

코퍼스 지표를 활용한 모의 토익시험의 유용성 검증 : 난이도와 변별도 분석을 중심으로

Verification of the Usefulness of the Mock TOEIC Test using Corpus Indices : Focusing on the Analysis of Difficulty and Discrimination

이예나

차의과학대학교

Yena Lee(eren01@cha.ac.kr)

요약

본 연구에서는 토익 시험의 정답률과 변별도에 영향을 미치는 구조적인 요인이 무엇인지 분석하기 위하여 문항 분석에서 도출된 각 파트별 코퍼스 지표들을 분석하였다. 이를 위하여 모의 토익 시험의 정답률과 변별도에 대한 코퍼스 요인들의 회귀 분석을 실시하였고, 분석 결과는 다음과 같다. 정답률에 대해서는 기초산출치 중에서 word_length, no_word_sentence1, sentence_length, 정합성 지표들 중에서는 LSA_overlap_adjacent_sentences, 어휘 다양성 지표들 중에서는 MTLD_VOCD, 접속사 지표들 중에서는 All_logical_causal_connectives_incidence, 상황모형 지표들 중에서는 casual_particles_causal_verbs_Ratio, 통사적복잡성 지표들 중에서는 Minimal_Edit_Distance1, Left_embeddedness, 통사적 패턴밀도 지표들 중에서는 Infinitive_density, Preposition_phrase_density 등이 음의 영향을 미치는 것으로 나타났다. 이러한 코퍼스 요인들은 토익의 정답률을 낮추는 효과가 있기 때문에 하위 레벨의 학습자들이 단기적인 학습 효과를 향상시킬 수 있는 중요한 정보들을 제공해 줄 수 있을 것이다. 변별도에 대해서는 어휘 다양성 지표들 중에서 MTLD_VOCD, 접속사 지표들 중에서는 All_logical_causal_connectives_incidence, Additive_connectives_incidence, 통사적 패턴밀도 지표들 중에서는 Infinitive_density, 어휘정보 지표들 중에서는 person1_2_pronoun_incidence 등이 정의 영향을 미치는 것으로 나타났다. 이러한 코퍼스 요인들은 변별도 증가에 영향을 주기 때문에 영어 능력 하위 그룹에 필요한 학습 프로그램을 개발하는 데 있어서 활용될 수 있을 것이다.

■ 중심어 : | 코퍼스 | 토익 | 유용성 | 문항분석 | 정답률 | 변별도 |

Abstract

In this study, in order to investigate the factors that affect the percentage of correct answers and the degree of discrimination of the TOEIC test, a regression analysis was performed using corpus indicators that influence correct answer rate and the degree of discrimination for each part derived from the item analysis. The basic calculation word_length, consistency index LSA_overlap_adjacent_sentences, lexical diversity MTLD_VOCD, conjunction All_logical_causal_connectives_incidence, situational model casual_particles_causal_verbs_Ratio, syntactic complexity Left_embeddedness, and syntactic pattern density Infinitive_density were found to have negative effects. These factors that lower the correct answer rate can be utilized when setting learning goals. Vocabulary diversity index MTLD_VOCD, conjunction Additive_connectives_incidence, syntactic pattern density Infinitive_density, and lexical information person1_2_pronoun_incidence were found to have a positive effect. Factors influencing the increase in discrimination may provide important information for developing a learning program.

■ keyword : | Corpus | TOEIC | Usefulness | Item Analysis | Item Difficulty | Degree of Discrimination |

접수일자 : 2021년 08월 12일
수정일자 : 2021년 09월 17일

심사완료일 : 2021년 10월 13일
교신저자 : 이예나, e-mail : erena01@cha.ac.kr

I. 서론

토익 시험의 목적은 일상생활 또는 비즈니스 업무에서 요구되는 영어능력을 평가하는 것에 중점을 두고 있지만, 국내에서는 이것을 대학 입학이나 기관의 취업, 장학생 선발 등의 다양한 곳에서 비교적 영향력 있는 영어능력 평가 방법으로 사용하고 있다. 이와 같이 토익 성적이 보편적인 영어능력 평가의 방법으로 사용되는 이유는 토익의 평가 성적이 0점부터 990점사이의 5점 단위로 부여되어 다른 시험들 보다 비교적 높은 변별도를 가진다는 것에 기인한다[1]. 토익 시험의 넓은 점수 영역대는 학생들의 성적을 평가함에 있어서 보다 높은 변별력을 가질 수 있다. 하지만, 듣기와 읽기 영역을 중요하게 평가하는 토익 시험의 특성상 토익 성적이 일반적인 언어능력의 평가 기준으로 사용되기에는 그 검증 과정이 충분치 않다. 또한 토익의 넓은 점수 영역대는 작은 점수의 차이가 그만큼 의사소통 능력의 차이를 나타내거나, 그것이 실제로 높은 변별도를 가지는지에 대해서는 충분한 검증이 이루어지지 않은 것이 사실이다. 따라서 국내에서의 토익 시험 성적의 평가와 활용을 위해서 이 시험이 실제 영어능력 및 의사소통능력을 적절하게 평가하고 있는지에 대한 검증이 필요하다. 이를 위하여 본 연구에서는 토익 시험의 정답률과 변별도에 영향을 미치는 평가 시험의 구조적인 요인이 무엇인지 분석하고 그 결과를 토대로 토익 시험을 준비하는 응시생들에게 학습 목표를 제시하고자 한다.

코퍼스 분석은 난이도에 영향을 줄 수 있는 어휘와 문장의 구조적인 특성을 파악할 수 있다. 그러나 코퍼스를 통한 유용성 분석은 영어 평가 시험의 구조적인 특성을 파악하고, 평가 시험의 어휘 수준을 측정하여 난이도와 변별력 수준을 분석하는데 그치고 있다[2]. 따라서 본 연구에서는 문항 분석에서 도출된 각 파트별 정답률과 변별도 측정값들에 영향을 주고 있는 코퍼스 지표들이 무엇인지 분석하고 그 의미들을 논의하고자 하였다. 정답률은 응시자들의 시험 수행 능력을 의미하므로 정답률이 낮은 경우 그 원인이 되는 코퍼스 지표들을 찾아내서 그 지표와 관련된 내용의 영어 교육을 실시해야 할 것이다. 또한 변별도의 경우 상, 하 레벨의 차이를 의미하므로 변별도가 높은 경우 그 원인이 되는

코퍼스 지표가 무엇인지 찾아내고 그 코퍼스 지표와 관련하여 하위 그룹들이 특별히 대비할 수 있는 영어 교육을 시행하는 것이 중요하다고 할 수 있다. 따라서 본 연구의 연구과제는 첫째, 파트별로 정답률과 변별도의 차이가 얼마나 되는지를 분석하고 둘째, 그 차이의 원인이 되는 코퍼스 지표들과 그 영향력을 분석하고자 하였다. 마지막으로 이러한 연구 결과를 토대로 평가 시험의 유용성과 학습에 대한 시사점을 제시하고자 한다.

II. 이론적 배경 및 선행 연구

1. 어학 시험의 유용성과 난이도

영어 시험의 난이도는 시험의 유용성을 결정하는 중요한 지표이다. 시험의 유용성을 구성하는 신뢰도와 타당도는 결국 평가 시험의 난이도에 따라서 달라질 수 있기 때문이다. 난이도는 평가 시험의 구조적인 요인에 의해서 결정되며 특히 토익 시험의 경우 성적은 각 파트별 출제 문항들의 난이도를 결정하는 구조적인 요인에 따라 달라질 수 있다. 일반적으로 토익 시험의 파트별 난이도는 듣기 영역인 파트 1이 가장 쉽고, 읽기 영역 중 문법 파트인 파트5와 독해파트인 파트 7이 다소 어렵게 출제된다[2]. 토익 시험의 각 파트별 난이도를 결정하는 구조적인 요인과 이러한 요인들이 파트별 시험 성적에 미치는 영향과 구조적 요인의 영향력을 분석하여 유용성을 측정할 수 있다.

1.1 난이도의 영향 요인

영어 시험에서의 난이도는 세부 문항 난이도의 합이라고 할 수 있다. 문항 난이도란 문항별 정답 비율을 통해 문항 난이도를 추측하는 기법으로, 시험의 난이도를 측정하는 방법 중 가장 합리적인 방법이라고 할 수 있다[3]. 이러한 문항 난이도 측정은, 해당 시험에서 응시자의 언어능력을 어떠한 수준으로 평가할 것인가에 대한 올바른 방법을 수립하는데 도움을 줄 수 있다[4].

어학 시험의 난이도에 영향을 주는 변인들이란 영어 시험의 난이도를 측정하기 위한 각 개별 요소를 의미하며, 언어 형태, 질문 형태, 배경 지식, 어휘 지식 및 문장 길이 등이 포함된다고 할 수 있다.

첫째, 언어 형태에 따라서 Alderson[5]은 질문을 제시하는 언어가 지문 자체보다 이해하기 어렵다면 응시자가 느끼는 문항에 대한 난이도는 훨씬 더 어려워질 것이라고 했고, Shohamy[6]는 선다형 문제의 질문이 모국어로 제시되었을 때 같은 문제를 제2언어로 제시한 것보다는 쉽다는 것을 연구를 통해 밝혔다.

둘째, 질문의 형태에 따라서도 문항의 난이도에 영향을 미칠 수 있는데, Pearson과 Johnson[7]에 따르면, 문장 내에서 핵심 정보를 찾을 수 있는 명시적 질문, 여러 문장을 통해 추측 가능한 암시적 질문, 응시자의 배경 지식을 통해 유추해야 하는 함축적 질문의 세 가지의 형태로 구분한다. 일반적으로 명시적 질문은 암시적 질문보다 낮은 난이도를 가진다[8].

셋째, 많은 학자들이 읽기 능력과 배경 지식이 지문의 난이도에 미치는 영향에 대해 연구한 결과, 두 변인 모두 영향을 주지만 읽기 능력이 더 영향을 준다는 분석[9]도 있고, 상대적으로 쉬운 지문에서는 언어 능력만 필요로 하지만 어려운 지문을 읽을 때는 배경 지식과 읽기 능력이 모두 필요하기도 하다는 해석도 나왔다[10]. 또한 Clapham[11]에 따르면 읽기 능력 수준이 낮은 응시자들은 자신의 전문 분야에 관한 지문도 어려워했지만, 읽기 능력 수준이 높은 응시자들은 자신의 전문 분야가 아니더라도 지문을 읽는 데 어려움을 덜 느낀다고 한다.

넷째, 어휘에 대한 지식 또한 중요한데, Qian[12]는 TOEFL시험을 바탕으로 어휘 지식과 읽기 능력의 상관관계는 .77($p < .01$)으로 높은 상관관계를 갖는다는 것을 밝혀냈다. 어휘 능력을 측정하는 시험에서 해당 어휘를 명시적으로 아는 경우도 있지만, 해당 어휘를 '모르겠다' 라고 응답한 응시자가 문항을 맞춘 다수의 경우도 있는데[13], 이는 자신도 모르게 어휘를 알고 있거나, 정답 이외의 다른 어휘에서 힌트를 얻었을 가능성이 있다. 이는 Johnston[14]의 어휘 평가 문항에 있어서 어휘 자체에 대한 지식도 중요하지만, 응시자의 기존 배경 지식 또한 중요하게 작용한다는 연구 결과와 일맥상통한다고 할 수 있다.

다섯째, 문장의 길에 있어서, Kyoung-Ae Jin과 Chung Park[4]의 연구에 따르면, 지문의 길이 자체보다는 지문 내에 포함된 어려운 단어의 개수가 많을수록

높은 문항 난이도를 가진다고 분석하였다.

1.2 구문의 복잡성과 난이도

광범위하게 말하자면, 구문 복잡성은 사용된 구문 구조의 다양성과 정교함 정도로 해석할 수 있고[15], 여기서 다양성은 구문 구조의 범위를 의미하고 정교함은 구문의 범위를 나타낸다. L2와 관련된 많은 작문 연구에서 구문의 복잡성과 외국어 숙련도와의 관계를 분석하고 있다[16][17], 이러한 연구들 중에서 L2 숙련 및 L2 작문에 대한 구문 복잡성 품질에 관한 연구[18][19], trajec-L2 학습자의 구문 개발에 관한 연구[20]에서는 구문의 복잡성이 L2 쓰기 능력에 영향을 미치는 것으로 나타났다[21][22].

본 연구에서는 토의 시험의 난이도를 정의하는 개념으로 구문의 복잡성 수준을 측정하였다. 구문의 복잡성은 코퍼스 분석을 통한 구조적인 지표로서 다양한 어학 시험 분야에서 시험 난이도를 결정하는 아주 중요한 요소로 밝혀지고 있다. 이는 주로 구문의 밀도로 측정되는 경우가 많다. 구문의 밀도는 각 품사별로 문장 내의 밀도를 통하여 측정되는데, 밀도가 높을수록 구문의 복잡성은 증가하고 난이도가 높아지는 결과를 초래하게 된다.

1.3 코퍼스를 통한 난이도 평가

코퍼스란 체계적인 연구를 목적으로 만들어진 말뭉치, 즉 컴퓨터가 읽을 수 있는 형태로 텍스트를 모아 놓은 언어 자료의 집합체를 가리키며, 코퍼스 언어학은 이러한 전자 분석을 바탕으로 하는 언어 연구나 연구 방법론을 말한다. 코퍼스 언어학이 언어 난이도 측면에서 유용하게 사용될 수 있는 이유는 어휘의 출현 횟수, 즉 빈도(frequency) 정보를 쉽게 얻을 수 있기 때문이다. 어떤 언어 항목이 유형(Type)이라면, 개별적으로 출현하는 단어 형태가 구현(Token)인데 코퍼스 언어학에서는 어휘의 유형뿐만 아니라 그 유형의 빈도, 즉 구현까지도 포착해 낸다[23]. 전체 텍스트에 사용된 어휘 수는 학습자의 학습 부담감과 관련하여 텍스트의 난이도를 결정하는 데 중요한 역할을 한다.

그러나 어휘 수가 아무리 많아 해도, 반복적으로 사용된 어휘의 비율이 높으면 그 텍스트의 난이도는 상대

적으로 낮아지게 된다[24]. 따라서 텍스트의 전체 어휘량만으로 텍스트의 난이도를 논하는 것은 의미가 없기 때문에, 진정성 있게 난이도를 논하는 방법으로 Schmitt[25]는 같은 양의 텍스트에 얼마나 많은 어휘가 사용되었는지를 알아봄으로써 텍스트의 난이도를 판단하는 것이 적절하다고 주장하였다.

코퍼스를 통한 영어 시험의 난이도를 평가하는 기존의 국내 연구는 국가적 교육방침에 따라 실시하는 청소년기의 고등교육과정 영어시험에 대해서만 일부 존재한다[26][27]. 하지만, 사회에서 실질적인 영어능력을 필요로 하는 성인들이 응시하는 공인영어시험에 대한 코퍼스 분석 활용은 매우 부족한 실정이다. 따라서 본 연구에서는 이러한 어휘 기반의 다양한 코퍼스 지표들이 시험의 난이도에 미치는 영향을 측정하고, 이러한 요인들에 대한 실증 분석을 통하여 공인영어시험 중 가장 대표적인 토익 시험의 유용성을 논의하고자 한다.

2. 코퍼스 분석

2.1 코퍼스 분석의 개념적 틀

코퍼스(Corpus)란 문장의 기본 요소인 텍스트 데이터를 기반으로 하나의 언어를 해석하여 이들 간의 유사성에 대한 해석 지표를 제공해 주는 문장 분석 기법이다. 코퍼스는 자연스러운 의사소통 과정에서 특정 어휘가 쓰이는 상황이나 그 어휘의 다양성 또한 텍스트의 모음 형태로 보여준다[28]. Biber, Douglas, Conrad & Reppen[29]은 코퍼스를 일정한 특정 원칙에 따라 대규모의 자연어 말뭉치들을 정규화하여 분류한 것이라고 정의하였다. 따라서 일정한 원칙에 따라 컴퓨터에 수집, 저장된 다량의 실제 언어 집합체를 코퍼스라고 정의할 수 있다. 이것은 실제로 사용되는 언어 데이터를 수집하여 해당 언어의 빈도수, 언어(Collocation) 그리고 의미적 차이 등 다양한 관점에서 정보 처리를 가능하게 한다. 따라서 코퍼스를 통한 어휘 학습은 어휘 빈도수를 파악하여 주어진 주제, 장르 등에 따른 데이터를 분석하여 사용 어휘의 차이를 알 수 있다[30].

코퍼스에서 각 단어들은 언어(collocation), 연결 범주(colligation), 의미적 선호(semantic preference), 의미적 운율(semantic prosody) 등의 네 가지의 의미 있는 관계로 설명된다[31]. 언어는 둘 혹은 그 이상의

단어들이 동시에 나타나는 현상으로 의사소통 능력 발달에서 그 중요성이 강조되고 있다. 연결 범주는 한 단어와 다른 단어들의 문법적 관계인 통사적 범주로 본다. 이는 문법적으로 강하게 연결하여 함께 나타나는 단어들을 통해 교사와 학습자가 어휘뿐만 아니라 문법적인 환경에 대한 상호 관계를 이해하도록 돕는다[32]. 의미적 선호는 각 단어들의 관계를 나타내는 의미적 범주를 설명하는 것이다. 의미적 운율은 특정한 환경에서 함축적인 의미를 보이는 단어가 나타난다[33].

이와 같이 코퍼스는 난이도에 영향을 줄 수 있는 어휘와 문장의 구조적인 특성을 파악하는 데 매우 유용한 도구이며, 다양한 장르의 텍스트들을 대상으로 어휘의 종류와 수준, 응시자들의 영어 능력의 특성을 측정하는데 많이 활용되어왔다[34]. 그러나 코퍼스를 통한 시험 유용성 분석은 영어 평가 시험의 구조적인 특성을 파악하고, 평가 시험의 어휘 수준을 측정하는 데 그치고 있다. 코퍼스를 활용한 영어 평가 시험의 구조적인 특성이 실제로 영어 시험 성적에 미치는 영향에 대해서는 연구 결과가 매우 미비하다. 특히 시험의 유용성을 구성하는 신뢰도와 타당도의 경우에는 코퍼스의 구조적인 요인들이 어떤 영향을 미치는지 아직 알려진 바가 없다. 따라서 본 연구에서는 코퍼스 분석에 의한 구조적인 요인들이 토익 시험의 유용성을 결정하는 정답률과 변별도에 어떤 영향을 미치고 있는지를 연구하고자 하였다. 이러한 연구 결과는 토익 시험의 유용성을 향상시키기 위한 문항 개발과 코퍼스 지표를 활용하는 방안들에 대한 시사점을 제공해 줄 수 있을 것이다.

2.2 Coh-Matrix

Coh-Matrix는 Jurafsky & Martin[35]이 정의한 전산 언어학적 방법들을 이용하여 텍스트의 다양한 언어학 및 심리 언어학적 측정치를 제공하는 언어 분석 프로그램이다[36][37]. Coh-Matrix는 현재 전산 언어학 분야에서 측정할 수 있는 거의 모든 언어학적 측정치를 제공하며[38], 주요 요소는 다음과 같다[39].

먼저 기초 산출치(Descriptive)는 문장 수, 단어 수, 문장당 평균 단어 수를 포함한 평균 문장 길이에 대한 정보를 제공한다. 문장의 길이는 난이도를 간접 측정하는 가장 기본적인 지표이다. 어휘 다양성(Lexical

Diversity)은 단위 문장 내 전체 단어 수(Type) 대비 고유한 단어 수(Token)의 비율을 의미하는 Type-Token Ratio(TTR)을 제공한다. TTR 값이 높을수록 보편적으로 높은 난이도의 문장이라고 추측할 수 있다. 어휘 정보(World Information)는 Baayen, Piepenbrock, & Gulikers의 CELEX[40], MRC(Medical Research Council Psycholinguistic Database)[41], WordNet [42][43]등의 언어 데이터베이스를 이용하여 단어 빈도(word frequency), 습득 나이(age of acquisition), 친숙도(familiarity), 구체성(concreteness), 상상성(imagability), 다의성(polysemy), 명사에 대한 상하의성(hypernymy), 동사에 대한 상하의성에 대한 정보를 제공한다. 단어 빈도수는 Baayen 등[40]의 CELEX 데이터베이스에 기반하여 특정 단어들의 출연 빈도를 나타내주는 정량적 수치다. 읽기 난이도 지수는 일반적으로 사용되는 FRE(Flesch Reading Ease)[44]와 FKGL(Flesch Kincaid Grade Level)[45]의 읽기 난이도 지수에 대한 정량적 수치를 제공한다. FRE가 높을수록 이해도가 높고, 난이도가 낮은 문장임을 의미한다. 통사적 복잡성 지수는 Charniak[46]의 통사구문 분석기에 기반한 명사구 내에 포함된 수식어들의 평균 개수, 본동사 앞에 위치한 단어의 평균 개수, 문장의 통사 구조(Syntactic structure)에 포함된 문장 성분(Constituents)의 평균 개수의 수치를 제공한다. 통사 구조 밀도는 명사구 밀도(noun phrase density), 동사구 밀도(verb phrase density), 부사구 밀도(adverbial phrase density), 전치사구 밀도(preposition phrase density), 수동태 밀도(agentless passive voice density), 부정어 밀도(negation density), 동명사 밀도(gerund density), 부정사 밀도(infinitive density) 등에 대한 정보를 제공하는데, 이들 통사 구조 밀도는 텍스트에서 나타나는 1,000 단어 당 발생 횟수(incidence)로 나타내어 제공한다. 참조적 응집성은 하나의 명사, 대명사, 혹은 명사구가 텍스트 내에 인접해 있는 다른 구성 요소(constituent)를 지칭할 때 발생하는 현상이다[36][47]. Coh-Matrix에서 제공하는 인접 문장들에 대한 참조적 응집성 측정치는 공통의 논항들(arguments)을 포함하고 있는 인접 문장들의 수를 전체 인접 문장들의 수로 나눈 비율로 산출된다. 의미적 응집성은 Coh-Matrix

에 사용되는 서로 인접한 문장 간의 의미적 연결 정도를 의미한다. 두 개 이상의 텍스트에 나타난 이 값은 코사인 값을 통해 0~1사이의 값으로 계산될 수 있으며, 값이 클수록 상호 간의 높은 의미적 응집성을 갖는다.

III. 연구 방법

1. 데이터 수집

본 연구에서 사용된 표본은 일반적인 4년제 대학교의 토의 유형의 시험이며, 데이터 수집에 대해 요약하면 다음 [표 1]과 같다.

표 1. 데이터 표본 요약

소속 학과	중간고사		기말고사	
	인원	평균점수	인원	평균점수
약학	1	925.0	1	935.0
의생명과학	61	635.9	59	640.6
간호학	105	627.2	105	672.5
바이오공학	74	551.2	76	572.9
스포츠의학	109	528.2	109	549.2
보건의료산업학	62	518.8	62	542.0
상담심리학	52	512.0	53	552.2
식품생명공학	64	507.5	66	529.3
미술치료학	49	498.5	48	531.2
의료홍보미디어학	82	481.4	80	503.6
데이터경영학	61	465.3	60	516.5
보건복지행정학	70	460.1	71	471.4
전체	790	531.5	789	557.9

2020년 교양 영어 수업 시간에 진행된 다양한 전공과 성적 편차를 보이는 1학년 학생들의 중간고사와 기말고사 성적 자료를 사용하였다. 중간고사와 기말고사에 사용된 문제지는 P 출판사에서 출시된 100문항 토의 유형의 모의시험으로, 중간고사 및 기말고사의 2회에 걸쳐 10개 유형의 각 유형 총합 1,000개의 서로 다른 문항으로 구성되어 있다. 정규 토의는 200문항으로 구성되지만, 동일한 비율과 난이도를 분배하여 본 연구에서는 100문항으로 특수 제작된 시험을 사용했으며, 이 시험은 정규 토의 시험문제를 토대로 만들어졌기 때문에 실제 토의와 비슷한 난이도와 변별도로 출제되었다. 각 유형의 시험은 토의 시험과 마찬가지로 파트 1~4의 듣기영역, 파트5~7의 읽기 영역으로 구성된다. 파트1은 사람등장 사진묘사와 사물묘사 문제로 구성되며 주로 사람 등장 사진 묘사 문제가 70%이상을 차지

한다. 파트1 문제의 난이도는 사람 등장 사진 묘사보다 사물등장 사물 묘사가 조금 더 높은 편이지만 전체 시험에서 봤을 때 가장 쉬운 난이도로 구성되어 있다. 파트2 역시 듣기 파트인데, 질문을 하면 가장 알맞은 응답을 고르는 문제 유형이다. 듣기 영역 시험 점수에 가장 큰 영향을 미치는 파트이며 고도의 순발력이 필요하다. 질문 1문장과 보기 3문장으로 구성되며, 문제지에 아무런 단서가 없어서 순간적으로 방송을 놓치면 실수하기 쉽기 때문에 반복적인 연습이 필요하다. 파트3은 남자와 여자 또는 남자2 여자1, 여자2 남자1로 구성된 대화파트이다. 문제지에 질문과 보기가 있고 한 세트당 3문제로 구성되며 방송을 듣기전 질문과 보기를 미리 읽어두면 도움이 된다. 동의어 표현을 많이 익히고 키워드를 잡는 연습을 하면 단시간에 고득점도 가능한 파트이다. 파트4는 뉴스나 일기예보, 토크쇼 진행, 교통방송 등의 모놀로그 파트이다. 독해 능력이 향상되면 저절로 파트4가 잘들리지만 독해 능력이 저조한 학생에게는 매우 어렵게 느껴질 수도 있다. 파트3과 마찬가지로 한 세트당 3문제로 구성되며 문제지에 문제와 보기가 나와 있어 문제를 먼저 읽어두고 키워드 잡는 연습과 동의어 표현을 정리해 두면 도움이 된다.

RC 영역인 파트5는 문법 문제와 어휘 문제로 구성된 빈칸 채우기 영역이다. 최근 몇 년 사이에 어휘 문제가 강화되어 어려운 단어들도 출제되며, 문법 문제들은 골고루 출제된다. 한 문제당 사용할 수 있는 시간이 10-20초 정도로 짧기 때문에 반복된 연습이 필요하다. 파트6은 파트5와 파트7가 섞여져 있는 느낌의 문법 어휘 문제와 독해가 어우러져 빈칸이 뚫려 있다. 문제의 길이는 짧은 독해지문정도이고 한 세트당 4문제로 구성되며 문법 문제, 어휘 문제, 문맥 파악 문제 등으로 출제된다. 파트7은 독해영역으로 단문독해와 장문독해, 장문독해는 단일 지문, 이중 지문, 삼중 지문으로 골고루 출제되며 최근 몇 년 간 토익 시험의 경향은 지문이 점점 길어지는 추세이다. 따라서 다양한 어휘와 기본적인 문법 지식, 직독직해 연습이 충분이 되어야 제한 시간 안에 문제를 모두 풀 수 있는 오랜시간 집중을 요하는 파트이다.

본 연구에 사용할 코퍼스 구축을 위해서 각 시험의 지문을 문단 및 문장 형식으로 연결하여 컴퓨터 텍스트

로 추출하였으며, 난이도 및 변별도 분석을 위해서 개별 문항의 선지 개수, 정답 번호와 실제 개별 학생들이 선택한 정답 번호를 추출하여 일정한 형식에 따라 배열함으로써 입력 데이터를 수집 및 구성하였다.

2. 코퍼스 지수에 대한 요인 분석

Coh-Matrix의 11개 유형의 지표들은 한 유형 안에 많은 지표들을 포함하고 있기 때문에 Coh-Matrix의 11개 유형, 106개 지표들을 모두 사용할 경우 회귀 분석 결과에 대한 해석이 어려울 가능성이 있다. 또한 11개 유형의 Coh-Matrix 지표들은 각각의 유형 안에서 서로 유사한 개념의 지표들이 중복되어 사용되고 있기 때문에 그것들을 모두 변수로 사용할 경우 독립변수 간의 상관성이 높아지는 다중공선성의 문제가 발생할 가능성이 높다. 따라서 Coh-Matrix를 구성하는 11개 유형, 106개 지표들에 대해서 각 유형별로 요인 분석을 실시하고 전체 106개 지표들을 총 39개 요인으로 축소하여 회귀 분석을 실시하였다.

요인 분석은 Scree Plot 그래프에 나타난 아이겐 값(Eigen Value) 1을 기준으로 그래프의 평탄화 시점을 참조하여 총 요인 수를 결정하였다. 토익 시험의 정답률과 변별도에 대한 회귀 분석에서는 각 요인에 대한 요인 점수를 사용하여 코퍼스 유형별로 각기 다른 요인들이 토익 시험의 정답률과 변별도에 미치는 영향을 분석하고자 하였다. 다음은 주요 코퍼스 지표들에 대한 요인분석결과, 각각의 요인에 대한 성분행렬을 [표 2-표 5]를 통해 확인할 수 있다.

표 2. Descriptives(기초산출치) 성분행렬

	중간고사		기말고사	
	word length	paragraph length	no of word / sentence	sentence length
DESWLsy	.954			
DESWLsyd	.889			
DESWLlt	.854			
DESWLltd	.830			
DESPL		.932		
DESPLd		.887		
DESPC		-.803		
DESSC			.901	
DESSC			.890	
DESWC				.450
DESSL				-.913

표 3. Text Easability Principle Component Scores (문서 용이성) 성분행렬

	성분				
	referential cohesion	narrativity	temporality	deep cohesion	connectivity
PCREFz	.899				
PCVERBp	.859				
PCREFp	.800				
PCVERBz	.770				
PCCNCz	.665				
PCCNCp	.587				
PCSYNz	-.789				
PCSYNp	-.861				
PCNARz		.920			
PCNARp		.934			
PCTEMPz			.879		
PCTEMPp			.839		
PCDCz				.966	
PCDCp				.964	
PCCONNz					.973
PCCONNp					.985

표 4. Lexical Diversity(어휘 다양성) 성분행렬

	성분	
	Lexical diversity, MTLD/VOCD	Lexical diversity, TTR
LDMTLD	.846	
LDVOCD	.931	
LDTTRc		.983
LDTTRa		.842

이러한 요인분석 결과를 토대로 회귀 분석에서 사용되고 있는 코퍼스 요인들은 소속된 코퍼스 유형의 범주를 이해하기 쉽도록 하기 위해서 다음과 같이 요인 명칭 대신 일련번호를 사용하여 변수명으로 사용하였으며, 데이터 수집에서 밝힌 바와 같이 각 시험의 지문을 컴퓨터 텍스트 문서로 추출하여 입력함으로써 각 요인에 대한 결과값을 얻을 수 있도록 하였다.

표 5. 요인 명칭 대응 변수명

범주	요인	요인 명칭
기초산출치	기초산출치 1	word length
	기초산출치 2	paragraph length
	기초산출치 3	no of word, sentence
	기초산출치 4	sentence length
문서용이성	문서용이성 1	referential cohesion
	문서용이성 2	narrativity
	문서용이성 3	temporality
	문서용이성 4	deep cohesion
	문서용이성 5	connectivity
정합성	참조적정합성 1	noun/stem overlap
	참조적정합성 2	content word overlap
	의미적정합성 1	LSA overlap, adjacent sentences, mean
	의미적정합성 2	LSA overlap, all sentences
어휘다양성	어휘다양성 1	lexical diversity,

		MTLD/VOCD
	어휘다양성 2	lexical diversity, TTR
접속사	접속사 1	all/logical/causal connectives incidence
	접속사 2	additive connectives incidence
	접속사 3	temporal connectives incidence
상황모형	상황모형 1	causal/intentional verbs incidence
	상황모형 2	temporal cohesion
	상황모형 3	LSA verb overlap
	상황모형 4	casual particles/ causal verbs ratio
통사적 복잡성	통사적복잡성 1	minimal edit distance
	통사적복잡성 2	left embeddedness
	통사적복잡성 3	sentence syntax similarity
통사적 패턴밀도	패턴밀도 1	noun/verb phrase density
	패턴밀도 2	negation/infinitive density
	패턴밀도 3	adverbial phrase density
	패턴밀도 4	gerund density
어휘정보	어휘정보 1	imagability/concreteness
	어휘정보 2	CELEX word/log frequency
	어휘정보 3	pronoun incidence
	어휘정보 4	familiarity
	어휘정보 5	second person pronoun incidence
	어휘정보 6	first person singular pronoun incidence
	어휘정보 7	hypernymy for verbs
가독성	가독성 1	flesch reading ease
	가독성 2	flesch-kincaid grade level
	가독성 3	coh-matrix L2 readability

2. 문항 분석

2.1 문항 분석의 목적

본 연구에서는 문항 분석을 실시하여 토익 시험의 신뢰도를 분석하고자 하였다. 문항 분석의 목적은 토익의 문항별 정답 및 오답 처리를 통하여 문항별 이상 현상을 분석하는 기법이다. 문항 분석은 여러 시험지 유형 및 다양한 문제들을 통해 토익의 신뢰도를 검증하는데 중요한 요소이다. 본 연구에서 사용하는 문항 분석은 고전 검사 이론에 근거하며 이 이론에 근거한 세부 분석 항목은 다음과 같다.

2.2 문항 분석 방법

다양한 문항 분석 도구 중 본 연구에서는 Ohio University의 TAP(Test Analysis Program)을 사용한다. TAP은 문항의 난이도, 변별도, 신뢰도, 응시자의 성적 분석뿐만 아니라 문항으로써 매력도가 떨어지는

문항을 분석해 준다. 또한, 그 원인을 제시해 주는 기능 등을 포함하여 다양한 문항 분석이 가능한 장점이 있는 문항 분석 도구이다.

TAP을 사용하여 문항의 난이도 및 변별도를 분석하기 위해서 본 연구에서 사용하는 실험 데이터를 TAP이 요구하는 조건에 맞춰 입력하는 과정이 필요하다. 데이터 수집에서 밝힌 바와 같이 중간고사 및 기말고사 10개 유형의 총 1,000문항에 대해 790여명의 응시생이 선택한 개별 답안과 함께 각 문항 선지의 개수, 그리고 실제 답을 TAP이 요구하는 형식에 맞춰 TAP에 입력하여 결과 분석을 얻는 과정을 [그림 1]과 같이 수행하였다.

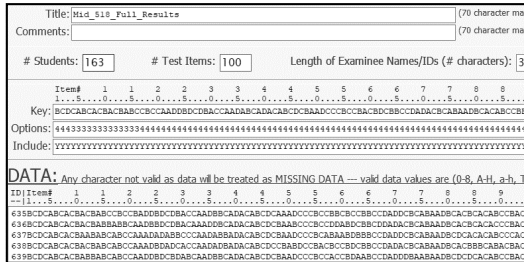


그림 1. TAP 분석을 위한 데이터 입력

2.2.1 문항 난이도 분석

문항 난이도는 단일 문항의 난이도를 계산하며 응시자의 문항 정답률이 상승할수록 이에 비례하여 낮은 문항 난이도를 가진다고 가정한다. 전체 응시자 대비 해당 문항을 맞힌 응시자의 비율을 통해 개별 문항의 난이도를 계산할 수 있으며 그 계산식은 아래와 같다.

$$P = \frac{R}{N} \tag{1}$$

- P : 문항 난이도
- N : 응시자의 수
- R : 문항을 맞힌 응시자의 수

문항 난이도의 계산 결과값은 0과 1사이의 값을 가지며, 0에 가까울수록 어려운 문항, 1에 가까울수록 쉬운 문항으로 분류된다. Cangelosi[48]은 이 문항 난이도 지수가 0.25미만이면 어려운 문항, 0.25이상 0.75미만이면 적절한 난이도의 문항 그리고 0.75 이상이면 쉬운 문항으로 구분하였다. 하지만 응시자의 응시 능력 및

환경, 그리고 출제 유형에 따라 서로 다른 문항이 같은 문항 난이도 지수를 가진다고 해도 정확히 같은 난이도의 문제라고는 해석하기 어렵다. 그러므로 문항 난이도의 정확한 측정을 위해서는 개별 문항을 해결하는 데 주어지는 시간, 응시자 집단, 시험 환경 등의 외부 요소를 엄격히 통제해야 한다.

2.2.2 문항 변별도 분석

문항 변별도는 응시자들의 성취도 그룹에 따른 특정 문항에서 정답률의 차이를 의미한다. 구체적으로 모든 문항에 대해서 높은 성취도를 가진 집단이 낮은 성취도를 가진 집단에 비해 특정 문항에 대해 얼마만큼의 정답률의 차이를 보이는지 분석하는 방법이다. 아래 수식과 같이 상하 두 집단 간의 정답을 맞춘 응시자 수의 차를 통해 계산이 가능하다.

$$D = \frac{R_h - R_l}{f} \tag{2}$$

- D : 문항 변별도
- R_h : 높은 성취도 집단의 정답 수
- R_l : 낮은 성취도 집단의 정답 수
- f : 두 집단의 응시자 수

계산된 개별 문항의 변별도 값은 -1과 1사이의 값을 가지며, -1에 가까울수록 낮은 문항 변별도를, 1에 가까울수록 높은 문항 변별도를 가진다고 해석할 수 있다. Ebel[49]은 문항 변별도 지수가 0.2미만이면 변별도가 없고, 0.2이상 0.3미만이면 낮은 변별도, 0.3이상 0.4미만이면 보통 변별도, 0.4 이상이면 높은 변별도로 변별도 지수에 따른 해석을 정의하였다. 그러나 문항 변별도 또한 문항 난이도와 같이 시험 유형 및 출제 경향에 따라 구간별 해석에 차이가 있을 수 있다.

문항 변별도 지수는 0이하의 값이 계산될 수도 있는데, 이는 낮은 성취도 그룹의 정답률이 높은 성취도 그룹의 정답률보다 높다는 것을 의미하므로, 이는 해당 문제의 변별도에 크게 문제가 있음을 의미한다. 이러한 문항 변별도 또한 개별 문항을 해결하는 데 주어지는 시간, 응시자 집단, 시험 환경 등의 외부 요소에 따라 크게 변동될 수 있기 때문에 이러한 외부 요소를 엄격히 통제해야 정확한 문항 변별도 측정이 가능하다.

IV. 분석 결과

1. 문항 분석

1.1 문항 난이도 분석 결과

TAP 문항 분석 프로그램을 사용하여 분석에 사용된 객관식 문항은 10개 유형의 총 1,000개 문항이다. 분석 대상이 되는 문항들은 정규 토익과 같은 유형의 하프 토익(100문항) 모의시험으로 각 유형은 파트 1부터 파트 7까지 총 7개의 파트로 구성되며 듣기 영역인 파트 1~4는 각각 3, 12, 20, 15개 정도의 문항 개수로 구성되며, 읽기 영역인 파트 5~7은 각각 15, 8, 27개 정도의 문항 개수로 구성된다.

표 6. 파트별 평균 문항 난이도

구분	문항수	평균 문항 난이도	난이도 표준편차	
LC	Part 1	3	0.786	0.07125
	Part 2	12	0.693	0.05802
	Part 3	20	0.629	0.05340
	Part 4	15	0.570	0.03911
RC	Part 5	15	0.563	0.06379
	Part 6	8	0.530	0.08112
	Part 7	27	0.531	0.06498

파트별 문항 난이도의 경우 위의 [표 6]과 같은 분포를 보이는데, 시험의 앞부분에 위치한 파트 1에서 파트 4까지 듣기 영역의 경우 평균 문항 난이도는 0.786에서 0.570이었으며, 뒷부분에 위치한 파트 5부터 파트 7까지 읽기 영역의 경우 평균 문항 난이도는 0.563에서

0.531이었다. 이를 통해 읽기 영역이 듣기 영역 보다 다소 어려운 문항이 포함되어 있음을 알 수 있다. 반대로 듣기 영역의 경우 0.6이상은 물론 파트 1의 0.786의 평균 문항 난이도가 측정됨으로써 매우 쉬운 문항이 다수 포함되어 있는 것을 확인할 수 있다.

1.2 문항 변별도 분석 결과

문항 변별도 분석은 시험 유형별, 파트별 문항 변별도와 하위 성취도 그룹이 상위 성취도 그룹보다 높은 정답률을 보이는 예외적인 경우를 찾아 분석하였다.

문항 변별도는 문항 분석 프로그램인 TAP에서 D-index 지표로 분석 결과를 제공한다. TAP에서 제공하는 D-index 지표는 상위 서술한 바와 같이 상위 그룹의 정답 빈도에서 하위 그룹의 정답 빈도를 뺀 값의 수를 나눈 산출하는 방식으로 상위 그룹과 하위 그룹 간의 상호 변별도를 판단할 수 있다. 표 8은 총 10개 유형의 시험에 대한 문항 변별도 분석 결과를 파트별로 분류해 나타낸 것으로, Ebel의 문항 변별도 분류에 따르면 파트 1의 변별도는 다른 모든 파트에 비해 가장 낮으며, RC파트에서는 파트 5의 변별도가 가장 낮은 것을 [표 7]을 통해 확인할 수 있다.

1.3 파트별 정답률 및 변별도 코퍼스 지표 분석

정답률은 응시자들의 시험 수행 능력을 의미하므로 정답률이 낮은 경우 그 원인이 되는 코퍼스 지표들을

표 7. 파트별 정답률과 변별도의 기술통계

	N	평균	표준편차	표준오차	평균에 대한 95% 신뢰구간		최소값	최대값	
					하한	상한			
정답률	1	10	0.786	0.071	0.023	0.735	0.837	0.680	0.893
	2	10	0.694	0.058	0.018	0.652	0.735	0.582	0.772
	3	10	0.627	0.053	0.017	0.589	0.665	0.539	0.723
	4	10	0.570	0.039	0.012	0.541	0.598	0.507	0.646
	5	10	0.563	0.064	0.020	0.518	0.609	0.493	0.699
	6	10	0.530	0.088	0.028	0.467	0.593	0.364	0.659
	7	10	0.530	0.062	0.020	0.486	0.575	0.440	0.598
	전체	70	0.614	0.108	0.013	0.588	0.640	0.364	0.893
변별도	1	10	0.280	0.129	0.041	0.188	0.372	0.043	0.473
	2	10	0.399	0.066	0.021	0.351	0.446	0.240	0.458
	3	10	0.380	0.068	0.022	0.331	0.429	0.298	0.493
	4	10	0.379	0.044	0.014	0.348	0.410	0.313	0.457
	5	10	0.347	0.050	0.016	0.312	0.383	0.233	0.430
	6	10	0.388	0.064	0.020	0.343	0.434	0.313	0.488
	7	10	0.397	0.058	0.018	0.356	0.439	0.331	0.473
	전체	70	0.367	0.010	0.010	0.348	0.386	0.043	0.493

찾아내서 그 지표와 관련된 교수법 개발이 필요하다고 볼수 있다. 또한 변별도의 경우 상, 하 레벨의 차이를 의미하므로 변별도가 높은 경우 그 원인이 되는 코퍼스 지표가 무엇인지 찾아내고 그 코퍼스 지표와 관련하여 하위 그룹들이 특별히 대비할 수 있는 방법을 찾는 것이 중요하다고 할 수 있다.

따라서 본 연구에서는 파트별 정답률과 변별도의 차이가 얼마나 되고, 그 차이의 원인이 되는 코퍼스 지표들과 그 상관성을 분석하고자 하였다. 먼저 각 파트별로 정답률과 변별도의 차이가 존재하는지를 분석하기 위하여 일원배치 ANOVA(Analysis of Variance)를 실시하였다.

표 8. 파트별 정답률과 변별도 차이 ANOVA 분석

구분	제곱합	자유도	평균 제곱	F	유의 확률	
정답률	집단-간	0.546	6	0.091	22.304	0.000
	집단-내	0.257	63	0.004	-	
	전체	0.803	69	-	-	
변별도	집단-간	0.107	6	0.018	3.318	0.007
	집단-내	0.338	63	0.005	-	
	전체	0.445	69	-	-	

[표 8]의 ANOVA 분석 결과에서 나타난 각 파트별 정답률과 변별도의 차이가 주로 어떤 파트 간의 차이 인지를 분석하기 위하여 사후 검정을 실시하였다. 사후 검정은 Tukey 검정을 실시하고, 두 검정 결과를 비교하고자 하였다.

먼저 정답률에 대한 Tukey 검정을 실시한 결과 아래 [표 9]와 같은 검정 결과를 도출하였다. 정답률의 경우 파트 1은 다른 모든 파트보다 높은 것으로 나타났고, 파트 2는 파트 1보다는 낮고 파트 4, 5, 6 7보다는 높은 것으로 나타났다. 파트 3은 파트 1보다는 낮고, 파트 6, 7보다 높게 나타났다. 파트 4와 5는 파트 1, 2 보다 낮은 것으로 나타났고, 파트 6과 7은 파트 1, 2, 3보다 낮은 것으로 나타났다.

표 9. 파트별 정답률 차이에 대한 Tukey 사후 검증 결과

종속변수	기준파트(I)	비교파트 (J)	평균차이 (I-J)	표준오차	유의확률
정답률 (Tukey HSD)	1	2	0.091900*	0.029	0.032
		3	0.158700*	0.029	0.000
		4	0.216100*	0.029	0.000
		5	0.222200*	0.029	0.000
		6	0.255400*	0.029	0.000

	2	7	0.255500*	0.029	0.000
		1	-0.091900*	0.029	0.032
		3	0.067	0.029	0.242
		4	0.124200*	0.029	0.001
		5	0.130300*	0.029	0.000
		6	0.163500*	0.029	0.000
		7	0.163600*	0.029	0.000
	3	1	-0.158700*	0.029	0.000
		2	-0.067	0.029	0.242
		4	0.057	0.029	0.419
		5	0.064	0.029	0.298
		6	0.096700*	0.029	0.020
		7	0.096800*	0.029	0.020
		4	1	-0.206100*	0.029
	2		-0.124200*	0.029	0.001
	3		-0.057	0.029	0.419
	5		0.006	0.029	1.000
	6		0.039	0.029	0.812
	7		0.039	0.029	0.810
	5		1	-0.222200*	0.029
		2	-0.130300*	0.029	0.000
3		-0.064	0.029	0.298	
4		-0.006	0.029	1.000	
6		0.033	0.029	0.905	
7		0.033	0.029	0.904	
6		1	-0.255400*	0.029	0.000
	2	-0.163500*	0.029	0.000	
	3	-0.096700*	0.029	0.020	
	4	-0.039	0.029	0.812	
	5	-0.033	0.029	0.905	
	7	0.000	0.029	1.000	
	7	1	-0.255500*	0.029	0.000
2		-0.163600*	0.029	0.000	
3		-0.096800*	0.029	0.020	
4		-0.039	0.029	0.810	
5		-0.033	0.029	0.904	
6		0.000	0.029	1.000	

다음은 변별도의 파트별 차이를 검증하기 위한 Tukey 사후 분석을 [표 10]과 같이 수행하였다. 파트 1의 변별도는 파트 4와 5를 제외한 모든 변인들에 대해서 변별도가 낮은 것으로 나타났고, 파트 2와 3은 파트 1에 대해서만 변별도가 높고 다른 파트와는 유의한 차이가 없는 것으로 나타났다. 파트 4와 5는 다른 파트와 유의한 차이가 없는 것으로 나타났고, 파트 6과 7은 파트 1에 대해서만 변별도가 더 높고 다른 파트와는 유의한 차이가 없는 것으로 나타났다.

표 10. 파트별 변별도 차이에 대한 Tukey 사후 검증 결과

종속변수	기준파트(I)	비교파트 (J)	평균차이 (I-J)	표준오차	유의확률
변별도 (Tukey HSD)	1	2	-0.119000*	0.033	0.010
		3	-0.100500*	0.033	0.047
		4	-0.099	0.033	0.052

		5	-0.067	0.033	0.392
		6	-0.108300*	0.033	0.025
		7	-0.117300*	0.033	0.011
	2	1	0.119000*	0.033	0.010
		3	0.019	0.033	0.998
		4	0.020	0.033	0.997
		5	0.052	0.033	0.697
		6	0.011	0.033	1.000
		7	0.002	0.033	1.000
		3	1	0.100500*	0.033
	2		-0.019	0.033	0.998
	4		0.001	0.033	1.000
	5		0.033	0.033	0.949
	6		-0.008	0.033	1.000
	7		-0.017	0.033	0.999
	4	1	0.099	0.033	0.052
		2	-0.020	0.033	0.997
		3	-0.001	0.033	1.000
		5	0.032	0.033	0.957
		6	-0.009	0.033	1.000
		7	-0.018	0.033	0.998
5		1	0.067	0.033	0.392
	2	-0.052	0.033	0.697	
	3	-0.033	0.033	0.949	
	4	-0.032	0.033	0.957	
	6	-0.041	0.033	0.871	
	7	-0.050	0.033	0.728	
	6	1	0.108300*	0.033	0.025
2		-0.011	0.033	1.000	
3		0.008	0.033	1.000	
4		0.009	0.033	1.000	
5		0.041	0.033	0.871	
7		-0.009	0.033	1.000	
7	1	0.117300*	0.033	0.011	
	2	-0.002	0.033	1.000	
	3	0.017	0.033	0.999	
	4	0.018	0.033	0.998	
	5	0.050	0.033	0.728	
	6	0.009	0.033	1.000	

사후 검증 결과를 종합하면 정답률은 파트 1, 2에서 높게 나타나고 파트 3은 중간 정도이며, 파트 4, 5, 6, 7은 상대적은 낮은 것으로 나타났다. 변별도는 파트 1의 경우 다른 파트보다 낮은 것으로 나타났을 뿐 다른 파트 간에는 변별도 차이가 없는 것으로 나타났다. 정답률이 파트 1, 2에서 높게 나타나고 있는 것은 이 두 파트가 다른 파트보다 비교적 난이도가 낮다는 증거이다. 이는 토익의 파트별 난이도 배분에 있어서 불균등이 발생할 가능성이 있고 그 결과 응시자의 특정한 능력에 따른 파트별 점수의 풀림 현상이 나타날 가능성이 높은 것으로 판단된다. 변별도의 경우 파트 1은 파트 4, 5를 제외한 파트 6, 7 등 보다 더 낮게 나타나고 있고, 다른 파트들 간에는 별다른 차이를 보이지 않고 있다.

이는 파트 1의 경우 영어 능력이 하위 그룹에 속하는 응시자들에게 상대적으로 유리한 문항들이 있다는 증거이다. 반면에 파트 6, 7 등은 파트 1보다 변별도가 더 높은 것으로 나타나 영어 능력 하위 그룹에 속하는 응시자들에게 추가적인 교육이 필요하다는 점을 말해 주고 있다.

2. 코퍼스 지표에 대한 회귀 분석 결과

정답률과 변별도는 토익 시험의 유용성을 결정하는 중요한 요인이기 때문에 본 연구에서는 정답률과 변별도에 영향을 주는 코퍼스 요인들이 무엇인지 알아보기 위하여 정답률, 변별도, 코퍼스 지표 간의 회귀 분석을 실시하였다.

회귀 분석의 목적을 좀 더 구체적으로 제시하면 다음과 같다. 첫째, 정답률에 정의 영향을 주는 코퍼스 지표와 음의 영향을 주는 코퍼스 지표들을 따로 구분할 경우 그 지표들이 토익 성적 향상에 직접적인 영향을 줄 수 있는 아주 중요한 지표이기 때문에 영어 교육의 목표를 설정하는 데 도움을 줄 수 있을 것이다. 특히 정답률과 역의 인과관계를 갖는 코퍼스 지표들은 토익 시험을 준비하는 응시자들의 교육 프로그램 내용을 결정하는 데 있어서 주로 어떤 파트의 어떤 문항들을 대상으로 중점 교육을 실시해야 하는지 많은 정보들을 제공해 줄 수 있을 것이다. 둘째, 변별도에 정의 영향을 주는 코퍼스 요인들의 경우 하위 그룹에 속하는 토익 응시자들에게는 파트별로 집중적인 교육이 필요한 부분이 무엇인지 알려주는 단서로 작용하게 될 것이다. 따라서 변별도와 정의 인과관계를 갖는 코퍼스 지표들이 무엇인지 알게 되면 하위 그룹들의 시험 성적 향상을 위한 교육 프로그램 개발에 대한 시사점들을 제공해 줄 수 있을 것이다.

본 연구는 이러한 파트별 정답률과 변별도의 차이를 가져오는 원인이 코퍼스 지표인 것으로 가정하고 정답률과 변별도에 대한 코퍼스 지표들의 인과관계를 분석하였다.

먼저 문항 분석결과 정답률에 대한 코퍼스 요인들의 영향력을 [표 11]과 같이 분석하였다. 기초산출치의 4개 요인들에 대해서 정답률과 변별도에 대한 인과관계를 분석하였다. 정답률에 대해서는 word_length(b=-.518,

표 11. 문항 분석 결과 정답률에 대한 코퍼스 지표 회귀 분석 결과

종속변수 : 정답률		비표준화 계수		표준화 계수	t	유의 확률	공선성 통계량		R ²
		β	표준오차	베타			공차	VIF	
기초산출치	(상수)	.614	.008		80.352	.000			0.669
	word_length	-.056	.008	-.518	-7.254	.000	1.000	1.000	
	no_word_sentence1	-.036	.008	-.334	-4.678	.000	1.000	1.000	
	paragraph_length	-.005	.008	-.044	-0.622	.536	1.000	1.000	
	sentence_length	-.058	.008	-.536	-7.506	.000	1.000	1.000	
문서용이성	(상수)	.614	.009		65.039	.000			0.51
	referential_cohesion	.048	.010	.447	5.063	.000	1.000	1.000	
	temporality	.018	.010	.167	1.890	.063	1.000	1.000	
	connectivity	.020	.010	.185	2.101	.040	1.000	1.000	
	narrativity	.023	.010	.215	2.436	.018	1.000	1.000	
	syntactic_simplicity	.036	.010	.338	3.836	.000	1.000	1.000	
	deep_cohesion	.032	.010	.296	3.359	.001	1.000	1.000	
정합성	(상수)	.614	.011		56.079	.000			0.32
	content_word_overlap	.063	.013	.584	4.729	.000	.686	1.457	
	noun_stem_overlap	.010	.015	.092	.641	.524	.513	1.950	
	LSA_overlap_adjacent_sentences	-.058	.017	-.541	-3.444	.001	.423	2.362	
	LSA_overlap_sall_sentences	-.018	.011	-.165	-1.583	.118	.957	1.045	
접속사	(상수)	.614	.010		59.691	.000			0.391
	all_logical_causal_connectives_incidence	-.065	.010	-.604	-6.289	.000	1.000	1.000	
	connectives_incidence	.006	.010	.059	.617	.539	1.000	1.000	
	additive_connectives_incidence	-.016	.010	-.148	-1.536	.129	1.000	1.000	
상황모형	(상수)	.614	.010		61.045	.000			0.426
	casual_intentional_verbes_incidence	.058	.010	.542	5.765	.000	1.000	1.000	
	casual_particules_causal_verbs_ratio	-.036	.010	-.337	-3.586	.001	1.000	1.000	
	WordNet_verb_overlap	.015	.010	.136	1.445	.153	1.000	1.000	
	LSA_verb_overlap	-.003	.010	-.026	-.277	.783	1.000	1.000	
통사적 복잡성	(상수)	.614	.009		65.270	.000			0.49
	minimal_edit_distancel	-.025	.009	-.233	-2.657	.010	1.000	1.000	
	sentence_syntax_similarity	.053	.009	.487	5.546	.000	1.000	1.000	
	left_embeddedness	-.048	.009	-.445	-5.065	.000	1.000	1.000	
통사적 패턴밀도	(상수)	.614	.010		61.749	.000			0.439
	infinitive_density	.009	.010	-.588	-6.004	.000	1.000	1.000	
	preposition_phrase_density	.061	.010	-.349	-3.757	.000	1.000	1.000	
	noun_phrase_density	.023	.010	.077	.829	.410	1.000	1.000	
	negation_density	.053	.010	.017	.181	.857	1.000	1.000	
어휘정보	(상수)	.614	.008		74.155	.000			0.611
	person1_2_pronoun_incidence	.009	.008	.087	1.125	.265	1.000	1.000	
	familiarity_imagability_concreteness	.061	.008	.565	7.300	.000	1.000	1.000	
	person3_pronoun_incidence	.023	.008	.215	2.780	.007	1.000	1.000	
	hypernymy_for_nouns	.053	.008	.488	6.312	.000	1.000	1.000	
가독성	(상수)	.614	.009		64.702	.000			0.466
	readability	.074	.010	.682	7.696	.000	1.000	1.000	

p<.001), no_word_sentence1 (b=-.334, p<.001), sentence_length (b=-.536, p<.001) 등이 음의 영향을 주는 것으로 나타났다(p<.001). 문서용이성 지표들 중에서 referential_cohesion (b=.447, p<.001), connectivity(b=.185, p<.05), narrativity (b=.215, p<.05), syntactic_simplicity(b=.338, p<.001), deep_cohesion(b=.296, p<.001) 등이 정의 영향을

미치는 것으로 나타났다. 정합성 지표들 중에서 content_word_overlap은 정의 영향을 미치고 (b=.584, p<.001), LSA_overlap_adjacent_sentences는 음의 영향을 미치는 것으로 나타났다 (b=-.541, p<.001).

어휘 다양성 지표들 중에서 TTR은 정의 영향을 미치고(b=.241, p<.05), MTLD_VOCD은 음의 영향을 주

는 것으로 나타났다($b=-.613, p<.001$). 접속사 지표들 중에서 All_logical_causal_connectives_incidence 이 음의 영향을 미치는 것으로 나타났다($b=-.604, p<.001$). 상황모형 지표들 중에서 casual_Intentional_verbs_incidence는 정의 영향을 미치고($b=.542, p<.001$), casual_particles_causal_verbs_Ratio는 음의 영향

을 미치는 것으로 나타났다($b=-.337, p<.001$).

통사적복잡성 지표 중에서 Minimal_Edit_Distance1 ($b=-.233, p<.01$), Left_embeddedness ($b=-.445, p<.001$) 등은 음의 영향을 미치고, Sentence_syntax_similarity는 정의 영향을 미치는 것으로 나타났다 ($b=.487, p<.001$). 통사적 패턴밀도 지표들 중에서

표 12. 문항 분석 결과 변별도에 대한 코퍼스 지표 회귀 분석 결과

종속변수 : 정답률		비표준화 계수		표준화 계수	t	유의 확률	공선성 통계량		R ²
		β	표준오차	베타			공차	VIF	
기초산출치	(상수)	.367	.009		39.628	.000			0.122
	word_length	.006	.009	.079	.678	.500	1.000	1.000	
	no_word_sentence1	.018	.009	.219	1.884	.064	1.000	1.000	
	paragraph_length	.018	.009	.220	1.889	.063	1.000	1.000	
	sentence_length	.011	.009	.139	1.197	.236	1.000	1.000	
문서용이성	(상수)	.367	.009		41.413	.000			0.221
	referential_cohesion	-.028	.009	-.347	-3.123	.003	1.000	1.000	
	temporality	.004	.009	.047	.418	.677	1.000	1.000	
	connectivity	-.015	.009	-.184	-1.651	.104	1.000	1.000	
	narrativity	.009	.009	.112	1.011	.316	1.000	1.000	
	syntactic_simplicity	-.004	.009	-.044	-.392	.696	1.000	1.000	
deep_cohesion	-.018	.009	-.222	-1.999	.050	1.000	1.000		
정합성	(상수)	.367	.010		38.428	.000			0.066
	content_word_overlap	-.016	.012	-.205	-1.415	.162	.686	1.457	
	noun_stem_overlap	-.010	.013	-.121	-.725	.471	.513	1.950	
	LSA_overlap_adjacent_sentences	.011	.015	.137	.742	.461	.423	2.362	
	LSA_overlap_all_sentences	.018	.010	.218	.218	.080	.957	1.045	
어휘다양성	(상수)	.367	.009		41.119	.000			0.159
	TTR	-.020	.009	-.249	-2.224	.030	1.000	1.000	
	MTLD_VOCD	.025	.009	.312	2.782	.007	1.000	1.000	
접속사	(상수)	.367	.009		42.235	.000			0.215
	all_logical_causal_connectives_incidence	.031	.009	.380	3.483	.001	1.000	1.000	
	connectives_incidence	-.005	.009	-.058	-.535	.595	1.000	1.000	
	additive_connectives_incidence	.021	.009	.259	2.378	.020	1.000	1.000	
상황모형	(상수)	.367	.009		39.789	.000			0.129
	casual_intentional_verbs_incidence	-.019	.009	-.234	-2.017	.048	1.000	1.000	
	casual_particles_causal_verbs_ratio	.003	.009	.032	.277	.782	1.000	1.000	
	WordNet_verb_overlap	-.021	.009	-.264	-2.277	.026	1.000	1.000	
	LSA_verb_overlap	-.005	.009	-.061	-.531	.597	1.000	1.000	
통사적 복잡성	(상수)	.367	.009		41.238	.000			0.177
	minimal_edit_distance1	.016	.009	.200	1.790	.078	1.000	1.000	
	sentence_syntax_similarity	-.030	.009	-.369	-3.306	.002	1.000	1.000	
	left_embeddedness	-.001	.009	-.013	-.119	.906	1.000	1.000	
통사적 패턴밀도	(상수)	.367	.009		43.189	.000			0.261
	infinitive_density	.040	.009	.496	4.650	.000	1.000	1.000	
	preposition_phrase_density	.002	.009	.022	.210	.834	1.000	1.000	
	noun_phrase_density	.005	.009	.059	.556	.580	1.000	1.000	
	negation_density	.008	.009	.103	.966	.338	1.000	1.000	
어휘정보	(상수)	.367	.009		42.967	.000			0.253
	person1_2_pronoun_incidence	.024	.009	.295	2.752	.008	1.000	1.000	
	familiarity_imagability_concreteness	-.023	.009	-.291	-2.711	.009	1.000	1.000	
	person3_pronoun_incidence	-.022	.009	-.269	-2.514	.014	1.000	1.000	
	hypernymy_for_nouns	-.008	.009	-.094	-.875	.385	1.000	1.000	
가독성	(상수)	.367	.010		38.389	.000			0.021
	readability	-.012	.010	-.145	-1.207	.232	1.000	1.000	

Infinitive_density ($b=-.558, p<.001$), Preposition_phrase_density ($b=-.349, p<.001$) 등이 음의 영향을 미치는 것으로 나타났다.

어휘정보 지표 중에서 Familiarity_Imagability_Concreteness ($b=.565, p<.001$), person3_pronoun_incidence ($b=.215, p<.01$), Hypernymy_for_nouns ($b=.488, p<.001$) 등이 정답률에 정의 영향을 미치는 것으로 나타났다. 마지막으로 가독성 지표 Readability는 정의 영향을 미치는 것으로 나타났다($b=.682, p<.001$).

다음은 문항 분석 결과 변별도에 대한 코퍼스 지표들의 영향력을 분석하였으며 그 결과는 [표 12]와 같다.

먼저 기초산출치의 4개 요인들에 대해서 변별도에 대한 인과관계를 분석한 결과 모두 유의하지 않은 것으로 나타났다. 문서용이성 지표 중에서는 referential_cohesion ($b=-.347, p<.01$), deep_cohesion ($b=-.222, p<.05$) 등이 음의 영향을 미치는 것으로 나타났다. 정합성 지표들에 대해서는 모두 변별도에 대해 유의한 영향을 미치지 못하는 것으로 나타났다.

어휘 다양성 지표들 중에서 TTR은 음의 영향을 미치고 ($b=-.249, p<.05$), MTLD_VOCD는 정의 영향을 주는 것으로 나타났다($b=.312, p<.01$). 접속사 지표 중에서 All_logical_causal_connectives_incidence($b=.380, p<.001$), Additive_connectives_incidence ($b=.259, p<.05$) 등이 정의 영향을 미치는 것으로 나타났다. 상황모형 지표들 중에서 casual_Intentional_verbs_incidence ($b=-.234, p<.05$), WordNet_verb_overlap ($b=-.264, p<.05$) 등의 지표들이 음의 영향을 미치는 것으로 나타났다.

통사적 복잡성 지표 중에서 Sentence_syntax_similarity가 음의 영향을 미치는 것으로 나타났다 ($b=-.369, p<.01$). 통사적 패턴밀도 지표들 중에서 Infinitive_density ($b=.496, p<.001$)만이 정의 영향을 미치는 것으로 나타났다.

어휘정보 지표 중에서 person1_2_pronoun_incidence($b=.295, p<.01$)는 정의 영향을 미치고, Familiarity_Imagability_Concreteness($b=-.291, p<.01$), person3_pronoun_incidence ($b=-.269, p<.05$) 등은 변별도에 음의 영향을 미치는 것으로 나타

났다. 마지막으로 가독성 지표 Readability는 변별도에 유의한 영향을 미치지 못하는 것으로 나타났다.

V. 결론 및 논의

1. 분석 결과 요약

본 연구에서는 문항 분석 및 코퍼스 지표에 대한 회귀 분석을 통해 토익 시험의 정답률과 변별도에 영향을 미치는 코퍼스 지표와 그 영향력에 대해서 분석하고자 하였다. 분석 결과를 요약하면 다음과 같다.

첫 번째, 문항 분석 결과 시험 유형별 평균 문항 난이도는 큰 편차 없이 보통 난이도를 보여 주고 있지만 변별력이 떨어져서 최상위권 학생과 상위권 학생, 중위권 학생을 구분하기가 어려울 수 있고 유형별로 난이도가 상이하여 신뢰도에 좋지 않은 영향을 주고 있는 것을 확인할 수 있었다. 변별력이 저하될 경우 토익 시험 응시자들의 학습 목표가 불확실해지고 성적에 대한 예측 가능성이 낮아져서 의사소통 능력의 측정 도구로서 토익의 신뢰도가 낮다고 볼 수 있을 것이다. 또한 듣기 영역 대비 읽기 영역이 낮은 문항 난이도를 가짐으로써, 읽기 영역에 다소 어려운 문항이 포함되어 있음을 알 수 있다. 독해 파트 후반부로 갈수록 지문이 길어지고 한 지문당 여러 문제를 풀어야 하기 때문에 시간도 많이 걸리고 집중력도 떨어지는 등 난이도 증가 현상이 나타나고 있다. 또한 듣기 영역과 읽기 영역, 각 파트별 난이도 차이가 크게 나타나고 있어서 신뢰도를 떨어뜨리는 원인이 되고 있다는 점을 발견하였다. 이는 파트별 정답률 및 변별도에 대한 코퍼스 지표 분석에서도 또한 마찬가지로 나타나는 현상인데, Tukey 검증 결과 파트가 뒤로 갈수록 어려워지는 것을 확인할 수 있었다. 이는 읽기 영역에 어려운 문제가 집중되어 있고 듣기 영역 파트 1에는 쉬운 문제가 밀집되어 있다는 걸 의미하는데, 이러한 구성은 전반적인 토익 시험의 신뢰도를 떨어뜨리는 원인이 된다고 할 수 있다.

변별도 분석 결과에서는 문항 난이도에서 모든 유형의 문항 난이도 지수가 전반적으로 높게 나와 대다수의 문항이 낮은 변별도를 갖는 것으로 나타났다. 변별도가 낮다는 것은 신뢰도가 낮다는 뜻이기 때문에 전체적으

로 쉬운 난이도가 계속 유지된다면 평가의 신뢰도가 떨어질 수 있음을 알 수 있다. 하위권 학생들의 정답률이 상위권 학생들의 정답률보다 높은 경우와 같이 잘못 출제된 문항이 다수 있는 것으로 분석되었는데, 이러한 문항이 포함된 시험은 시험 자체의 신뢰도를 떨어뜨려 어학 능력의 정확한 평가를 불가능하게 만드는 원인이 될 수 있다.

두 번째, 코퍼스 지표에 대한 회귀 분석 결과 지표에 따라서 토익 시험의 정답률과 변별도에 미치는 영향력이 서로 다르고 코퍼스 지표 관리를 통하여 토익 시험의 유용성을 향상시킬 수 있다는 사실이 밝혀졌다.

기존의 연구 결과에서 코퍼스 분석은 난이도에 영향을 줄 수 있는 어휘와 문장의 구조적인 특성을 파악할 수 있는 도구인 것으로 밝혀졌다. 그러나 정확히 코퍼스의 어떤 요인이 어떤 영향을 미치는지 에 대해서는 검증된 바가 없다. 또한 코퍼스를 통한 유용성 분석은 영어 평가 시험의 구조적인 특성을 파악하고, 평가 시험의 어휘 수준을 측정하여 난이도와 변별력 수준을 분석하는데 그치고 있다[2]. 따라서 본 연구에서 밝혀진 바와 같이 코퍼스 요인들이 평가 시험의 난이도와 변별도에 어떤 영향을 주는지에 대한 검증 결과를 토대로 평가 시험의 유용성 증대에 도움을 줄 수 있을 것이다. 특히 다양한 평가 시험의 난이도 및 변별도 조절과 시험에 대한 학습법에 대한 시사점을 제공할 수 있을 것이다.

특정 코퍼스 지표들은 토익의 정답률에 음의 영향을 미치고 그 결과 성적을 저하시키는 요인으로 작용하고 있다. 문장의 길이, 워드 수 등 기본적인 산출치가 많거나, 어휘 다양성과 접속사 등과 관련된 코퍼스 요인들의 증가는 정답률을 낮추는 것으로 나타났다. 따라서 이러한 코퍼스 요인들에 대한 학습 프로그램을 개발하여 많은 학습자들에게 정답율을 향상시킬 수 있는 유용한 정보를 제공해 줄 수 있을 것이다. 또한 변별도의 회귀 분석 결과에서 나타난 바와 같이 변별도에 정의 영향을 주는 지표들은 성적을 떨어뜨리는 원인이 된다. 따라서 하위 그룹들의 성적 저하를 가져오는 코퍼스 요인들은 효과적인 교육 프로그램을 개발하는 데 중요한 정보를 제공할 수 있을 것이다.

2. 시사점 및 연구 한계

본 연구에서 밝혀진 정답률과 변별도에 대한 코퍼스 요인들의 인과관계 분석 결과가 시사하는 바는 다음과 같다. 정답률은 시험 난이도의 결과로써 정답률에 역의 영향을 주는 코퍼스 요인들을 반영하여, 교육적 성과를 높일 수 있는 영어 학습 프로그램 개발에 활용할 수 있을 것이다. 그러나 본 연구는 실제 토익 시험 문항에 대한 수집이 어려운 국내 현실로 인하여 모의 토익시험을 대상으로 분석하고 있다.

본 연구에서는 모의 토익 시험에 대한 정답률과 변별도를 사용하고 있으므로, 연구 결과를 전체 토익 시험에 대해서 일률적으로 적용할 수는 없을 것이다. 또한 모의 토익 시험의 일부 유형에 대해서만 분석을 하고 있고 문항 수도 실제 토익과 차이가 있는 관계로 다른 유형의 시험에 대해서는 코퍼스 지표의 영향력이 다르게 나올 수도 있을 것이다. 단지 본 연구 결과는 평가 시험의 유용성을 결정하는 요인으로써 코퍼스 지표의 역할과 기능에 대해서 밝히고 있다는 점이 유용하다고 할 수 있다. 또한 정답률과 변별도에 의한 실제 성적의 결과물에 대해서는 연구 범위를 벗어나고 있다. 따라서 정답률과 변별도의 차이가 실제 성적에 어떻게 나타나고 있고 이에 대한 코퍼스 지표의 영향력을 분석 대상으로 하지 않고 있다. 추후에는 실제 성적에 대한 정답률과 변별도의 영향력을 분석하여 코퍼스 지표에 의한 난이도 조절이 과연 가능한지에 대한 심도 깊은 연구가 필요할 것이다. 이러한 연구는 평가 시험의 유용성을 높이는 도구로써 코퍼스 지표의 활용 가치에 대해서 많은 시사점을 제공할 수 있을 것이다.

참고 문헌

- [1] C. W. Liao, Y. Qu, and R. Morgan, "The relationships of test scores measured by the TOEIC listening and reading test and TOEIC speaking and writing tests," *TOEIC Compendium Study*, Vol.13, pp.1-15, 2010.
- [2] Y. N. Lee, *A validation study of TOEIC based on test usefulness*, Unpublished doctoral dissertation, Korea University, 2020.

- [3] S. K. Shin, "The prediction of item difficulty in a listening test," *Language Research*, Vol.41, No.3, pp.683-692, 2005.
- [4] K. A. Jin and C. Park, "The prediction of English item difficulty in college scholastic ability test," *English Teaching*, Vol.59, No.1, pp.267-278, 2004.
- [5] J. C. Alderson, *Assessing reading*, UK: Cambridge University Press, 2000.
- [6] E. Shohamy, "Does the testing method make a difference?," *Language Testing*, Vol.1, No.2, pp.147-170, 1984.
- [7] M. Pearson and D. Johnson, *Teaching reading comprehension*, NJ: Holt, Rinehart and Winston, 1978.
- [8] B. Davey and C. Lasasso, "The interaction of reader and task factors in the assessment of reading comprehension," *The Journal of Experimental Education*, Vol.52, No.4, pp.199-206, 1984.
- [9] T. S. Hock, "The role of prior knowledge and language proficiency as predictors of reading comprehension among undergraduates," In J. de Jong and D. Stevenson (Eds.), *Individualizing the assessment of language abilities*, England: Multilingual Matters, pp.214-244, 1990.
- [10] J. C. Alderson and A. H. Urquhart, "The effect of students' academic discipline on their performance on ESP reading tests," *Language Testing*, Vol.2, No.2, pp.192-204, 1985.
- [11] C. Clapham, *The development of IELTS: A study of the effect of background on reading comprehension (Vol. 4)*, UK: Cambridge University Press, 1996.
- [12] D. D. Qian, "Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective," *Language Learning*, Vol.52, No.3, pp.513-536, 2002.
- [13] N. Schmitt, "The relationship between TOEFL vocabulary items and meaning, association, collocation and word-class knowledge," *Language Testing*, Vol.16, No.2, pp.189-216, 1999.
- [14] P. Johnston, "Prior knowledge and reading comprehension test bias," *Reading Research Quarterly*, Vol.19, No.2, pp.219-239, 1984.
- [15] L. Ortega, "Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing," *Applied Linguistics*, Vol.24, No.4, pp.492-518, 2003.
- [16] H. Ai and X. Lu, "A corpus-based comparison of syntactic complexity in NNS and NS university students' writing," In A. Diaz-Negrillo and N. Ballier, and P. Thompson (Eds.), *Automatic treatment and analysis of learner corpus data*, PA: John Benjamins Publishing, pp.249-264, 2013.
- [17] X. Lu, "A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development," *TESOL Quarterly*, Vol.45, No.1, pp.36-62, 2011.
- [18] N. Taguchi, W. Crawford, and D. Z. Wetzel, "What linguistic features are indicative of writing quality? A case of argumentative essays in a college composition program," *TESOL Quarterly*, Vol.47, No.2, pp.420-430, 2013.
- [19] W. Yang, X. Lu, and S. C. Weigle, "Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality," *Journal of Second Language Writing*, Vol.28, pp.53-67, 2015.
- [20] B. Bulté and A. Housen, "Conceptualizing and measuring short-term changes in L2 writing complexity," *Journal of second language writing*, Vol.26, pp.42-65, 2014.
- [21] X. Lu and H. Ai, "Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds," *Journal of Second Language Writing*, Vol.29, pp.16-27, 2015.

- [22] H. J. Yoon and C. Polio, "The linguistic development of students of English as a second language in two written genres," *T2ESOL Quarterly*, Vol.51, No.2, pp.275-301, 2017.
- [23] 권인숙, "중학교 영어 교과서의 코퍼스 언어학적 어휘 비교 분석," *영어교육*, 제57권, 제4호, pp.409-444, 2002.
- [24] 이상희, *개정 7차 교육과정 고등영어 교과서 학습활동책의 수준별 난이도 분석*, 고려대학교 교육대학원, 석사학위논문, 2009.
- [25] N. Schmitt, *Vocabulary in language teaching*, UK: Cambridge University Press, 2000.
- [26] 황이수, 이제영, "수학능력시험 영어 읽기 지문의 응집성과 문항 난이도 간의 상관관계 분석," *한국콘텐츠학회논문지*, 제20권, 제5호, pp.344-350, 2020.
- [27] 최민주, 김정렬, "수능 영어 문항 유형간 응집력과 어휘정보 분석," *한국콘텐츠학회논문지*, 제17권, 제12호, 2017.
- [28] J. Sinclair, "Corpus and text-basic principles," In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice*, UK: Oxbow Books, pp.1-16, 2005.
- [29] D. Biber, B. Douglas, S. Conrad, and R. Reppen, *Corpus linguistics: Investigating language structure and use*, UK: Cambridge University Press, 1998.
- [30] M. Lewis, *Teaching collocation: Further developments in the lexical approach*, England: Language Teaching Publication, 2000.
- [31] J. Sinclair, *Preliminary recommendations on corpus typology*, UK: Corpus Linguistics Group, 1996.
- [32] S. Hong, "How can corpus data be applied for EFL learners and teachers?," *English Language Teaching*, Vol.22, No.2, pp.55-77, 2010.
- [33] B. Louw, "Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies" In M. Baker and G. Francis, and E. T. Bonelli (Eds.), *Text and technology: In honour of John Sinclair*, PA: John Benjamins Publishing, pp.157-176, 1993.
- [34] 이석재, 정채관, "한국 예비 대학생의 영어 사용 특성 파악을 위한 대규모 공개 영어 학습자 코퍼스 구축 및 분석," *한국콘텐츠학회논문지*, 제14권, 제11호, pp.1019-1029, 2014.
- [35] D. Jurafsky and J. H. Martin, *Speech and language processing: An introduction to speech recognition, computational linguistics and natural language processing*, NJ: Prentice Hall, 2008.
- [36] A. C. Graesser, M. Jeon, Y. Yan, and Z. Cai, "Discourse cohesion in text and tutorial dialogue," *Information Design Journal*, Vol.15, No.3, pp.199-213, 2007.
- [37] A. C. Graesser, M. Jeon, Z. Cai, D. S. McNamara, J. Auracher, and W. van Peer, "Automatic analyses of language, discourse, and situation models," In J. Auracher and W. van Peer (Eds.), *New beginnings in literary studies*, UK: Cambridge Scholars Publishing, pp.82-88, 2008.
- [38] M. G. Jeon and I. J. Lim, "A corpus-based linguistic analysis of high school English textbooks," *Foreign Languages Education*, Vol.17, No.1, pp.209-233, 2010.
- [39] A. C. Graesser, D. S. McNamara, and J. M. Kulikowich, "Coh-Matrix: Providing multilevel analyses of text characteristics," *Educational Researcher*, Vol.40, No.5, pp.223-234, 2011.
- [40] R. H. Baayen, R. Piepenbrock, and L. Gulikers, *The CELEX lexical database (release 2)*, PA: Linguistic Data Consortium, 1995.
- [41] M. Coltheart, "The MRC psycholinguistic database," *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, Vol.33, No.4, pp.497-505, 1981.
- [42] C. Fellbaum, *WordNet: An electronic lexical database*, Cambridge MA: MIT Press, 1998.
- [43] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller, "Introduction to WordNet: An on-line lexical database," *International Journal of Lexicography*, Vol.3, No.4, pp.235-244, 1990.

- [44] J. P. Kincaid, R. P. Fishburne, R. L. Rogers, and B. S. Chissom, *Derivation of new readability formulas (automated readability index, fog count, and flesch reading ease formula) for Navy enlisted personnel*, Naval Technical Training Command Millington TN Research Branch, 1975.
- [45] N. J. Farr, J. J. Jenkins, and D. G. Paterson, "Simplification of Flesch Reading Ease Formula," *Journal of Applied Psychology*, Vol.35, No.5, pp.333-337, 1951.
- [46] E. Charniak, *A maximum-entropy-inspired parser*, Association for Computational Linguistics conference (NACCL-1), 2000.
- [47] M. Jeon, *Automatic analyses of cohesion and coherence in tutorial dialogue*, Unpublished doctoral dissertation, University of Memphis, 2008.
- [48] J. S. Cangelosi, *Designing tests for evaluating student achievement*, New York: Longman, 1990.
- [49] R. L. Ebel, *Measuring educational achievement*, Englewood Cliffs, NJ: Prentice Hall, 1965.

저 자 소 개

이 예 나(Yena Lee)

정희원



- 2020년 8월 : 고려대학교 영어교육학(교육학박사)
- 2018년 3월 ~ 현재 : 차의과학대학교 조교수

<관심분야> : 영어교육, 교육평가, 멀티미디어 영어교육