할 것 같습니다. 그러한 면에서 제가 보기에는 선생님들의 연구는 ‘오류’ 관찰을 위한 ‘오류 음성 코퍼스’보다는 ‘오류’ 관찰을 위한 ‘학습자 발음(혹은 음성) 코퍼스’를 구축하시려는 것으로 보입니다.

두 번째 질문은 발음의 진단·평가 시스템 설계에 관한 것입니다. 학회 당일 발표 중에 더 구체화된 내용을 들을 기회가 있기를 기대합니다만, 일단은 질문을 드려 보고자 합니다. 근전도 검사와 같은 생체 신호가 발음의 정확도를 판정하는 데 어떠한 정보를 제공하는지 등이 궁금합니다. 검사 도구를 목 근육에 부착하는 것으로 보아서는 학습자가 조음하는 자음의 발성 유형(phonation type)을 관찰할 수 있지 않을까 상상을 해 볼 뿐입니다. 근전도 검사 외에도 또 어떠한 검사들을 통해 의공학적 기술을 빌어 조음상의 특징이나 발음의 정확도를 측정할 수 있을지, 또 그러한 알고리즘은 어떻게 구현되는지 설명을 더 들어보는 기회가 있었으면 좋겠습니다.

세 번째 질문은 코퍼스 공유에 관한 것입니다. 컴퓨팅 플랫폼을 구축하여 연구의 결과물을 배포하실 예정이신 것으로 보입니다. 원 음성 자료 데이터 베이스를 공유받을 수 있는 것만으로도 학습자 발음 연구자와 교육자에게는 큰 도움이 될 것입니다. 그런데 나아가서 전문가에 의해 처리된 코퍼스, 그러니까 segmentation을 비롯한 억양 단위나 음운 현상 등에 대한 정보가 처리된 코퍼스를 연구자나 교육자들이 접할 수 있다면, 연구는 물론 교수·학습 현장에서 효용성 있는 교수 방안 개발에 참으로 의미있는 기여가 될 것으로 기대합니다. 원 코퍼스와 처리된 코퍼스의 단계별 완성 시기와 단계별 공유 계획 등에 대한 말씀을 더 들어 보고 싶습니다.

## 선별(Screening) 목적의 인터넷 기반 한국어 진단평가 과제 개발 기초 연구

한운정(경희대학교)

### 1. 서론

대학 입학 직후 학문 목적 학습자들을 위한 한국어 진단평가의 필요성은 꾸준히 제기되어 왔다. 국가통계포털(KOSIS)에서 한국 유학 생활의 어려움을 조사한 결과 2020년 12월 기준으로 학업 관련 문제, 경제적 문제 등 12개 항목 중 언어 문제가 61.3%로 월등히 높았다. 다른 항목이 10% 미만임을 볼 때 매우 높은 비율이라고 할 수 있다. 한국어능력시험(TOPIK)이 대학 입학의 당락을 결정하는 고부담 평가로서 계속 활용되어 오고 있으나, 개별 학습자들의 언어 능력을 구체적으로 진단하여 강점과 약점을 측정하고 학업 수행에 도움이 되는 진단적 피드백을 제공하는 ‘진단 도구’의 역할은 할 수 없다. Fox 외(2016:44)에서는 영어가 모국어인 학생이 아니라면 숙달도 평가 점수만으로 영어권 대학에서 성공적인 학문적 참여를 위한 준비가 되었는지를 가늠하기는 어렵다고 언급하고 있다. 또한 학업 수행을 계속 유지하기 위해 입학 후 첫 해의 중요성을 강조하며 이 시기에 언어 능력을 진단하는 것이 필요하다고 언급한다. 이러한 맥락으로 한국어교육 분야에서도 대학이나 대학원 입학 시기의 진단평가가 필요하다고 본다.

따라서 본 연구에서는 진단평가 과제 개발을 위한 기초 연구로서 쟁점을 논의해 보고자 한다. 특히 본고에서 중점적으로 다룰 부분은 인터넷 기반으로 선행될 수 있는 ‘선별(Screening)’ 목적의 평가 과제이다. Alderson(2005:8-9)에서는 진단평가가 유용하려면 수행에 대한 결과 프로파일이 매우 중요하고 평가를 구성하는 다양한 요소들을 아우르는 시험 수행에 대한 구체적인 정보가 요구된다고 언급한다. 또 Fox 외(2016:50)에서는 진단평가의 쟁점 요소를 크게 ‘평가의 질, 후속 교육의 지원, 전달과 마케팅’으로 제시하고 요소 간의 균형을 유지하는 것이 필요하다고 언급하기도 하였다. 즉 진단평가는 학문 목적 한국어 상황에서 매우 필요하지만 평가의 개발과 시행에 매우 복잡한 요소들이 영향을 미쳐 어려움이 존재한다고 할 수 있다. 따라서 본격적인 진단평가를 치르기 위한 선별 목적의 간단한 평가 과제 개발이 먼저 논의되고 연구되어야 한다고 본다. 이에 본 기초 연구에서는 선별 시험의 특징과 과제 개발과 관련한 쟁점을 종합적으로 다루고자 한다.

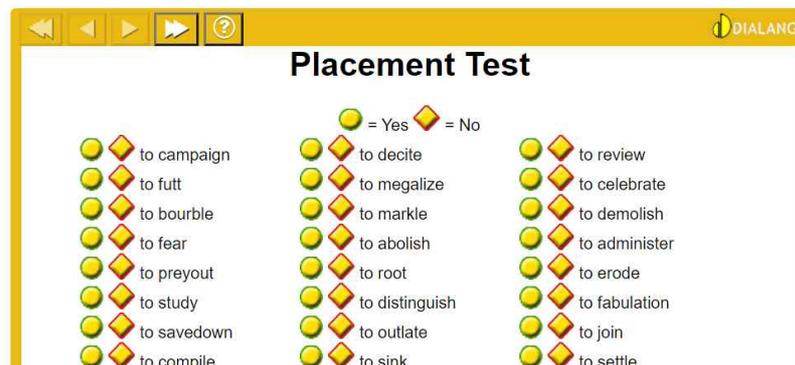
### 2. 학문 목적 언어 진단평가의 선별 시험

먼저 <표 1>과 같이 해외 주요 진단평가의 전반적인 내용을 살펴보았다.

〈표 1〉 관련 평가의 개요(한윤정, 2021:60-61의 일부)

시험 이름	DELNA (Diagnostic English Language Needs Assessment)	DELA (Diagnostic English Language Assessment)	PAAL (Post-Entry Assessment of Academic Language)	DIALANG <sup>1)</sup> (DIAGnostic LANGuage)
시험 개요 및 목적	영어권 대학에서의 학업 수행을 위한 신입생의 준비 상태를 평가하고 이후의 언어적 필요를 확인하기 위한 시험. 성공적인 학업 수행을 위한 대학 수준의 영어 능력을 진단함.	대학 수준의 학습 과정에 필요한 읽기, 쓰기, 듣기 등 학문적 언어 능력을 평가함. 성공적인 학업 수행을 위한 대학 수준의 영어 능력을 진단함.	학문적인 성공과 관련된 언어 능력을 평가함. 학업적인 언어 지원의 필요 여부를 판단하기 위한 짧은 온라인 시험임.	온라인 언어 진단시험 시스템으로 14개의 유럽 언어를 대상으로 언어 능력을 진단하는 평가임.
주관 기관	뉴질랜드 Auckland대학교 DELNA office	호주 Melbourne대학교 LTRC(Language Testing Research Centre)	호주 Melbourne대학교 LTRC(Language Testing Research Centre)	Alderson (영국 Lancaster 대학교) 외 연구팀 (SOCRATES programme과 유럽 25개 기관 지원)
시험 방식	인터넷 기반 & 지필	지필	인터넷 기반	인터넷 적응형
영역 및 소요 시간	<ul style="list-style-type: none"> <li>• Screening(30분) 어휘/빠르게 읽기</li> <li>• Diagnosis(2시간) 읽기(45-55분) 듣기(30분) 학술적 쓰기(30분)</li> </ul>	<ul style="list-style-type: none"> <li>• 읽기(45분)</li> <li>• 쓰기(30분)</li> <li>• 듣기(30분)</li> </ul>	<ul style="list-style-type: none"> <li>• 배경 질문(5분)</li> <li>• 텍스트 완성(15분)</li> <li>• 빠르게 읽기(10분)</li> <li>• 쓰기(30분)</li> </ul>	<ul style="list-style-type: none"> <li>• 배치평가</li> <li>• 읽기</li> <li>• 듣기</li> <li>• 쓰기</li> <li>• 어휘</li> <li>• 구조(문법)</li> </ul>

먼저 DELNA, DELA, PAAL은 뉴질랜드, 호주에서 시행되고 있는 입학 후 언어 평가(Post-Entry/Admission Language Assessment, PELA)이다. 세부 시행 방식은 차이가 있으나 컴퓨터, 인터넷 평가가 대부분임을 확인할 수 있으며 평가 후 진단 피드백이나 후속 교육 지원에 대한 안내가 전달된다. DELNA와 DELA, PAAL은 대학 입학 초에 학문적 목적으로 언어 능력을 진단하여 교육기관 차원에서 학업 수행을 위한 언어적 지원을 적극적으로 지원하고자 하기 위한 평가다. DIALANG은 학문적 목적으로 공식 활용되는 것은 아니나 언어 진단평가 중 가장 주목받는 대표적인 시험이다. 특징은 수험자의 대략적인 어휘 크기를 측정하는 Yes/No 형식의 배치시험(placement test)이 있고 그 결과에 따라 본 시험의 수준이 결정되어 문제가 출제된다. 〈그림 1〉과 같이 실재하는 단어와 가상의 단어를 75개 나열하여 존재한다고 생각하는 어휘에 ‘yes’를 표시하는 방식이다.



〈그림 1〉 DIALANG의 배치시험 화면 일부

1) <https://dialangweb.lancaster.ac.uk/>를 참고하였다.

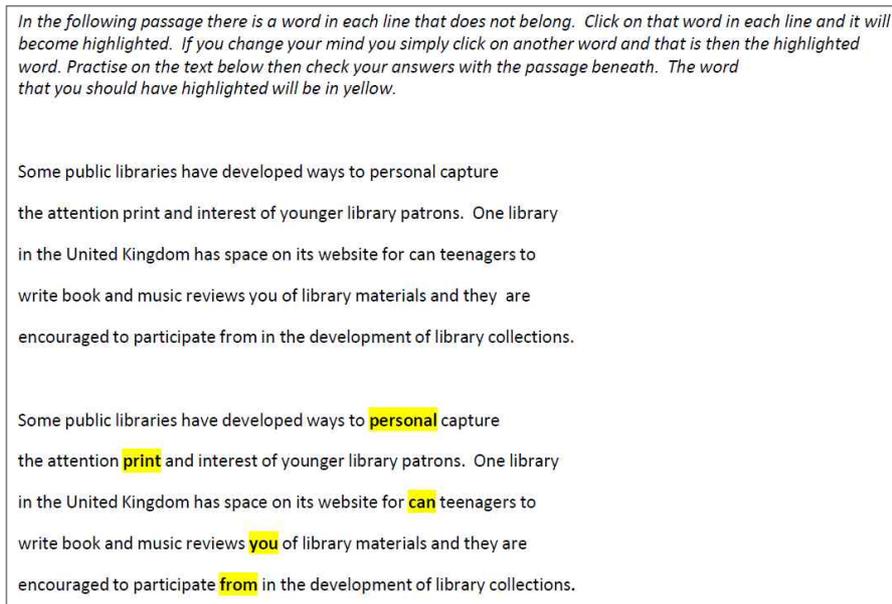
Yes/No 형식은 어휘 규모를 간단히 측정하는 체크리스트(checklist) 검사 도구 일종으로, 보통 아는 어휘에 체크하는 단순한 형식에 시험이 쉽고 시간이 많이 소요되지 않는다는 장점을 가진다. 하지만 정말 자신이 아는 단어에만 정확하게 체크하지 않으면 그 결과가 달라질 수 있다. 즉, 같은 수준의 학생이라도 시험수행에 따라 점수가 달라질 수 있다는 것이다. 이렇게 수험자가 자신의 어휘 능력을 과대평가할 수 있어 어휘 목록의 25~33% 정도 그럴듯한 가짜 단어(plausible nonword)를 포함한다. 이 형식이 학문 목적 진단평가에 활용된 예는 오만에서 대학 입학생을 대상으로 한 TYN 시험(Timed Yes/No Vocabulary test)이다. TYN 시험은 RocheHarrington(2013) 이후 Roche를 중심으로 지속적인 연구를 통해 개발·연구되었으며, 오만(Oman)의 두 대학에서 사용되었으나 확인 결과<sup>2)</sup> 현재는 공식적으로 시행되지 않고 있다. 이 시험은 어휘 지식의 깊이와 속도를 측정하며 수험자가 목표 단어의 세트를 아는지 판단하는 컴퓨터 기반의 Screening 방식으로 구성되어 있다. Roche 외(2016:174)에서는 기존의 대학 입학 시 치러지는 언어 평가가 시간과 비용 측면에서 비효율적임을 지적하며 TYN 시험이 비용 대비 효율적인 대안이 될 수 있다고 주장한다. 이런 측정 방식은 최근의 고부담 숙달도 평가에서는 수용되지 않지만, 배치 혹은 진단의 목적을 가진 PELA에서는 포함될 수 있다. Read(2016:12)는 잘 구성된 어휘 시험은 높은 신뢰성을 가진 효율적인 평가 도구이며, 독해 능력과 일반적인 언어 숙달도의 좋은 예측 변수가 된다는 Alderson(2005:206)의 연구 결과를 바탕으로 Roche 외(2016)의 주장이 유의미하다고 밝힌다. 또한 이러한 평가 도구들은 PELA의 진단적인 목적에 잘 부합하여 숙달도 평가에서 선호되는 의사소통적 언어 사용 과제들과는 분명 차별성을 갖는다고 언급하였다.

DIALANG의 배치시험과 비슷한 맥락으로 뉴질랜드 오클랜드대학교의 DELNA에서는 진단(diagnosis) 시험을 치를 대상을 파악하기 위해 선별(screening) 시험이 먼저 진행된다. 대학 내 모든 신입생은 입학 후 2주 이내 온라인으로 선별 시험에 응시해야 하며, 심층 진단이 필요하다는 결과가 나온 학생들은 듣기, 읽기, 쓰기 영역으로 구성된 지필 진단 시험에 응시해야 한다. 이처럼 응시자를 선별할 수 있는 인터넷 기반 시험을 먼저 치르도록 하고 대상자에 한해서만 지필 시험에 응시하도록 하는 것은 경제적인 면에서나 효율적인 면에서 매우 유용하다고 할 수 있다. 물론 모든 절차를 인터넷 기반으로 구축하는 것이 이상적일 수 있으나 그렇게 되더라도 선별 과제를 통해 수험자의 대략적인 수준을 파악하여 그 이후의 조치를 결정하는 방식은 유용하다 할 수 있다. DELNA의 선별 시험은 어휘 과제(Vocabulary task)와 빨리 읽기 과제(Timed reading task)로 구성되어 있다. 어휘 과제는 <그림 2>와 같이 단어나 구를 제시하고 의미를 찾게 하거나 의미를 보고 단어나 구를 찾게 하는 형식으로 27문항을 준비시간을 제외하고 7분 안에 완료해야 한다. 또한 빨리 읽기 과제는 <그림 3>과 같이 73줄의 텍스트를 최대한 빨리 읽고 필요 없는 단어를 삭제하는 cloze-elide 형식으로, 준비시간을 제외하고 10분 안에 완료해야 한다. 이 과제는 어휘의 질적 지식, 통사적 지식, 의미 지식, 사용 양상을 간단하게 측정하기 위해 제작되었다고 소개된다.

2) 2020년 6월 연구진 중 한 명인 Sohar 대학 언어학부 Yogesh Kumar Sinha 교수님에게 메일을 보내 개인적으로 확인하였다.



〈그림 2〉 DELNA 선별 시험의 어휘 과제 예시(DELNA Handbook)



〈그림 3〉 DELNA 선별 시험의 빨리 읽기 과제 예시(DELNA Handbook)

두 가지 과제로 구성된 선별 시험에서 대학의 학업 수행에 있어 언어 능력이 충분하다(sufficient)는 결과가 나오면 특별한 후속 조치가 없으나, 적합(adequate)으로 판정되면 언어 능력을 향상할 수 있도록 언어 강화 옵션이 안내된다. 그리고 학업 시작 시 언어적 지원이 필요하다(you may need language support)는 판정이 나오면 DELNA 진단평가를 예약해 구체적인 심층 진단을 받도록 안내된다.

한편 호주 멜버른대학교의 PAAL은 짧은 시간 동안 온라인으로 학문적 언어 지원의 필요 여부를 판단하기 위한 목적으로 시행되고 있다. 같은 대학에서 DELA<sup>3)</sup>가 학문적 언어 능력 진단의 목적을 가진 평가로 긴 시간 동안 본격적으로 시행되고 있음을 보았을 때 PAAL은 선별의 목적에 맞추어 시행되고 있음을 알 수 있다<sup>4)</sup>. 제한된 시간 내 과제를 수행해야 하며 시험 결과 프로파일에 언어 지원 옵션을 지원받을 수 있는지의 여부와 담당자 연락처 등이 제시된다. 짧게 이루어지는 시험이기는 하지만 선별 시험으로서 어휘, 읽기, 쓰기와 같이 여러 영역을 함께 다루고 있다는 점이 특징이다. PAAL은 간단한 배경 질문(Background questionnaire) 후 텍스트 완성(Text completion)과 빠르게 읽기(Speed reading)

3) 호주 멜버른대학교의 DELA는 뉴질랜드 오클랜드대학교의 DELNA와 함께 연구개발되어 시행되는 진단평가이므로 선별 시험을 제외한 나머지 영역과 과제 구성은 모두 동일해 〈표 1〉에 포함하지 않았다.

4) 현재 PAAL의 시험 명칭으로 Academic English Screening Test(AEST)도 함께 사용되고 있다. 공식 홈페이지에도 AEST/PAAL이라고 병기되어 있으며 LTRC 담당자에게 동일한 시험임을 확인했다.

과제가 있다. 두 과제 모두 학술적 문어적 영어의 이해를 평가하며 ‘텍스트 완성’의 경우 산출까지도 평가 내용으로 다룬다. ‘텍스트 완성’은 짧은 3개의 텍스트가, ‘빠르게 읽기’는 1000단어 정도의 좀 더 긴 하나의 텍스트가 학문적인 내용, 구성, 문체로 구성되어 있다. 학문적인 내용을 다루나 특정 분야의 전문지식을 요구하지 않으며 전자는 하나의 텍스트에 5분, 후자는 10분의 제한 시간이 있다. 쓰기 과제는 일부 필요한 상황에서 응시자에게 더 많은 진단 정보를 제공하기 위해 추가 과제로 구성되어 있으며 DELA의 쓰기 과제와 동일하다. 제시된 몇 개의 문장을 읽고 특정 질문에 대한 300단어 정도의 짧은 논설문 쓰기 과제다. 쓰기를 제외한 두 가지 과제를 통해 학생들은 능숙함(Proficient), 경계에 있음(Borderline), 위험 수준에 있음(At Risk)과 같은 판정을 통해 후속 언어 지원에 대한 선별이 이루어진다. ‘경계에 있음’과 ‘위험 수준에 있음’으로 판정된 응시자는 학업 수행을 위해 추가 언어 지원이 필요하거나 강하게 권유되며 이에 대한 간단한 안내 사항을 제공받는다. ‘빠르게 읽기’는 <그림 3>의 형식과 동일하며, ‘텍스트 완성’은 전형적인 cloze test의 형식으로 아래 <그림 4>와 같다.

In this sample text completion task, parts of some words are missing. The letters needed to complete these words correctly are shown in the **Answers** below.

Early public buildings in Australia were constructed around the importance of influencing community and civic identity. There w\_\_\_ a senti\_\_\_ attachment t\_\_\_ the id\_\_\_ of pub\_\_\_ space wi\_\_\_ a ci\_\_\_ square rin\_\_\_ by ci\_\_\_ buildings. I\_\_\_ the foun\_\_\_ of t\_\_\_ first buil\_\_\_ in Aust\_\_\_, a dua\_\_\_ of appro\_\_\_ existed: th\_\_\_ which domi\_\_\_ the land\_\_\_ and th\_\_\_ designed t\_\_\_ blend i\_\_\_ . Early resid\_\_\_ architecture w\_\_\_ a resp\_\_\_ to the Australian landscape and the climate with its unique flora and fauna and intense sunlight. Early buildings needed to respond to these discrete climatic elements.

<그림 4> PAAL의 텍스트 완성 과제 예시(PAAL Handbook)

이 외에도 학문 목적 학습자 또는 일반 성인 학습자 대상은 아니지만 초·중등 학습자를 대상으로 한 대화형 일본어 평가(DLA, Dialogic Language Assessment), 미국 WIDA Screener, ELPAC(English Language Proficiency Assessments for California)와 같은 진단평가 예<sup>5)</sup>를 보아도 단계적 진행이 이루어지고 있음을 알 수 있다. DLA의 경우 어휘력 확인 후 말하기 과제가 이어지고 그 수준에 따라 시험이 종료될 수도 있고 각 영역의 과제를 추가로 진행할 수도 있다. WIDA Screener의 경우 적응형 시험(adaptive test)의 방식으로 구성되어 시작 과제(Entry Task)와 간단한 말하기 과제를 통해 이후 진행될 영역별 과제의 수준을 결정한다. ELPAC의 경우에도 최초 평가(Initial ELPAC)와 총괄 평가(Summative ELPAC)로 구성되어 영어 학습의 도움을 필요로 하는지 아닌지를 선별하는 목적의 단계를 거쳐 본격적인 진단을 위한 과제가 진행된다.

숙달도 평가나 성취도 평가 등 다른 목적의 평가에 비해 연구나 시행 자료가 충분하지는 않지만, 지금까지 살펴본 내용을 바탕으로 선별 시험의 개념을 명확히 하고 학문 목적 한국어교육 맥락에서 선별 목적의 과제로서 어떤 유형이 가능한지를 다음 장에서 살펴보려고 한다.

5) 이준호 외(2020)의 평가 조사 자료 및 분석 내용을 참고하였다.

### 3. 선별 목적 한국어 평가의 특징 및 구성 논의

학문 목적 한국어교육의 맥락에서 대학이나 대학원 입학의 당락을 결정하는 데 큰 영향을 미치는 TOPIK과 같은 속달도 평가가 학업 수행을 위한 언어 능력 진단의 목적으로 사용되는 것에 큰 한계가 있음은 계속 지적되어 왔다(신인철 외, 2018; 정연숙, 2018; 이선영·나윤주, 2018 등). 필요성이 매우 높은 상황에서 학문 목적 언어 진단평가를 다룬 한국어교육 분야의 연구는 어휘 영역을 중심으로 이효경(2015), 한운정(2021)이 전부이므로 앞으로 계속 연구가 이루어져야 하는 분야라고 할 수 있다. 본고에서는 개발의 실효성과 시행의 경제성을 염두에 두고 진단과 더불어 선별 목적에 중점을 둔 선별 시험에 대한 논의를 진행하고 있다. 먼저 학문 목적 한국어 진단평가에서 선별과 진단의 목적을 세분화하여 평가를 개발하고 시행한다고 할 때 선별 시험은 다음과 같은 목적과 특성을 가진 것으로 정리할 수 있다.

#### 〈학문 목적 한국어 선별 시험의 목적과 특징〉

- 목적: 대학이나 대학원 등에서 한국어로 학업을 수행하는 학습자를 대상으로 교육기관 차원의 언어적 지원이나 교육이 필요한 학습자를 선별하는 시험.
- 특징:
  - 영역별, 하위 요소별 구체적인 진단 피드백을 제공하는 진단 시험과 차이가 있다.
  - 내용적으로 학업 수행에 기본이 되는 학술기본어휘가 충분히 평가에 반영되어야 한다.
  - 제한된 시간 내에 빠르게 과제를 해결하도록 하는 속도 평가(speed test)가 적절하다.
  - 대학 및 대학원 신입생의 접근성이 중요하므로 인터넷 기반으로 시행되도록 하며 응시 시간이 너무 길어지지 않도록 문항 수 조절이 필요하다.
  - 과제 구성 시 학술 텍스트를 활용하나 특정 전공 지식을 요구하지는 않는다.
  - 응시자가 큰 부담감 없이 응시할 수 있도록 저부담 평가로 개발되어야 하며 시험 결과로 인해 성적이나 과목 선정 등에 불이익이 없도록 해야 한다.
  - 채점이나 결과 보고 체계가 간편해야 한다.

이러한 내용을 바탕으로 이제 학문 목적 선별 시험에서 ‘무엇을’, ‘어떻게’ 평가해야 할지에 대한 구체적인 논의가 필요하다. 특히 선별 시험에서 ‘학문 목적 한국어능력’의 어느 부분을 어떻게 다루어야 하는지 고려해야 한다. 먼저 학술적 내용과 형식을 갖춘 학술 텍스트를 평가에 활용하되 학술기본어휘<sup>6)</sup>가 충분히 반영되어야 한다. 앞 장에서 살펴본 선별 시험에서 ‘어휘’ 영역이 포함되어 학술적 어휘 능력을 평가하고자 한 의도 역시 학업 수행에 있어 그 중요도가 높기 때문이다. Saville-Troike(1984:199, 216)는 학습자의 학문적 성취를 위한 언어 능력의 가장 중요한 영역이 어휘라고 강조하였고, Jordan(1997:149)은 학문 목적 학습자들이 늘 어휘에 대한 부족과 어려움을 느끼며 어휘를 언어 실력 향상의 척도로 여기는 현상이 있다고 지적하였다. 또한 한국어 학술어 목록의 학술 텍스트 점유율에 관한 연구인 윤새롬·김영규(2017)에서 경영학 전공 개론서 2권의 텍스트에 나타난 학술 어휘 유형(type) 텍스트 점유율이 25.6%<sup>7)</sup> 정도였다. 이 연구에서는 학술 어휘가 학술 텍스트에서 높은 텍스트 점유율을 나타낼 뿐만 아니라 TOPIK 초·중급 어휘만 알고 있을 때 느낄 수 있는 어휘 공백을 효율적으로 보충해 주고 있음을 확인하였다(윤새롬·김영규, 2017:208). 이러한 내용은 선별 시험에서

6) ‘학술기본어휘(academic vocabulary)’는 ‘학술전문어휘(technical vocabulary)’와 대비되는 학문 목적 한국어 교육용 어휘군을 말하며 ‘사고도구어’나 ‘교양교과어휘’ 등으로 불리기도 한다.

7) 해당 연구에서는 ‘경영학 개론’과 ‘마케팅’이라는 개론서를 분석 대상으로 삼아 각각 어휘 유형 기준으로 26.42%, 24.71%의 텍스트 점유율이 도출되었으며, 이 값의 평균값을 본 연구에서 제시한 것이다. 두 권의 다소 제한된 말뭉치에서 도출된 값이기는 하지만, 대략적인 현황을 살펴볼 수 있어 여기 제시하였다.

우선으로 학술기본어휘를 다루어야 한다는 본 연구의 주장에 뒷받침이 될 수 있다. 학술기본어휘 과제 구성에 대해서는 다음 장에서 좀 더 다루고자 한다.

한편 선별 시험에서 활용할 학술 텍스트의 범주는 개론서나 대학(대학원) 교재, 보고서 등에서 자주 볼 수 있는 정보성 글을 포함해 학술 논문도 포함될 수 있다. 대학과 대학원의 차별성을 둔다면 대학 신입생을 대상으로 하는 경우 좀 더 정보성 글이 많이 포함되도록 하고, 대학원 신입생을 대상으로 하는 경우 학문과 연구를 수행을 염두에 두고 논문 형식을 더 많이 포함되도록 할 수 있다. 또한 선별 시험은 접근성과 채점 및 결과 보고의 간편성이 매우 중요하기 때문에 채점 절차가 복잡하고 주관이 개입될 수 있는 말하기나 쓰기 영역은 제외되거나 필요한 경우 추가로 제시되는 것이 적절하다<sup>8)</sup>. 또한 같은 이유로 학문 목적의 듣기 영역을 선별의 목적으로 다루기에는 응시 소요 시간이 길어진다는 단점이 있어 제외되는 것이 바람직하다<sup>9)</sup>. 그렇다면 선별 시험은 학술기본어휘를 중심으로 어휘, 읽기 과제를 중점으로 개발하되 학술 텍스트를 활용해 다양한 과제 유형으로 구성하는 것이 타당하다고 본다.

다음으로 기존 영어권 선별 시험에서 활용한 빨리 읽기(불필요한 단어 삭제하기)와 텍스트 완성하기(빈칸 채우기)가 한국어 평가 상황에서 적절한지를 확인할 필요가 있다. <그림 3>과 <그림 4>에서 제시한 영어 예시 문항을 보면 공통적으로 관사나 조동사, 시제 등 문법 요소를 주로 다루고 어휘 일부를 채워 쓰게 하거나 맥락상 불필요한 어휘를 삭제하도록 한다. 이 두 가지 형식을 한국어의 경우로 바꾸어 생각해 보면 내용적인 면과 형식적인 면으로 나누어 살펴볼 수 있다. 먼저 내용적인 면에서 문법을 다루고자 한다면 학문 목적 학습자들의 언어 사용 양상을 참고할 수 있고 어휘를 다루고자 하면 학술기본어휘를 제시하면 될 것이다. 하지만 형식적인 면에서 보면 관사나 조동사 등과 같은 일부 문법 항목을 삭제하는 cloze-elide 형식이 한국어에서는 적절하지 않고, 텍스트 완성의 경우에도 단어의 일부만 보여주고 채우는 방식은 적절하지 않다. <표 2><sup>10)</sup>에서 보는 바와 같이 내용어인 ‘지역’, ‘시민권자’, ‘공정한’을 제외하고 관사나 조동사는 한국어에서는 실현되지 않기 때문이다. 또한 <표 3>과 같이 텍스트 완성의 경우에도 한국어로 바꿔 보았을 때 모아쓰기를 하는 한국어의 특성상 단어의 일부만 쓰도록 하는 것이 효율적이지 않으며 조사나 연결어미, 종결어미를 결합하여 써야 한다. 또한 가능한 답안이 여러 개 제시될 수 있어 선별 시험으로 구현하기에 어려움이 있다.

8) 학술적 말하기 및 쓰기 영역이 객관적 시험(objective testing)의 형식으로 제시되는 것이 적절하지 않으므로 주관적 시험(subjective testing)의 형식이라고 가정하고 언급한 것이다.

9) 선별 시험에서 제외하는 게 적절하다고 언급한 말하기, 쓰기, 듣기 영역은 학문 목적 한국어 능력에서 그 중요도가 낮다는 의미가 아니라 ‘선별’의 목적과 시험의 특징에 초점을 맞추어 고려한 판단임을 다시 한번 강조하고자 한다.

10) PAAL의 빨리 읽기 과제 중 첫 문단을 한국어로 바꿔 본 것이다. <그림 3>에서 제시한 유형과 동일한 유형이다.

〈표 2〉 PAAL 빨리 읽기 과제의 한국어 구현 모습

The simplest conception of citizenship is that it is the state of being a citizen of a particular country. From this point of **the** view, the concept of citizenship is quite straightforward and is often associated with a person's country of **area** birth. Usually, for example, a person born in Canada is thought of as a Canadian citizen; a person born in Japan is a Japanese **of** citizen, one born in **citizenship** Australia is an Australian citizen, and so on. However, citizenship is a multi-dimensional concept **a** which does not lend itself to one, simple **fairly** definition. Therefore, rather than attempting to define the term, this discussion will **of** focus on describing some of the dimensions of citizenship, specifically, the political and legal dimensions.

시민권의 가장 단순한 개념은 그것이 특정 국가의 시민이 되는 상태라는 것이다. 이런 관점에서 볼 때 시민권의 개념은 매우 간단하며 종종 한 개인의 **지역** 출생국과 관련이 있다. 예를 들어, 보통 캐나다에서 태어난 사람은 캐나다 시민으로 생각되고, 일본에서 태어난 사람은 일본 시민권자, 호주 **시민권자**에서 태어난 사람은 호주 시민권자이다. 그러나 시민권은 하나의 단순한 **공정한** 정의에 부합하지 않는 다차원적 개념이다. 따라서 이 논의는 용어를 정의하려고 시도하기보다 시민권의 일부 차원, 특히 정치적, 법적 차원을 설명하는 데 초점을 맞출 것이다.

〈표 3〉 PAAL의 텍스트 과제의 한국어 구현 모습

Early public buildings in Australia were constructed around the importance of influencing community and civic identity. There **was** a **sentimental** attachment **to** the **idea** of public space with a **civic** square ringed **by** **civic** buildings. **In** the **founding** of **the** first buildings in Australia, a **duality** of approaches existed: **those** which **dominated** the landscape and **those** designed **to** blend **in**. Early residential architecture **was** a **response** to the Australian landscape and the climate with its unique flora and fauna and intense sunlight. Early buildings needed to respond to these discrete climatic elements.

호주의 초기 공공건물은 지역사회와 시민 정체성에 영향을 미치는 중요성에 따라 건설되었다. **시민** 건물들로 둘러싸인 **인** **시민** 광장이 있는 **공공의** 공간에 대해 **감상적인** 애착이 있었다. **호주**에서 첫 번째 건물이 세워졌을 때 **두 가지 접근법**이 존재했다. 즉, 경관을 **지배하는** 방식과 혼합되도록 설계된 접근 방식이었다. 초기 **주거** 건축은 독특한 동식물과 강렬한 햇빛으로 호주의 풍경과 기후에 **부응했다**. 초기 건물들은 이러한 개별적인 기후 요소에 대응해야 했다.

다음으로 기존 문항 유형 외에 본 선별 시험에 적절한 평가 요소를 좀 더 고려해 볼 필요가 있다. 이를 위해서는 선별 시험에서 ‘무엇을’ 다루어야 하는지에 대해 좀 더 구체적으로 논할 필요가 있다. 즉 응시자의 학업 수행을 위한 언어 능력이 어떠한지, 언어적 지원이 필요한지 등에 대한 여부를 간단한 선별 시험에서 가려내기 위해 과연 학술적 언어 능력의 어느 부분을 얼마나 어디까지 다루어야 하는지에 대한 것이다. 이는 평가 개발 시 반드시 고려해야 할 목표 언어 사용(TLU, Target Language Use) 영역을 고려하기 위함이기도 하다. 박진욱(2014:54)에서 언급한 것과 같이 학문적 의사소통은 일반적 상황에서의 의사소통보다 고등의 사고 과정을 필요로 하는 복합적인 과정이다. 이에 해당 연구에서는 역량(competencies) 중심의 접근을 주장하며 ‘한국 대학에서 한국어로 학문을 수행할 수 있는 능력’을 한국어 학습역량(Korean study competence)이라 명명하고 이와 관련한 학문의 주요 과업을 수행할 수 있는 수행 능력을 세부역량(competencies)으로 설정하고 있다(박진욱, 2014:60). 박진욱(2014:162-171)에서 도출한 한국어 학습역량 모형에서는 각 세부 역량과 세부 활동별로 중요도를 함께 제시하고 있는데 이를 살펴 선별 시험에서 다루어야 할 내용에 참고하고자 한다. 박진욱(2014:174)에서는 연구 결과로 도출된 세부 역량 16개를 세 개의 군으로 분류한다. 먼저 수업과 관련된 주요 역량군, 수업 시간의 활동을 수행하는 데 도움을 주는 보조 역량군, 마지막으로 학습 결과물로 생산해야 하는 보고서 관련 역량군이 그것이다. 그중 ‘예습복습하기, 발표시험 준비하기, 스터디, 학업 문제 상담하기’가 포함된 보조 역량군 다소 학업 수행의 내면적인 부분과 관련되어 있다면 수업과 직결된 주요 역량군과 학습 결과물과 관련된 보고서 관련 역량군은 학업 수행에서 표면적으로 드러나는 부분과 맞닿아 있다고 할 수 있다(박진욱, 2014:175). 각 역량군은 모두 중요성을

지니고 있어 한국어 학습역량으로 도출되었다고 할 수 있으나 본고의 선별 시험에 적용하기 위해 세부 역량 중 주요 활동 일부를 선택할 필요가 있다. 예를 들어 <표 4>와 같이 ‘강의 듣기’의 한국어 학습역량 모형이 제시된다고 할 때 선별 시험에서는 가장 중요도가 높은 항목(●)과 문항으로 구현이 가능한 세부 활동(기술)을 고려할 수 있다. 중요도가 높은 ‘답화표지에 유의하며 집중하여 강의 듣기’와 ‘중요 내용 메모 및 밑줄 긋기’, ‘모국어 및 한국어 모두 사용해서 중요 내용 노트하기’ 중 선별 시험에서 다룰 수 있는 내용은 ‘듣기’를 제외한 나머지 부분이 될 수 있다.

<표 4> ‘강의 듣기’의 한국어 학습역량 모형(박진욱, 2014:162)

세부역량	주요활동	세부활동	우선순위 (중요도)	
강의듣기 ●	수업 내용 이해하기	진도 확인 후 관련 교재 및 자료 준비하기	○	
		수업 직전 배울 부분 훑어보기	◎	
		모르는 단어 사전을 이용해 찾아두기	◎	
		답화표지에 유의하며 집중하여 강의듣기	●	
		중요 내용 메모 및 밑줄 긋기	●	
		강의 들으며 의문 나는 내용 메모하기	◎	
		모국어 및 한국어 모두 사용해서 중요내용 노트하기	●	
		판서된 내용 사진으로 찍기	○	
		이해가 어려운 부분 녹음하기	○	
		수업 내용 점검하기	예습 시 메모한 질문들 해결 확인하기	◎
			담당 교수에게 모르는 내용 질문하기	◎
			배운 내용과 관련하여 더 찾아볼 내용 정리	◎
			배운 부분 진도 체크해두기	○

※ [비고] 중요도: (높음) ● ◎ ○ (낮음) / 핵심항목: 음영

주요 역량군과 보고서 관련 역량군의 한국어 학습역량 모형에서 선별 시험에 적용 가능한 세부 역량과 학습 활동을 모아보면 다음과 같다. 본격적인 진단평가 과제가 아닌 만큼 ‘선별’ 목적으로 필수적으로, 간단한 형식으로 다룰 수 있는 항목만 제한적으로 추출하였음을 다시 한번 밝힌다.

<표 5> 선별 시험에 적용 가능한 주요 세부 역량 및 활동

역량군	세부 역량	주요 활동	세부 활동
주요 역량군	강의 듣기	수업 내용 이해하기	중요 내용 메모 및 밑줄 긋기
			모국어 및 한국어 모두 사용해서 중요 내용 노트하기
	시험 치기	답안 작성하기	문제 파악하기(포인트가 되는 부분에 밑줄 긋기)
			답안 내용 메모 및 개요 작성하기
보고서 역량군	자료 제시 보고서	보고서 작성하기	보고서 개요 작성하기
			자료를 근거로 혹은 인용하여 자신의 의견 작성하기

<표 5>의 한국어 학습역량을 살펴보면 선별 시험에서 다룰 수 있는 항목은 중요 내용에 밑줄을 긋거나, 메모하기 등의 학습 기술과 개요를 작성하거나 근거를 가지고 의견을 작성하는 것 등 읽기쓰기 영역과 관련이 높고 분류나 선택, 종합 등의 사고 기술이 포함됨을 알 수 있다. 이러한 내용은 이준호(2009:186)의 학문 목적 학습자를 위한 사고 기술(thinking

skills)<sup>11)</sup>과 연결이 되는 내용이라고 할 수 있다. 이 중 <표 5>를 고려하여 선별 과제에서 과제화할 수 있는 사고 기술을 살펴보면 이해하기, 분석하기, 분류하기, 순서화하기, 통합하기, 논증하기를 꼽을 수 있다. 박진욱(2014)이나 이준호(2009)의 내용을 평가로 구현하기 위해서는 이준호(2009)에서 제시한 바와 같이 포트폴리오 과제나 보고서, 글쓰기와 같은 종합적인 활동을 제시하고 있어 본 선별 시험의 방향성과는 맞지 않는 것처럼 보인다. 그러나 간단한 과제를 통해 이러한 종합적인 활동의 가능성을 판별하기 위한 참고 자료로 활용할 수 있다면 선별 시험에서 충분히 다룰 수 있지 않을까 생각한다<sup>12)</sup>. 따라서 학술기본어휘의 크기 측정뿐 아니라 역량별 활동과 사고 기술을 참고하여 반영할 수 있는 것을 포함하면 선별에 더 도움이 될 것이다. 이제 관건은 이러한 학습 기술을 어떻게 평가 문항으로 구현할 수 있을지에 대한 것이라고 할 수 있다.

#### 4. 선별 시험 과제 개발의 실제

지금까지 논의한 내용을 바탕으로 학문 목적 한국어 선별 시험은 30분 정도의 짧고 접근성 높은 인터넷 기반 평가로 구성될 수 있다. 또한 평가 과제는 학습자의 배경을 간단히 확인하는 질문에서부터 어휘 과제, 맥락을 활용한 짧은 텍스트 완성 과제, 학습 기술과 관련된 과제로 <표 6>과 같이 크게 네 부분으로 구성될 수 있다.

<표 6> 학문 목적 선별 시험의 구성

① 배경 질문(시간제한 없음)
② 어휘 과제(15분)
③ 텍스트 완성 과제(6분)
④ 학습 기술 관련 과제(9분)

먼저 ① 배경 질문에서는 소속 학과나 전공, 한국어 학습 기간 및 한국어로 학업 또는 업무를 수행해 본 기간을 묻고, 추가로 TOPIK 등급<sup>13)</sup>을 알아볼 수 있다. ①은 참고사항이지 선별 시험의 판정에 영향을 끼치는 내용이 아니므로 시간 제한을 따로 두지 않는 것이 좋을 것이다.

② 어휘 과제는 선별 시험에서 가장 중요한 부분 중 하나로 크게 1부와 2부로 나누어 Yes/No 형식의 시험과 기타 어휘 과제로 구성할 수 있다. 이 부분에서는 선별 시험의 특성상 분리적(discrete)이고 맥락 독립적(context-independent), 선택적(selective) 성격의 평가와 수용적(receptive)<sup>14)</sup> 성격의 어휘 과제로 구성하고자 한다. 1부인 Yes/No 형식은 진단평가의 '일부'로 많이 사용되는 대표적인 유형으로 부담 없이 빠르게 참여할 수 있고 많은 양의 학술기본어휘를 다룰 수 있

11) 이준호(2009)에서는 이해하기, 분석하기, 분류하기, 순서화하기, 비교하기, 통합하기, 추론하기, 논증하기, 정체성 파악하기, 평가하기의 10가지 사고 기술 목록을 제시하고 있다.

12) 학문적 맥락의 '선별' 시험에서 학습 기술을 포함하는 것이 좋을지에 대해서는 더 많은 전문가의 의견 반영이 필요한 부분이라고 본다. 선별 시험이 학습자들에게 접근성이 높고 짧은 시간 내 이루어져야 하는 시험인 만큼 어휘 외에 맥락 의존적인 문항 활용은 최소화해야 한다는 것이 본 연구자의 입장이다. 하지만 그렇다고 해서 평가 목적과 TLU를 생각해 보았을 때 타당도 높은 선별이 이루어지기 위한 요소가 포착되었다면 기초 연구 단계에서 이를 빼는 것은 바람직하지 않다고 보았다.

13) TOPIK 등급의 경우 대학 입학 시 대부분의 학교에서 요구하는 조건이라 실제 대학 기관에서 사용한다면 불필요한 질문이 되겠지만, 연구 목적으로는 관련 내용이 필요하다고 보고 항목에 포함했다.

14) Read(2000:9)의 어휘 평가 유형과 Nation(2013:558-560)의 난이도에 영향을 끼치는 세 요인에 따른 분류를 참고해 제시하였다. 필요한 경우 한윤정(2021:59,130)을 참고할 수 있다.

다는 장점이 있다. 따라서 수험자의 다양한 배경과 상황을 선별하기 위한 장치로 활용하고자 하나 앞서 언급한 대로 추측해서 맞힐 가능성이 있어 시험 과제의 일부로만 사용하고 30% 정도의 가짜 단어를 포함해 신뢰도를 높일 수 있는 장치를 마련해 둘 수 있다. 가능한 방식으로는 두 가지가 있는데 Meara:Jones(1990)의 EVST(Eurocentres Vocabulary Size Test)와 같이 아는 어휘에 체크 표시하는 방식과 DIALANG의 배치시험과 같이 존재하는 단어라고 생각하는 것이 체크 표시하는 방식이다. 전자는 ‘아는 단어’라는 표현이 학습자 스스로 안다고 과대평가할 수 있는 가능성이 후자보다 높아 신뢰도에 영향을 미칠 수 있다고 보고 본 연구에서는 후자의 방식을 선택해 제시하고자 한다.

㉒ 어휘 과제의 2부는 두 가지 정도의 어휘 과제로 구성될 수 있다. 하나는 VLT(Vocabulary Levels Test)나 VST(Vocabulary Size Test), 정의문 완성하기(A definition completion test)와 같은 의미와 형태를 연결하는 어휘 과제 유형이다. 이러한 유형은 배치와 진단 목적에 유용하고 문항의 신뢰도와 타당도가 높은 것으로 여러 연구에서 확인된(Schmitt, 2010:199) 유형이다. 문항의 예시는 다음과 같다.

〈표 7〉 어휘 의미, 형태 연결 과제 유형 예시

<b>VLT 유형 활용</b>	<p>※ 다음 단어를 보고 알맞은 의미를 골라 번호를 쓰세요. (Schmitt 외, 2001 참고)</p> <table border="1" style="width: 100%;"> <tbody> <tr> <td>1. 가정</td> <td></td> </tr> <tr> <td>2. 대립</td> <td style="text-align: center;">2</td> </tr> <tr> <td>3. 발생</td> <td style="text-align: center;">3</td> </tr> <tr> <td>4. 실시</td> <td style="text-align: center;">5</td> </tr> <tr> <td>5. 유지</td> <td></td> </tr> <tr> <td>6. 조절</td> <td></td> </tr> </tbody> </table> <p style="text-align: right;">서로 반대되는 것 일이 새로 생겨나는 것 변하지 않게 계속하는 것</p>	1. 가정		2. 대립	2	3. 발생	3	4. 실시	5	5. 유지		6. 조절			
1. 가정															
2. 대립	2														
3. 발생	3														
4. 실시	5														
5. 유지															
6. 조절															
<b>정의문 완성하기 유형 활용</b>	<p>※ 빈칸에 들어갈 알맞은 단어를 고르세요. (Read, 1995 참고)</p> <table border="1" style="width: 100%;"> <tbody> <tr> <td>1. 서로 반대되는 것은 ‘ ② ’이다.</td> <td style="text-align: right;">① 유지</td> </tr> <tr> <td>2. 변하지 않게 계속하는 것은 ‘ ① ’이다.</td> <td style="text-align: right;">② 대립</td> </tr> <tr> <td>3. 어떤 일이 새로 생겨나는 것은 ‘ ④ ’이다.</td> <td style="text-align: right;">③ 증명</td> </tr> <tr> <td>4. 어떤 일을 실제로 해 나가는 것은 ‘ ⑦ ’이다.</td> <td style="text-align: right;">④ 발생</td> </tr> <tr> <td>5. 균형이 맞게 바로잡고 맞추는 것은 ‘ ⑤ ’이다.</td> <td style="text-align: right;">⑤ 조절</td> </tr> <tr> <td></td> <td style="text-align: right;">⑥ 가정</td> </tr> <tr> <td></td> <td style="text-align: right;">⑦ 실시</td> </tr> </tbody> </table>	1. 서로 반대되는 것은 ‘ ② ’이다.	① 유지	2. 변하지 않게 계속하는 것은 ‘ ① ’이다.	② 대립	3. 어떤 일이 새로 생겨나는 것은 ‘ ④ ’이다.	③ 증명	4. 어떤 일을 실제로 해 나가는 것은 ‘ ⑦ ’이다.	④ 발생	5. 균형이 맞게 바로잡고 맞추는 것은 ‘ ⑤ ’이다.	⑤ 조절		⑥ 가정		⑦ 실시
1. 서로 반대되는 것은 ‘ ② ’이다.	① 유지														
2. 변하지 않게 계속하는 것은 ‘ ① ’이다.	② 대립														
3. 어떤 일이 새로 생겨나는 것은 ‘ ④ ’이다.	③ 증명														
4. 어떤 일을 실제로 해 나가는 것은 ‘ ⑦ ’이다.	④ 발생														
5. 균형이 맞게 바로잡고 맞추는 것은 ‘ ⑤ ’이다.	⑤ 조절														
	⑥ 가정														
	⑦ 실시														
<b>VST 유형 활용</b>	<p>※ 밑줄 친 단어의 의미로 알맞은 것을 고르세요. (Nation:Beqar, 2007 참고)</p> <p>1. 이 부분은 결론에 해당한다.</p> <p>① 어떤 내용에 반대된다. ② 어떤 일이 실제 상황에 이른다. ③ 어떤 일이 다른 것에 비교된다. ④ 어떤 범위나 조건에 들어맞는다.</p>														

이러한 유형 중 가장 간단하고 부담이 적은 VLT 형식이 선별 시험에서 활용하기 좋다고 본다. VST 형식은 하나의 단어를 위해 읽어야 하는 선택지가 많이 경제성이 가장 떨어지며 정의문 완성하기 형식은 여러 의미를 다룰 수는 있으나 문장 형식으로 되어 있어 간단한 VLT 형식이 더 적절하다고 보았기 때문이다.

다음으로 한국어의 학술기본어휘 특징 중 유의 관계를 이루는 어휘가 많다는 것에 주목하여 유의어 관련 과제도 포함하고자 한다. 선별 시험에서 맥락을 활용해 푸는 문제는 시간이 많이 소요될 수 있으므로 아래와 같은 유형이 가능하다고 본다.

〈표 8〉 유의어 과제 유형 예시

<p><b>유의어 고르기 유형 활용</b></p>	<p>※ 다음 단어와 의미가 비슷한 것을 고르세요. (Schmitt, 2000 참고)</p> <p>1. 예상                  ① 보완                      ② 선택                      ③ <b>전망</b>                      ④ 회상</p> <p>2. 나누다                  ① 기여하다                      ② <b>분류하다</b>                      ③ 적용하다                      ④ 처리하다</p>
<p><b>의미 관계 파악하기 유형 활용</b></p>	<p>※ 다음 두 단어가 서로 관련된 것이면 ‘네’, 관련되지 않은 것이면 ‘아니요’에 ✓ 하세요. (Bogaards, 2000 참고)</p> <p>1. 예상 : 전망                      <input checked="" type="checkbox"/> 네                      <input type="checkbox"/> 아니요</p> <p>2. 선택 : 회상                      <input type="checkbox"/> 네                      <input checked="" type="checkbox"/> 아니요</p> <p>3. 포함 : 제외                      <input checked="" type="checkbox"/> 네                      <input type="checkbox"/> 아니요</p>
<p><b>구 TOPIK 유형 활용</b></p>	<p>※ 다음 밑줄 친 부분과 의미가 비슷한 것을 고르세요. (어휘 및 문법, 고급 10~11번 참고)</p> <p>1. 경제 전문가들은 그 회사의 미래가 밝을 것으로 <u>예상한다</u>.                  ① 보완한다                      ② 선택한다                      ③ <b>전망한다</b>                      ④ 회상한다</p> <p>2. 방식을 변경한 후 시간이 더 걸린다는 단점이 생겼다.                  ① <b>소요된다는</b>                      ② 전환된다는                      ③ 초월된다는                      ④ 형성된다는</p>
<p><b>GRE 언어적 추론 유형 활용</b></p>	<p>※ 다음 빈칸에 단어를 넣었을 때 문장의 의미가 가장 비슷해지는 것 2개를 고르세요.                  (언어적 추론, Section2, 1-5번 참고)</p> <p>1. 경제 전문가들은 그 회사의 미래가 밝을 것으로 _____.</p> <p>① 도출한다                      ② 보완한다                      ③ 선택한다                      ④ <b>예상한다</b></p> <p>⑤ 적용한다                      ⑥ <b>전망한다</b>                      ⑦ 회상한다</p>

〈표 8〉의 유형 중 35회 이전 구 TOPIK 유형은 언어 시험의 대표적인 유의어 평가 유형이기는 하나 맥락 의존적 문항으로 구성되어 선별 시험에서 빠르게 활용하기 어려운 문제가 있다. GRE<sup>15)</sup> 유형도 유의 관계에 있는 어휘 2개를 모두 골라야 한다는 점에서 학문 목적 학습자의 어휘 능력 진단에 유용하나, 역시 맥락 의존적 문항이고 2개의 어휘를 모두 찾아야 한다는 점에서 부담이 높다. 따라서 〈표 8〉의 위 두 유형 중 선별 시험에 더 적합하고 어휘의 의미 관계를 명시적으로 보여주는 ‘의미 관계 파악하기 유형’을 활용해서 문항을 구성할 수 있다. 다만 Bogaards(2000)에서는 여러 계열적 의미 관계를 유형 안에 모두 다루고 있는데 본 연구에서는 학술기본어휘의 유의 관계로 제한하여 적용하고자 한다.

다음으로 ③ 텍스트 완성 과제에 대해 논하고자 한다. 학술 텍스트를 활용하여 맥락 의존적인 문항이 구성된다면 한국어의 조사나 어미만 쓰도록 하는 것이 불가능하므로 내용어 위주로 이루어져야 한다. 또한 학술기본어휘의 종류<sup>16)</sup>를 고려해 평가 대상이 되는 어휘를 선정할 수 있다. 특히 〈표 9〉와 같이 연구 과정과 관련된 어휘(Martin, 1976)가 텍스트 완성 과제에서 활용될 수 있다고 본다.

15) GRE(Graduate Record Examination)는 널리 알려진 미국 대학원 진학을 위한 시험을 가리키는 것이며 언어와 관련된 영역은 ‘분석적 쓰기’와 ‘언어적 추론’이다.

16) Martin(1976:93)에서는 학술기본어휘의 종류를 연구 과정 어휘(define, examine, simulate 등), 분석 어휘(consist of, group, result from, derive, base on 등), 평가 어휘(exhaustive, controversial, coherent, indispensable 등)로 제시하고 있다.

〈표 9〉 텍스트 완성 과제 유형 예시

<b>연구 과정 에서 사용되는 어휘 연습 유형 활용<sup>17)</sup></b>	※ 다음을 읽고 밑줄 친 부분에 들어갈 말을 골라 번호를 쓰세요. 1. 이 실험은 대학에 재학 중인 학생 30명을 대상으로 두 단계로 실시했다. 1차 실험에는 13명이, 2차 실험에는 17명이 <u>(가)</u> . 1, 2차 실험의 시험지는 동일하게 <u>(나)</u> . 본 연구에서는 실험 결과 중 세 문항의 응답을 자세히 <u>(다)</u> .
	① 분석하다                      ② 생략하다                      ③ 구성하다                      ④ 참여하다
	(가) <u>④</u> (나) <u>③</u> (다) <u>①</u>

〈표 9〉는 Martin(1976:93)의 연구 과정에서 사용되는 어휘 연습 유형을 활용하여 재구성한 것이다. 기존 선별 시험에서는 〈표 3〉과 같이 빈칸에 학습자가 직접 단어를 완성하는 생산적(productive) 과제 유형으로 구성되어 있다. 한국어 상황에서 이를 실현하면 조사를 붙이거나 어미를 결합해 써야 하는데, 그럴 경우 정답으로 인정할 수 있는 항목이 늘어 채점 절차가 복잡해진다는 단점이 있다. 따라서 〈표 9〉와 같이 밑줄에 들어갈 수 있는 단어를 보기에서 고르게 하는 방법을 선택하고자 한다.

마지막으로 ④는 과제의 범주를 ‘학습 기술’이라 명명하고 앞에서 언급한 박진욱(2014), 이준호(2009)의 내용을 다룬다. 그런데 이 범주에서 다룰 수 있는 평가 항목이 무엇인지를 먼저 정리하는 것이 필요하다. 앞에서 구현 가능한 세부 활동과 사고 기술을 정리해 보았을 때 ‘작성하기’는 직접적으로 다루기 어려우므로 간접적인 방식으로 검사가 가능하다. 예를 들어 ‘개요 작성’의 경우 직접 작성하는 본격적인 활동이 선별 시험에서는 적절하지 않으니 다음 〈표 10〉과 같이 보고서의 세부 내용이 어느 부분에 해당하는지 연결하는 방법을 고려해 볼 수 있다. 이는 이해하기, 분석하기, 분류하기, 순서화하기의 사고 기술을 필요로 한다.

〈표 10〉 개요 작성 관련 과제 유형 예시

※ ‘한류의 현황과 과제’라는 주제로 보고서를 쓰려고 합니다. (가)~(사)의 세부 내용이 서론, 본론, 결론 중 어디에 포함되는 내용인지 연결해 보세요 <sup>18)</sup> .	
1. 서론 2. 본론 3. 결론	

이와 더불어 수업 내용이나 자료의 중요 내용을 파악하는 활동에도 초점을 맞춰볼 수 있다. 이는 어떠한 글 또는 자료의 주제나 중심 내용을 파악하는 활동과 일맥상통한다. 또한 자료를 근거로 자신의 의견을 작성하는 활동은 자료의 내용을 잘 분석해 적절한 의견을 도출하는 활동으로 변형해 볼 수 있다. 이는 이해하기, 분석하기, 통합하기, 추론하기, 논증하기의 사고 기술을 필요로 한다.

17) 이 과제 유형은 한운정(2021)의 학문 목적 어휘 진단평가 파일럿 테스트에서도 활용된 바 있다. 본 연구에서는 진단뿐만 아니라 선별의 목적에서도 활용될 수 있다고 보아 선정하였다.

18) 이 과제의 형식은 『서울대 한국어+ 학문 목적 쓰기(서울대학교 언어교육원, 2017:55)』를 참고하고 내용은 『바로 쓰는 한국어 6(이정희 외, 2019:92)』를 참고하였다.

〈표 11〉 주요 내용 찾기 및 의견 제시 관련 과제 유형 예시

※ 다음 글을 읽고 질문에 답하세요<sup>19)</sup>.

보통 어린아이의 기억력은 어른의 기억력보다 떨어진다고 여겨져 재판에서 아이들의 증언이 증거로 채택되지 않는 경우가 많다. 그러나 아이들의 증언이 어른들보다 신빙성이 높다는 연구 결과가 나왔다. 아이들은 한 번에 많은 내용을 기억하지는 못하지만 기억하는 내용은 정확하다는 것이다. 이에 비해 어른들의 경우, 자신들이 알고 있던 지식이나 경험이 기억과 섞이면서 실제와 허구를 혼동하는 경향을 보였다.

1. 윗글의 주제문으로 가장 알맞은 것에 밑줄을 그어 보세요.
2. 이 연구 결과를 근거로 주장할 수 있는 내용으로 가장 알맞은 것을 고르세요.
  - ① “재판에서 증언의 가치와 중요성이 더 확대될 필요가 있다고 봅니다.”
  - ② “기억의 정확성을 좀 더 객관적으로 측정할 수 있는 기준이 있어야 합니다.”
  - ③ “아이의 증언도 믿을 수 있으니 재판 시 증거로 인정될 수 있다고 생각합니다.”
  - ④ “아이나 어른의 증언은 완전히 정확하지 않으니 물질적인 증거가 꼭 필요하다고 봅니다.”

## 5. 결론

(생략)

19) TOPIK 26회 고급 이해 영역 38번의 지문을 활용하였다.

■ 참고문헌 ■

- 박진욱(2014). 학습역량 기반 학문 목적 한국어 교육과정 연구 -전공 진입 전 과정을 중심으로-. 고려대학교 대학원 박사학위논문.
- 신인철 · 한지은 · 박효민(2018). 외국인 유학생의 학업중단 실태와 요인. *다문화사회연구*, 11-2, 105-133쪽.
- 윤새롬 · 김영규(2017). 한국어 학술어 목록의 학술 텍스트 점유율에 관한 연구. *외국어로서의 한국어교육*, 47, 193-212쪽.
- 이선영 · 나운주(2018). 외국인 유학생의 학업적응 실태 조사 -교양 교과목 개발을 위한 기초 연구. *교양교육연구*, 2-6, 167-193쪽.
- 이준호(2009). KAP 학습자를 위한 사고 기술 평가 방안- 사고 기술과 평가 과제 선정을 중심으로-. *한국어교육*, 20-2, 175-201쪽.
- 이준호 · 전형길 · 이인혜(2020). 초등 학습자 대상 한국어(KSL) 진단평가 연구 - DLA, WIDA, ELPAC 평가 분석을 바탕으로-. *Journal of Korean Culture*, 49, 43-83쪽.
- 이효경(2015). 한국어 어휘 지식 평가 도구 개발: KAP 학습자 대상 학술 기본 어휘를 중심으로. *이화여자대학교 외국어교육특수대학원 석사학위논문*.
- 정연숙(2018). 학문목적 한국어 교육에 대한 외국인 유학생 인식 조사: 한국어 어휘를 중심으로. *교사교육연구*, 57-4, 481-498쪽.
- 한운정(2021). 학문 목적 어휘 진단평가 연구. *경희대학교 대학원 박사학위논문*.
- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: the interface between learning and assessment*. NY : Continuum.
- Bogaards, P. (2000). Testing L2 vocabulary knowledge at a high level: The case of the Euralex French tests. *Applied Linguistics* 21(4), pp.490-516.
- Fox, J., Haggerty, J., & Artemeva, N. (2016). Mitigating Risk: The Impact of a Diagnostic Assessment Procedure on the First-Year Experience in Engineering. In Raed, J. (ed.), *Post-admission Language Assessment of University Student*. Switzerland: Springer International Publishing.
- Jordan, R. R. (1997). *English for Academic Purposes*. Cambridge: Cambridge University Press.
- Martin, A. V. (1976). Teaching Academic Vocabulary to Foreign Graduate Students. *TESOL Quarterly*, 10(1), pp. 91-97.
- Meara, P., & Jones, G. (1990). *Eurocentres Vocabulary Size Test* (version E1.1/K10,MSDOS). Zurich: Eurocentres Learning Service.
- Nation, I. S. P. (2013). *Learning Vocabulary in Another Language*, 2nd edition. NY: Cambridge University Press.
- Read, J. (2000). *Assessing Vocabulary*. UK: Cambridge University Press. 배도용·전영미 역(2015). 어휘 평가. 글로벌콘텐츠.
- Read, J. (2016). *Post-admission Language Assessment of University Student*. Switzerland: Springer International Publishing.
- Roche, T., & Harrington, M. (2013). Recognition vocabulary knowledge as a predictor of academic performance in an English as a foreign language setting. *Language Testing in Asia*, 3.
- Roche, T., Harrington, M., Sinha, Y. & Denman, C. (2016). Vocabulary recognition skill as a screening tool in English-as-a-Lingua-Franca university settings. In Raed, J. (ed.), *Post-admission Language Assessment of University Student*. Switzerland: Springer International Publishing.
- Saville-Troike, M. (1984). What Really Matters in Second Language Learning for Academic Achievement?. *TESOL Quarterly*, 18(2), pp.199-219.
- Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual*. UK: Palgrave Macmillan.
- University of Auckland. (n.d). *DELNA Handbook*. Retrieved from <https://cdn.auckland.ac.nz/assets/delna/delna/documents/delna-handbook.pdf>.
- University of Melbourne LTRC(Language Testing Research center) (n.d). *DELA Handbook for Candidates*. Retrieved from [https://students.unimelb.edu.au/\\_data/assets/pdf\\_file/0007/3284800/DELA-Handbook-2016.pdf](https://students.unimelb.edu.au/_data/assets/pdf_file/0007/3284800/DELA-Handbook-2016.pdf).
- University of Melbourne LTRC(Language Testing Research center) (n.d). *PAAL Handbook for Candidates*. Received by e-mail from the LTRC.

## 〈토론〉 “선별(Screening) 목적의 인터넷 기반 한국어 진단평가 과제 개발 기초 연구”에 대한 토론문

이향(한국조지메이슨대학교)

본 연구는 대학 입학 후 선별 목적의 진단평가 과제 개발을 위한 기초 연구로서 아직까지 한국어 교육에서는 드물었던 진단 평가라는 특수 평가에 대한 연구라는 데 그 의의가 있고, 개인적으로도 몇 전부터 관심을 가지고 공부하고 있던 분야라 흥미를 가지고 잘 읽었습니다. 발표문은 연구자께서 말씀하셨듯 기존에 영어 교육에서 사용되어 온 진단 평가를 분석하고, 이를 한국어에 적용할 수 있는지에 대한 가능성을 살펴본 기초 연구로서, 발표문에서는 완성된 평가 샘플을 볼 수 있는 것이 아니었기 때문에 읽으면서 궁금했던 점과 개인적으로 진단 평가에 대해 공부를 하면서 고민을 해 왔던 부분들을 나누는 것으로 토론을 대신하고자 합니다.

연구자께서 보여 주신 세 가지 진단 평가들은 현재 사용되고 있는 진단 평가들로 모두 10년 이상 실제로 사용되고 있는 평가들이기는 하지만 아직까지 각각의 시험들의 타당성에 대한 논의/논쟁이 진행 중인 평가들입니다. 특히 이 시험들이 “진단 평가”라는 용어를 사용하고는 있지만, 실제로 정말 수험자들의 언어 능력을 “진단”하고 수험자들에게 “진단 정보”를 제공하는지, 그게 아니라 오히려 “배치 평가”의 성격을 가지고 있는 것이 아닌가하는 본질적인 문제와 개발한 평가가 타당하다는 것을 어떻게 객관적으로 증명할 수 있는가 하는 타당화에 대한 문제들이 지적되어 왔습니다. 물론 DELNA 같은 경우 연구진들이 자체적으로 타당성을 증명한 연구를 진행하긴 하였으나 다른 숙달도 평가들 처럼 다각적으로 연구되었다고 보기는 힘들 것 같습니다. 연구자께서는 이러한 논쟁들에 대한 어떤 의견을 가지고 있으신지 궁금합니다. 연구자께서 생각하시는 ‘진단 평가’란 무엇이며 본 평가를 통하여 궁극적으로 학생들에게 제공하고자 하는 진단 정보는 무엇인지 말씀해 주시면 감사하겠습니다.

또한 본 연구에서는 선별 평가에 있어서 어휘와 읽기를 통한 평가를 해야 한다고 하셨습니다. 그럼, 선별된 어휘와 읽기 과제의 내용 타당성 또한 매우 중요한 요인일 텐데, 어휘나 읽기 과제의 선정은 어떤 기준을 가지고 선정해야 한다고 생각하시는지 알고 싶습니다. 그리고 어휘 과제의 1부 유형인 Yes/No 유형의 경우 수험자가 가지고 있는 어휘의 크기를 측정하기 위한 것이 그 목적인데 몇 개의 어휘를 제시하고 측정하는 것이 적절하고, 몇 개를 맞아야지 통과할 수 있는 기준으로 볼 것이며, 그 타당성의 객관적인 근거는 어떻게 증명할 수 있다고 보시는지 말씀해 주시면 감사하겠습니다.

그리고 어휘 과제 2나 텍스트 완성하기 과제의 경우 선택되지 않은 다른 유형들은 모두 표면적으로는 같은 능력을 측정하고 있는 것 같으나 사실 각각의 유형들이 측정하고 있는 세부 언어 능력은 다르다고 보이는데, 과제 선정의 기준이 시간적인 경제성을 중요한 선정 기준으로 보신 것 같습니다. 궁극적으로 이들 과제들 즉 어휘 과제2와 텍스트 완성하기 과제를 통하여 측정하고자 하는 것이 무엇인지 궁금합니다.

마찬가지로 학습 기술 관련 과제나 주요 내용 파악하는 활동을 통하여 측정하고자 하는 수험자의 능력이 무엇이며,

〈표 10〉이나 〈표 11〉의 과제를 성공적으로 수행한 수험자의 경우 어떤 선별 혹은 진단 정보를 제공해 준다고 보시지요.

마지막으로 본 연구에서 “학문 목적 선별 시험”, “선별 목적의 인터넷 기반 한국어 진단 평가” 등 제시하고자 하는 평가의 용어가 여러 개 혼재되어 사용되고 있는데, 용어를 영어 직역에서 벗어나 개발하고자 하는 혹은 제안하고자 하는 평가의 목적과 부합하는 정확한 용어를 사용하면 본 연구에서 제안하고자 하는 평가의 목적을 이해하는 데 도움을 줄 것 같습니다.