

<http://dx.doi.org/10.17703/JCCT.2021.7.3.439>

JCCT 2021-8-51

한국 대학생의 에세이에 나타난 영어 능력 수준과 통사적 복잡성 간의 관계 탐색

The Relationship between English Proficiency and Syntactic Complexity for Korean College Students

이영주*

Young-Ju Lee*

요약 본 연구는 자동화된 통사 구문 분석 프로그램인 TAASSC를 활용하여 한국 대학생의 에세이에 나타난 영어능력 수준과 통사적 복잡성간의 관계를 살펴본다. 본 연구에서는 ICNALE 코퍼스에 포함된 한국인 대학생이 작성한 에세이를 분석하였고, 구 복잡성 지표와 절 복잡성 지표가 영어 수준을 얼마나 잘 예측할 수 있는지를 살펴보았다. 단계적 회귀분석 결과 구 복잡성 지표는 영어 능력 수준 변인의 8%, 절 복잡성 지표는 약 11%를 설명하는 것으로 나타났다. 따라서 절 복잡성 지표가 구 복잡성 지표보다 한국인 대학생의 영어 수준을 조금 더 잘 예측한다고 볼 수 있으며, 이는 구 복잡성 지표가 쓰기 능력 발전의 표준이 된다는 Biber 외(2011)의 연구 결과와는 상반된다.

주요어 : 통사적 복잡성, TAASSC 프로그램, 쓰기 지도

Abstract This study investigates the relationship between syntactic complexity and English proficiency for Korean college students, using the recently developed TAASSC(the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity) program. Essays on the ICNALE(International Corpus Network of Asian Learners of English) corpus were employed and phrasal complexity indices and clausal complexity indices, respectively were used to predict English proficiency level for Korean students. Results of stepwise regression analysis showed that indices of phrasal complexity explained 8% of variance in English proficiency, while indices of clausal complexity accounted for approximately 11%. That is, indices of clausal complexity were slightly better predictors of English proficiency than indices of phrasal complexity, which contradicts Biber et al.(2011)'s claim that phrasal complexity is the hallmark of writing development.

Key words : Syntactic complexity, TAASSC(the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity) program, Writing pedagogy

*정회원, 한밭대학교 영어영문학과 교수 (주저자)
접수일: 2021년 7월 31일, 수정완료일: 2021년 8월 7일
게재확정일: 2021년 8월 9일

Received: July 31, 2021 / Revised: August 7, 2021

Accepted: August 9, 2021

*Corresponding Author: yjulee@hanbat.ac.kr

Dept. of English Language and Literature, Hanbat National
University, Korea

I. 서론

영어교육에서 통사적 복잡성(syntactic complexity)과 영어 능력의 발전은 많은 관심을 받은 분야이지만 통사적 복잡성의 정확한 정의에 대해서는 이견이 많으며 대체적으로 특정 에세이에 나타난 다양한 통사구문으로 받아들여지고 있다 [1]. 영어를 제 2 언어나 외국어로 공부하는 학습자의 쓰기 능력이 발전함에 따라 문장, 절, T-unit의 길이가 길어지고 [2, 3] 전치사구의 사용과 같은 구의 삽입(embedding)도 빈번해진다 [4].

쓰기 능력 발달의 중요한 구성요소인 통사적 복잡성은 지금까지 T-unit의 평균길이(mean length of T-unit), 절의 평균 길이(mean length of clause), T-unit에 포함된 절의 개수(clauses per T-unit)가 분석에 많이 사용되었다 [5]. 선행연구 결과에 의하면 언어 능력이 높아질수록 학습자는 T-unit과 절의 길이가 점점 길어지고 T-unit에 포함된 절의 개수가 많아지는 경향이 있다.

T-unit의 평균길이는 지금까지 통사구문의 발달을 살펴보는 많은 선행연구에서 지표로 선정되어 연구되어 왔지만, 통사구조를 정확히 반영하지 못한다는 단점이 있다 [6]. 이를 Kyle과 Crossley[6]의 연구에 제시된 예시를 통해 살펴보면 다음과 같다. 아래의 예시에서 (1)과 (2)는 모두 12개의 단어로 이루어졌고 1개의 T-unit으로 구성된 문장이므로 T-unit의 평균길이는 12가 된다.

(1) The athletic man in the jersey kicked the ball over the fence.

(2) Because he wanted to score a goal, the man kicked the ball.

(1)에서 주어 man은 형용사 athletic과 전치사구 in the jersey에 의해 정교화(elaborated) 되었고, 동사구 kicked the ball은 위치를 나타내는 전치사구 over the fence에 의해 정교화 되었다. 반면에 (2)에서 주절인 the man kicked the ball은 종속절인 Because he wanted to score a goal에 의해 정교화 되었다. T-unit의 평균길이에 의하면 (1)과 (2)는 모두 12가 되고 두 문장은 통사구문의 복잡성 측면에서 같다는 잘못된 결론에 이르게 된다. 하지만, 두 문장은 서로 다른 구조를

가지고 있으며 T-unit의 평균길이와 같은 절의 복잡성(clausal complexity) 지표만으로는 두 문장의 통사적 복잡성과 정교화가 제대로 표현되지 못한다는 한계가 있다 [6, 7].

Biber, Gray 그리고 Poonpon[4]는 절의 복잡성을 쓰기 능력의 표준으로 삼는 것에 대한 반응을 제시한다. [4]의 연구에 의하면 구의 복잡성(phrasal complexity)이 학문적 글쓰기(academic writing)의 결정적인 특징이며, 절의 복잡성은 대화문의 특징을 보이기 때문에 T-unit과 같은 절의 복잡성 지표를 쓰기 능력의 표준으로 삼는 것은 문제가 있다고 지적한다. 또한 Norris와 Ortega[8]는 절 지표(clausal indices)는 쓰기 능력의 발전에 있어서의 언어적 변이(linguistic variation)를 제대로 포착하지 못한다고 본다.

이러한 논리에 의하면, 영어 능숙도가 높아질수록 구의 복잡성 지표가 발전하게 된다고 가정해 볼 수 있으며, 영어 수준이 높은 학생이 작성한 에세이는 구의 복잡성 지표가 절의 복잡성 지표보다 높다는 가설을 세워 볼 수 있다. 본 연구에서는 이러한 가설을 자동화된 텍스트 분석 프로그램을 활용하여 검증해 보고자 한다.

본 연구에서는 통사적 정교화와 복잡성을 살펴볼 수 있는 최근에 개발된 프로그램인 TAASSC(the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity)를 활용하여 한국 대학생이 작성한 에세이에 나타난 통사 구문을 비교하고 통사구문의 복잡성과 영어 능력간의 관계를 살펴보고자 한다.

본 연구에서 살펴보고자 하는 구체적인 연구 질문은 다음과 같다. 첫째, 영어 능력 수준과 구의 복잡성 지표간의 관계는 어떠한가? 즉, 영어 능력 수준을 가장 잘 예측하는 구의 복잡성 지표는 무엇이고 영어 능력 수준 변인을 얼마나 잘 설명하는가? 둘째, 영어 능력 수준과 절의 복잡성 지표간의 관계는 어떠한가? 즉, 영어 능력 수준을 가장 잘 예측하는 절의 복잡성 지표는 무엇이고 영어 능력 수준 변인을 얼마나 잘 설명하는가?

II. 선행연구 및 이론적 배경

최근에 개발된 TAASSC 프로그램에는 구 복잡성 지표, 절 복잡성 지표와 관련된 통사적 복잡성 지표가 포함되어 있으며 <표 1>과 <표 2>에 예시가 각각 제시되어 있다.

표 1. 구 복잡성(phrasal complexity) 지표 예시 [6, p.341]
 Table 1. Examples of phrasal complexity features [6, p.341]

| 지표 이름 | 예시문 |
|------------------------------------|---|
| 명사구 주어(nominal subject) | <i>The man in the red hat</i> gave that tall man some money. |
| 전치사의 목적어(prepositional object) | The man <i>in the red hat</i> gave that tall man some money. |
| 관계절 수식어(relative clause modifiers) | The person <i>who brought pizza</i> is a hero. |
| 간접 목적어(indirect object) | The man in the red hat gave <i>that tall man</i> some money. |
| 직접 목적어(direct object) | The man in the red hat gave that tall man <i>some money</i> . |
| 명사구 보어(nominal complement) | He is <i>a tall man</i> . |
| 전치사구(prepositional phrases) | The man <i>in the red hat</i> gave that tall man some money. |

표 2. 절 복잡성(clausal complexity) 지표 예시 [6, p.339]
 Table 2. Examples of clausal complexity features [6, p.339]

| 지표 이름 | 예시문 |
|---|---|
| 부사구 절(adverbial clause) | The accident happened <i>as night fell</i> . |
| 보문절(clausal complement) | I am certain <i>that he did it</i> . |
| 절 주어(clausal subject) | <i>What he said</i> is not true. |
| 전치사 보어절(clausal prepositional complement) | They heard <i>about you missing classes</i> . |
| 등위절(clausal coordination) | Jill runs and <i>Jack jumps</i> . |

TAASSC 프로그램은 .txt 형태의 텍스트 파일을 처리하며 .csv 형태의 결과물 파일을 생성한다. TAASSC 프로그램의 화면 예시는 <그림 1>과 같다.

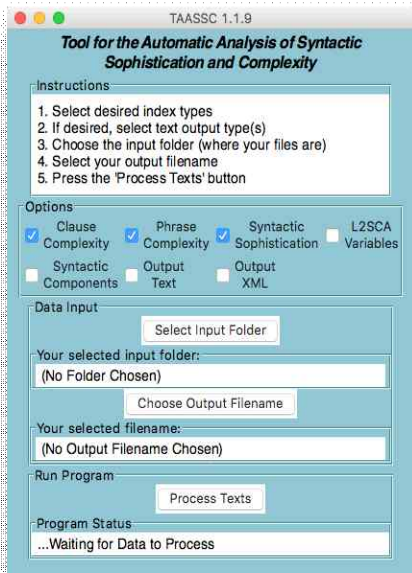


그림 1. TAASSC 1.1 프로그램
 Figure 1. TAASSC 1.1 program

지금까지 통사 구문의 복잡성을 살펴본 대부분의 선행연구[9, 10]는 해외에서 수행되었으며, Coh-Metrix 라는 프로그램을 활용하여 제한적으로 분석하였다. Guo, Crossley와 McNamara[9]는 어휘의 정교화, 통사적 복잡성, 응집력과 같은 언어자질이 TOEFL iBT 쓰기 과업의 에세이 점수를 얼마나 잘 예측할 수 있는지를 Coh-Metrix 프로그램을 활용하여 분석하였다. TOEFL iBT 쓰기 과업은 독립형 과업(independent task)과 통합형 과업(integrated task)으로 구분된다. 연구결과를 살펴보면 3인칭 단수형 동사의 사용은 통합형 과업의 에세이 점수를 예측하는 유의미한 변인이나 독립형 과업의 에세이 점수를 예측하는 변인은 아닌 것으로 나타났다. 통합형 과업은 듣기 자료와 읽기 지문의 정보를 요약해서 제시해야 하기 때문에 현재 시제로 기술하는 것이 요구되며, 3인칭 단수형 동사의 사용이 에세이 점수를 예측하는 유의미한 변인이 된다.

국내의 경우 Coh-Metrix 프로그램을 활용하여 실제로 에세이의 통사적 복잡성을 분석한 연구는 많지 않으며 대표적으로는 Kim과 Jeon[11]이 있다. Kim과 Jeon[11]은 초등학교 6학년 학생의 에세이 176개를 Coh-Metrix 프로그램을 활용하여 분석하였다. 연구결과, 영어수준이 높아질수록 많은 수의 단어를 사용하여 긴 문장을 썼으나 통계적으로 유의미한 차이는 없었고, 통사적 복잡성의 경우 본동사 앞에 평균 1개의 수식어를 사용하여 영작했으나 영어 수준에 따라 유의미한 차이는 발견되지 않았다.

III. 연구방법

3.1 연구대상

본 연구에서는 ICNALE(International Corpus Network of Asian Learners of English) 코퍼스에서 추출한 에세이 600개를 분석하였다. 한국인 대학생 300명이 흡연과 파트 타임 직업의 주제에 대해 각각 작문했으며, 300명의 한국 학생의 영어 능숙도를 유럽공동 표준기준인 CEFR(Common European Framework of Reference)에 의해 분류하면 하 수준인 A2에는 75명, 중 수준인 B1에는 149명, 상 수준인 B2에는 76명이 속한다.

3.2 통사적 복잡성 지표 추출

Kyle과 Crossley(2018)의 연구방법론과 유사한 방법

으로 통사적 복잡성 지표가 추출되었다. TAASSC 프로그램이 생성한 132개의 구 복잡성 지표와 31개의 절 복잡성 지표 중에서 정규 분포를 따르지 않는 지표를 삭제하였다. 또한 영어능력 수준과의 상관분석 결과 유의미하지 않고 상관계수가 0.1보다 작은 변인을 삭제하였으며 다중공선성(multicollinearity)이 큰 지표를 삭제하였다. 이러한 통사적 복잡성 지표 추출과정을 거쳐, 본 연구에서는 5개의 구 복잡성 지표와 6개의 절 복잡성 지표가 통계분석에 투입되었다.

3.3 자료분석

본 연구에서는 SPSS 26.0의 프로그램을 활용하여 회귀분석이 수행되었다. 회귀분석을 통해 영어 능력 수준을 가장 잘 예측하는 구 복잡성 지표와 절 복잡성 지표가 각각 무엇인지를 살펴본다. 또한 구 복잡성과 절 복잡성 중에서 어떠한 지표가 영어 능력 수준 변인을 더 잘 설명할 수 있는지를 살펴본다.

IV. 연구결과

CEFR의 A2, B1, B2의 영어 수준을 잘 예측할 수 있는 통사구문의 복잡성 지표가 무엇인지를 살펴보기 위해 단계적 회귀분석이 수행되었고, 결과는 아래와 같다. 구 복잡성의 회귀분석 결과는 <표 3>에 제시되어 있다.

표 3. 구 복잡성 회귀분석 결과
Table 3. Results of Regression: Phrasal Complexity

| 모형 | 독립변인 | R | R제곱 | β | 유의 확률 |
|----|--------------|------|------|--------|----------|
| | | 제곱 | 변화량 | | |
| 1 | 명사구의 소유격 | .046 | .046 | 0.215 | 0.000 |
| 2 | 명사의 의존어구 | .057 | .010 | -0.102 | 0.01 |
| 3 | 직접 목적어의 의존어구 | .072 | .015 | 0.131 | 0.002 |
| 4 | 명사구 주어의 의존어구 | .080 | .008 | -0.095 | 0.022 |

구 복잡성의 경우, 투입된 5개의 변인 중에서 직접 목적어의 형용사 수식어구(adjectival modifiers per direct object) 변인은 영어 능력 수준을 예측하는 유의미한 변인이 아니었으며, 회귀모형에서 삭제되었다. 산

출된 회귀모형의 R 제곱은 0.08로 8%의 설명력을 보였으며 통계적으로 유의미한 모형이다(F=12.991, p=0.000).

영어 능력 수준을 예측하는 가장 강력한 구 복잡성 지표는 명사구에 포함된 소유격(possessives per nominal)으로 나타났으며 영어 능력 수준의 4.6%를 차지했다. R 제곱 변화량은 구 복잡성 지표가 회귀모형에 추가될 때 영어 능력 수준을 예측하는 설명력이 얼마나 증가 하는지를 나타낸다. 명사의 의존어구(dependents per nominal)와 직접 목적어의 의존어구(dependents per direct object)가 추가 됨으로써 전체 설명력은 각각 1% 증가했다. 또한 명사구 주어의 의존어구(dependents per nominal subject) 지표가 추가됨으로써 설명력은 0.8% 높아졌다.

절 복잡성의 회귀분석 결과는 <표 4>에 제시되어 있다.

표 4. 절 복잡성 회귀분석 결과
Table 4. Results of Regression: Clausal Complexity

| 모형 | 독립변인 | R | R제곱 | β | 유의 확률 |
|----|--------------|------|------|--------|----------|
| | | 제곱 | 변화량 | | |
| 1 | 절에 포함된 전치사 | .053 | .053 | 0.230 | 0.000 |
| 2 | 절에 포함된 의존어 | .092 | .039 | -0.199 | 0.000 |
| 3 | 절에 포함된 명사 보어 | .109 | .018 | -0.136 | 0.008 |

절 복잡성의 경우, 투입된 6개의 변인 중에서 절의 부사 수식어구(adverbial modifiers per clause), 절의 의존어 표준편차(dependent per clause in standard deviation), 절의 수동태 조동사(passive auxiliary verbs per clause)의 3개 변인은 영어 능력 수준을 예측하는 유의미한 변인이 아니었으며, 회귀모형에서 삭제되었다. 산출된 회귀모형의 R 제곱은 0.109로 약 11%의 설명력을 보였으며 통계적으로 유의미한 모형이다(F=24.410, p=0.000).

영어 능력 수준을 예측하는 가장 강력한 절 복잡성 지표는 절에 포함된 전치사(prepositions per clause)로 나타났으며 영어 능력 수준 변인의 5.3%를 차지했다. R 제곱 변화량은 절 복잡성 지표가 회귀모형에 추가될 때 영어 능력 수준을 예측하는 설명력이 얼마나 증가 하는지를 나타낸다. 절에 포함된 의존어(dependents per clause)와 절에 포함된 명사 보어(nominal complements

per clause)가 추가됨으로써 전체 설명력은 각각 3.9%, 1.8% 증가했다.

단계적 회귀분석 결과를 요약하면, 구 복잡성 지표는 영어 능력 수준 변인에 대해 8%의 설명력을 보였으며, 절 복잡성 지표는 약 11%의 설명력을 나타냈다. 따라서 절 복잡성 지표가 구 복잡성 지표보다 한국인 대학생의 영어 수준을 조금 더 잘 예측한다고 볼 수 있으며, 이는 구 복잡성 지표가 쓰기 능력 발전의 표준이 된다는 Biber 외(2011)의 연구 결과와는 상반된다.

V. 결 론

통사적 복잡성은 쓰기 능력 발달의 중요한 구성요소이고 지금까지 절의 복잡성 지표가 쓰기 능력의 표준으로 연구되었으며, Biber 외(2011)에 의하면 구의 복잡성 지표가 학문적 글쓰기의 주요 특징이 된다. 본 연구는 자동화된 텍스트 분석 프로그램인 TAASSC를 활용하여 한국 대학생의 에세이에 나타난 통사구문의 복잡성과 영어 수준간의 관계를 살펴보았다. 본 연구에서 살펴본 두 개의 연구질문은 영어 능력 수준과 구의 복잡성 지표간의 관계 그리고 영어 능력 수준과 절의 복잡성 지표간의 관계가 어떠한가 하는 것이다. 본 연구에서는 ICNALE 코퍼스에서 추출한 한국인 대학생 에세이 600개를 단계적 회귀분석을 활용하여 분석하였다.

연구분석 결과를 요약하면 다음과 같다. 첫째, 구 복잡성 지표는 영어 능력 수준 변인의 8%를 설명하는 것으로 나타났으며, 영어 능력 수준을 가장 잘 예측한 구 복잡성 지표는 명사구에 포함된 소유격, 명사의 의존어구, 직접 목적어의 의존어구이다. 둘째, 절 복잡성 지표는 영어 능력 수준 변인의 약 11%를 설명하는 것으로 나타났으며, 영어 능력 수준을 가장 잘 예측한 절 복잡성 지표는 절에 포함된 전치사, 절에 포함된 의존어, 절에 포함된 명사 보어 이다.

본 연구는 통사적 복잡성의 특성이 영어 능력과 어떠한 관계에 있는지를 면밀히 살펴보기 때문에 영어교육 분야에서 연구결과가 활용될 수 있다. 본 연구결과를 바탕으로 실효성 있는 쓰기교육의 자료를 제공할 수 있을 것이다. 통사 구문의 향상을 위해서는 절 복잡성과 구 복잡성 모두 중요하며, 이 중에서도 명사구에 포함된 소유격, 직접 목적어의 의존어구, 절에 포함된 전치사와 같은 통사구문을 교육현장에서 지도할 필요가 있다.

References

- [1] Biber, D., Gray, B., & Poonpon, K. Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45, pp. 5-35. 2011
- [2] Biber, D., Gray, B., & Staples, S. Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, 37, pp. 639-668. 2014.
- [3] Bulte, B., & Housen, A. Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & F. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21-46). Amsterdam: John Benjamins. 2012.
- [4] Crossley, S., & McNamara, D. S. Predicting second language writing proficiency: The role of cohesion, readability, and lexical difficulty. *Journal of Research in Reading*, 35, pp. 115-135. 2012
- [5] Guo, L., Crossley, S., McNamara, D. S. Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing*, 18, pp. 218-238. 2013
- [6] Jung, Y., Crossley, S., & McNamara, D. S. Linguistic Features in MELAB Writing Task Performances. *CaMLA Working Papers*, pp. 1-17. 2015.
- [7] Kim, Sojung & Jeon, Moongee. An analysis study of English writing of elementary school 6th grade English language learners using Coh-Matrix. *Modern English Education*, 17(3), pp. 263-287. 2016
- [8] Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication*. Doctoral dissertation, Georgia State University, Atlanta, GA.
- [9] Kyle, K., & Crossley, S.. Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, 102(2), pp. 333-349. 2018.
- [10] Larsen-Freeman, D.. An ESL index of development, *TESOL Quarterly*, 12, pp. 439-448. 1978.
- [11] Lu, X.. A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45, pp. 36-62. 2011.

- [12] Norris, J. M., & Ortega, I. Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30, pp. 555-578. 2009.