

다문화 학습자 통사적 복잡성 측정법에 대한 현황과 과제

김명광*

요약 이 글은 '언어 숙달도'의 하나인 '통사적 복잡성'을 계량적으로 측정하는 방법론에 대한 최근의 연구 현황과 앞으로의 과제를 살펴본다. '통사적 복잡성'이란 '문장으로 자신의 복잡한 생각을 명확하게 표현하고 이해할 수 있는 정도'를 뜻하며, 이를 계량적으로 측정할 수 있는 방법에 '최소종결단위 측정법'(T-unit: Minimal terminable unit), '의사소통단위 측정법'(C-unit: Communication unit), '문장성분단위 측정법'(CS: Constituents of sentence¹)이 있다. 이 세 유형의 측정법에 대한 소개, 연구 성과, 그리고 그 한계점에 대하여 각 절에서 논한다. 2절에서는 T-unit 측정법을 살펴보는데 T-unit 측정법은 보통 절 이상의 단위를 측정하여 문장의 통사적 복잡성을 측정하는 방법으로 T-unit 당 절 또는 종속절의 수, 그리고 '대등 지수'가 있다. 이 측정법은 문어 측정에는 적합하지만 구어에는 적용하는 데는 어려움이 있음을 논증한다. 3절에서는 C-unit 측정법에 대한 논의를 한다. C-unit 측정법이 T-unit 측정 방법과 유사하지만 구어 처리를 위한 측정 방법이 고안되어 T-unit과 다름을 밝힌다. 일련의 연구 성과들도, 절 처리에 있어서 문제가 있어 향후 보완되어야 할 과제임을 말한다. 4절에서는 CS 측정법에 대한 논의를 한다. 이 측정법은 주로 유형과 길이를 중심으로 통사적 복잡성을 산출하는 방법으로 '해석문법(Analytical Grammar)'에 기반한다. 이 측정 방법을 사용한 일련의 연구들은 기존에 간과하였던 문장 성분의 복잡성을 수치화하였다는 데 큰 의의를 갖는다. 하지만 내포절이나, 이중적 기능을 가지는 절 등을 계산하는 데 문제가 있음으로 해서 향후 방법론적인 정교성이 필요함을 지적하였다.

주요어 언어 숙달도, 통사적 복잡성, 최소종결단위, 의사소통, 문장성분, 종속절, 대등 지수, 해석문법

1 후술하겠지만 문장 성분으로 언어 숙달도를 평가하는 방법은 해석문법(Analytic Grammar)의 견해를 받아들여 김의수(2002) 등에서 만든 독특한 측정법이다. 그런데 어떤 단위를 측정하느냐를 보았을 때, 주로 문장 성분을 중심으로 하기 때문에 이 글에서는 임의적으로 CS라고 명칭을 붙인다.

* 대구대, 국제한국어교육학, 대구대학교 다문화사회정책연구소, kimmk0857@daegu.ac.kr

1. 서론

이 글은 ‘언어 숙달도’의 하나인 ‘통사적 복잡성’을 계량적으로 측정하는 방법론에 대한 최근의 연구 현황과 앞으로의 과제를 살펴보는 데 있다. ‘언어 숙달도’란 언어 화자들이 동일한 언어를 구사하는 사람들과 의사소통을 얼마만큼 잘 할 수 있느냐 하는 언어 수행 수준의 정도이다. 여기에는 언어 화자들이 해당 언어를 오류 없이 정확하게 표현할 수 있는 정도(정확성, accuracy), 상대방에게 물 흐르듯이 매끄럽게 표현할 수 있는 정도(유창성, fluency), 그리고 자신의 복잡한 생각을 명확하게 표현하고 이해할 수 있는 정도(복잡성, complexity)²와 같이 세 가지 하위 범주로 나눌 수 있다(이하 CAF). 아울러 ‘언어 숙달도’는 ‘언어 부문(Components)’에 따라서도 나눌 수 있는데, 언어 부문이 ‘음운, 어휘(또는 형태), 통사, 의미, 화용’ 등으로 구분되므로 각기 음운 숙달도, 어휘 숙달도, 통사 숙달도, 의미 숙달도, 화용 숙달도 등과 같이 세분화하여 볼 수 있다. 따라서 ‘통사적 복잡성’이란 ‘문장으로 자신의 복잡한 생각을 명확하게 표현하고 이해할 수 있는 정도’로 개략적으로 정의할 수 있으며,³ 이를 계량적으로 측정한다 함은 문장 또는 발화 구조를 구성하는 단위들의 연결이 복잡한 정도를 양적으로 계산하여 수치화하는 것이다. 기존 연구는 문장 안의 형태소, 단어, 어절의 수(절 미만의 단위)를 세는 ‘평균 발화 길이’(MLU: Mean Length of Utterance)로 측정하였으나, 이는 취학 전 아동의 ‘숙달도’를 측정하는 데는 비교적 정확하나 이후의 시기나 외국인 학습자들에게는 부정확하다는 단점을 가지고 있다.⁴ 이에 최근에는 형태소 단위에서 벗어나 문장 또는 발화 단위 안의 절 수 또는 문장 성분의 수로 ‘통사적 복잡성’을 측정하는

2 ‘복잡성’의 개념에는 두 가지 유형이 있는데, 하나는 인지적 복잡성(cognitive complexity)이며, 다른 하나는 언어적 복잡성(linguistic complexity)으로, 언어 발달의 지표가 되는 것은 주로 언어적 복잡성이다(Bulté & Housen, 2012). 언어적 복잡성과 관련하여 내포절과 수동태가 대등접속문보다 어려우며 종결어미보다는 연결어미가, 연결어미 중에는 병렬연결어미보다 종속연결어미가 더 복잡하다는 연구 결과가 있다(Dissel, 2004).

3 Skehan(1998), Robinson(2003), Larsen-Freeman(2006), Housen & Kuiken(2009), Wolfe-Quintero, Inagaki & Kim(1998) 등이 이와 같은 견해를 지지하고 있다.

4 김명광(2016) 참조.

여러 가지 방법론이 모색되고 있다. 여기에는 ‘최소종결단위 측정법’(T-unit: Minimal terminable unit), ‘의사소통단위 측정법’(C-unit: Communication unit), ‘문장성분단위 측정법’(CU: Constituents unit⁵)이 있다. 이 글은 ‘평균 발화 길이’ 측정법을 제외⁶ 이 세 유형의 측정법의 간략한 소개와 이를 활용한 최근의 연구 성과, 그리고 앞으로 보완되어야 할 과제를 밝힌다.

2. T-unit

MLU 측정법이 대개 절미만의 단위를 측정하는 반면 T-unit 측정법은 보통 절 이상의 단위를 측정하여 문장의 통사적 복잡성을 측정하는 것이다. T-unit(최소종결단위-Minimal terminable unit)란, 문장이 하나의 T-unit으로 대응된다. 그런데 이 T-unit은 내포문(관형 구조의 주절과 종속절)은 하나로 계산하지만, 대등문은 선행절과 후행절을 각각 독립된 단위로 보아 두 개로 처리한다. 하지만 대등절 내부에 주술 관계를 나타내는 성분이 없다면 하나의 T-unit으로 본다(박지순·서세정, 2009 참조). 예를 들어, ‘어제 내가 본 영화를 철수가 보았다.’는 ‘어제 내가 본 영화’와 ‘철수가 그 영화를 보았다’의 두 문장이 연결되어 있는 데 전자가 종속절이므로 전체는 하나의 T-unit이다. 반면에 ‘산이 높고 강이 푸르다’는 두 개의 절이 대등하게 연결되어 있으므로 각각 별개의 T-unit으로 계산된다. 반면에 ‘강이 푸르고 맑다’는 대등적으로 연결되어 있을지라도 ‘맑다’의 주어 성분이 나타나지 않으므로 하나의 T-unit으로 계산된다.⁷ 또한 T-unit 측정법에서는 대등접속과 종속접속의 비율을 측정하는 ‘대등 지수’가 있는데, 그 계산법은 아래와 같다.

5 후술하겠지만 문장 성분으로 언어 숙달도를 평가하는 방법은 해석문법(Alytic Grammar)의 견해를 받아들여 김의수(2008) 등에서 만든 독특한 측정법이다. 그런데 어떤 단위를 측정하느냐를 보았을 때, 주로 문장 성분을 중심으로 하기 때문에 이 글에서는 임의적으로 CS라고 명칭을 붙인다.

6 김명광(2016)에서 MLU를 집중적으로 다루었기 때문에 논의에서 제외한다.

7 영어에서 보면 ‘He came home and washed his hands’에서 wash의 주어가 없으므로 하나의 T-unit이다.

$$(1) \text{대등지수(CI)}^8(\text{Coordination Index}) = \frac{\text{대등 접속절의 수}}{\text{대등 접속절} + \text{종속 접속절의 수}}$$

T-unit 측정은 단문보다는 대등절이 대등절보다는 내포절이 심리 처리 과정이 더 어렵다는 인지적 측면을 반영한 것으로 MLU 측정법이 어휘 중심이었다면, 이 측정법은 통사적 구조 중심의 산출 방법이다.^{9,10} 즉 T-unit 측정법은 이 단위 안에 절이 몇 개 나타나는 지, 절 중에 종속 접속의 비율이 얼마나 높은지를 산출하여 이 지수가 높으면 높을수록 통사적 복잡성도 높다는 상관성을 밝히는 지표이다. T-unit을 활용한 국내 연구를 몇 가지 들어보면 다음과 같다.

- (2) 진대연(2004), 유재임(2005), 정희모·김성희(2008), 서세정(2009), 박지순·서세정(2009), 최지영(2012), 광미영·권도하(2012), 정주리(2014), 남주연 외(2016).

각각을 간략하게 설명해 보면, 우선 진대연(2004)에서는 외국인 학습자들의 한국어 쓰기 능력 평가하는 데 있어서 텍스트의 통사적 복잡도를 계산하는 여러 척도 가운데 T-unit가 밀도 측정에서(the density of the lexis)¹¹ 매우 유용하다는 점을 언급하였다. 유재임(2005)에서는 중·고등학생 그리고 대학생들의 영어 작문을 비교하여 문장성숙도(이글의 용어를 사용하면 문장 숙달도)를 측정하는 데 T-unit을 사용하여 산출한 결과 평균 절의 수가 10.2>21.0>23.5로 학력이 증가할수록 절 사용 수가 증가함을 밝혔다. 정희모·김성희(2008)에서

8 Bardovi-Harlig(1992), 남주연 외(2016), 남주연·김영주(2015), 남주연(2015) 등.

9 남주연 외(2016), 김명광(2008a, b), 임동훈(2005), 이익섭·채완(1999)에서는 통사적 구조와 복잡성에 대하여 살펴보았다. 여기서 남주연 외(2016)의 예를 들면 다음과 같다. 종결어미보다는 연결어미가, 연결어미 중에서는 병렬연결어미보다 종속연결어미가 더 복잡한 인지적 절차를 거쳐야 한다. 종속연결어미는 시제나 양태의 선어말 어미와 결합하는 데 제약이 더 크기 때문이다.

10 국외의 연구에는 Hunt(1965), Bygate(1999), Lord(2002), Bulté & Housen(2012) 등이 있다.

11 어휘 밀도란 '전체 어휘(lexis) 중에서 내용어(content word)의 비율(Cw/L)' 또는 '기능어(function word)에 대한 내용어의 비율(Fw/L)'이다.

도 대학생들의 작문을 대상으로 능숙한 학생과 그렇지 않은 학생 간의 응집성과 결속성을 살펴보는 가운데 T-unit 측정법을 사용하여 T-unit 측정이 유용함을 말하였다. 서세정(2009)에서는 외국인 학습자(대학생)들의 작문을 대상으로 T-unit을 활용하여 통사 구조의 복잡성을 산출하였다. 특히 서세정(2009)에서는 복잡성을 계산할 때 연결어미 사용 문장 중 무오류절만을 계산하여 통사적 정확도를 높이려고 하였으며 T-unit 이외에 통사적 문법형태(시제, 상 서법 부정, 존대)에 대한 빈도수도 함께 살펴보았다. 박지순·서세정(2009)에서도 초급, 중급, 고급 외국인 학습자들의 작문을 상호 비교하여 T-unit 당 '명사절, 관형절, 부사절, 인용절, 서술절'을 측정하였다. 여기에서는 한국어의 T-unit이 대체로 문장 단위와 일치한다고 보아 문장 중심의 T-unit을 기반으로 하였다. 더 나아가 한국어의 특성 상 문장 성분은 생략될 수 있다고 하여 문장 성분의 생략과는 상관없이 서술어가 있는 문장은 T-unit으로 보았다.¹² 또한 서세정(2009)와 마찬가지로 문법적인 오류문은 측정 범위에 제외시킨 후(단순한 철자 오류문은 포함) 숙달도 지표 결과를 정량화하였다. 최지영(2012)에서는 학령기 아동을 대상으로 작문의 복잡성을 평가하였는데, 습이나 억양에 따른 발화는 그 구분이 모호하다고 하여 T-unit 측정을 문어에만 한정시켰다. 최지영(2012)에서는 필리핀 여성결혼이민자 가정의 초등학교 저학년 아동 두 명의 설명담화를 종단적으로 연구하여(1학년에서 3학년까지의 발달 추이 연구), 내포절의 T-unit을 살펴보았다. 광미영·권도하(2012)에서는 약 6~10세 사이의 일반 아동을 대상으로 구어 자료(이야기 회상 자료)에서 T-unit을 활용하여 주어 표현과 절 길이가 연령 간 유의미한 차이가 있음을 밝혔다¹³. 여기서는 정주리(2014)에서는 한국 대학생들의 논문에 T-unit 검정을 실시하였는데, T-unit 이외에 단락 수, 문장의 문법적 오류 분석 등을 함께 산출하여 학생의

12 하지만 서술어가 없는 문장의 경우 예컨대 "이런 생각하다가 벌써 스물 다섯."은 T-unit의 단위로 보지 않았다.

13 여기에서는 김영태(2002)의 발화 구분 원칙에 따라, 시간의 경과나 주제의 변화가 있을 때는 구분하고, 새로운 뜻의 첨가 없이 반복된 경우나 너무 잦은 간투어, 자동적 구어, 관련 없는 내용, 혼자 중얼거림, 거의 명료성이 없는 발화 등은 절 분석대상에서 제외하였다. 또한 절 구분의 원칙을 아래와 같이 하였다(광미영·권도하, 2012: 63).

문장 숙달도를 분석하였다. 남주연 외(2016)에서는 외국인 학습자(미국인, 프랑스인, 중국인, 일본인 학습자 총 130명)의 문어를 대상으로 명사절, 관형사절, 부사절, 인용절과 같은 내포절 유형의 통사적 복잡성을 활용하였다. 여기에서도 '대등절/절(대등지수), 대등절/T-unit, T-unit/문장'과 같이 주로 T-unit 측정법을 이용하였지만 '대등절/문장'과 같이 '문장 당 대등절'도 같이 계산하여 통사적 복잡성을 좀 더 정밀화하였다. 그런데 이러한 일련의 연구를 잘 보면 광미영·권도하(2012)를 제외하고는 문어 중심의 자료와 성인을 대상으로 T-unit을 측정하였다는 공통점을 지닌다.¹⁴ 그 이유에는 여러 가지가 있겠

절로 인정하는 경우	예외
'주어-서술어'가 모두 존재하는 경우	주어를 말한 후 서술어 없이 미완성된 경우
서술어와 다른 문장 요소가 있는 경우	'보어+서술어(뺀이 되다)'로 이뤄진 관용적 구
서술어가 연이어 있는 경우(각각의 절로 인정)	'본용언+보조용언'으로 구성된 경우는 하나의 서술어로 분석
서술어가 종결형(개가 된다), 연결형(개가 뛰니까), 관형사형(열심히 뛰는 개), 명사형(개 똥이)으로 나타난 경우	서술어인 어휘 하나로 구성된 관형사형 절은 형용사로 분석(작은 개구리가 뛰었다)

14 이는 많은 연구에서 문장 중 오류문을 걸러내어 T-unit을 산출하거나(서세정: 2009, 박지순·서세정: 2009 등), 서술어'가 없는 문장(또는 절)이나 '주어'가 없는 문장(또는 절)을 인정하지 않는 견해(박지순·서세정:2009, 광미영·권도하:2012 등)에서, T-unit 측정법 이외의 다른 지표(Cw/L, Fw/L, MLU-w, w/T 등)를 병행해서 사용하였다는 점과 연관이 있다. 참고로 각 연구들에서 활용한 지표들을 살펴보면 다음과 같다.

- 진대연(2004) - Cw/L, Fw/L,
- 유재임(2005) - MLU-w, 단어/T,
- 정희모·김성희(2008): t-unit 수, 문장 화제 수, 어휘 수/t-unit 수, 병렬적 화제 진행 수, 순차적 화제 진행 수, 확장된 병렬적 화제 진행 수, 응집성 지표 수, 비문 수/전체 문장 수, t-unit 수 / 화제 수,
- 서세정(2009) - 가. 어절, 절, 형태소, 종속절/문장(구조적 복잡도)
나. 시제, 상 서법, 부정, 존재, 연결어미, 전성어미/문장(문장 당 빈도)
다. 무오류절/문장, 학습자 연결어미 오류 양상(통사적 정확도)
라. 접속부사, 지시사/문장(문장 이상).
- 박지순·서세정(2009) - 가. 문장, T, 무오류 T, 절, 무오류 절, 단어/작문
나. T, 무오류 T, 절, 무오류절, 단어/문장
다. 무오류 T단위, 절, 무오류 절, 단어, 총단어/T
라. 단어, 총 단어/무오류 T.
- 최지영(2012): 어절, 종속절/T-unit
- 광미영·권도하(2012): T-unit이나 C-unit의 평균 발화 길이.

나, 문어가 구어보다 조사나 문장 성분 그리고 적극적인 문장이 상대적으로 잘 나타나기 때문에 측정이 용이하고, 아동의 경우 대등 접속문 이외의 내포문(관형절, 보문 등)이 잘 나타나지 않기 때문에 T 당 종속절이나 대등지수를 측정하기 어렵다는 점이 주된 이유이다.¹⁵ 이를 달리 말하면 T-unit 측정법이 문장, 성분이 생략된 문장 연결 어미나 명사형으로 종결되어 서술어가 없는 문장, 절이 도치되거나 문장 성분으로만 종결된 문장, 반복어(반복절과 문장 포함) 문장 그리고 문법적으로 오류가 빈번한 문장 등에 대한 정밀한 측정 방법의 개발이나 점검이 필요하다는 점을 말해 준다.

3. C-unit

C-unit 측정법이 하나의 주절에 안긴 종속절이나 수식하는 말을 기본 단위로 보고, 보통 절 이상의 단위를 측정하여 통사적 복잡성을 산출한다는 점은 T-unit과 유사하다. 하지만 C-unit은 ‘Communication unit(의사소통단위)¹⁶’로, 구어에서 흔히 나타나는 문장 성분이 생략된 문장 재구성된 문장, 미완성 문장, 감탄사 등과 같은 발화 단위를 인정한다는 점에서 T-unit의 개념과 다르다(남주연, 2016; 2015; Foster et al., 1996; 2000; Owens, 2004; 이지현, 2008; 이재은, 2012 등). 예컨대, “학생이에요?”라는 질문에 언어 화자가 ‘네’라고 응답했을 때 T-unit에서는 ‘네’가 독립된 절로 인정되지 않지만 C-unit에서는 하나의 단위로 인정된다. 여기에 더 나아가 화자의 의도를 구분할 수 있는 ‘역양’이나 ‘휴지’, 띄어 말하기 단위까지도 인정할 수도 있는데 이를 C-

• 정주리(2014): 낱말 수, 단락 수, 낱말 수, t-unit 자체의 수.

15 이와 관련하여 T-unit 분석이 목표어 발달에 어느 정도 이상의 수준을 뛰어넘는 경우에만 유용하다고 말한 한 Gass & Selinker(2008: 55-56, 58)의 견해에 주목할 필요가 있다. 이 논문에서는 낮은 숙달도 단계에서 T-unit 분석이 어려운 이유가 “숙달도가 낮은 학습자의 언어는 특성상 문법적·어휘적 오류가 매우 빈번하기 때문”이라고 하였다. 이러한 견해는 이는 본고에서 조사한 국내 연구들이 대부분 문어와 성인 중심의 연구 경향의 이유와 일치한다.

16 남주연(2016: 7-9)

unit 단위와 구별하여 부를 때는 As-unit(Analysis of Speech Unit, Foster et al., 2000: 361; 남주연, 2016: 7-9, 47-50 등)라고 한다. 예를 들어 “그는 공부를 하면서 논다.”를 보면, 후행절의 주어가 생략되어 있어서 T-unit의 단위에서는 하나의 unit이지만, As-unit의 경우 이 두 문장 사이에 0.5초 이상의 휴지가 있을 때는 두 개의 As-unit으로 그렇지 않을 때는 하나의 unit으로 달리 처리한다. 또한 대화 중간에 다른 화자의 개입으로 발화가 일시적으로 중단되더라도 다시 발화가 계속 이어진다면 이어진 발화까지를 한 화자의 발화 단위(speech unit)으로 산정한다¹⁷. C-unit 과 As-unit이 모두 구어의 통사적 복잡성을 주로 다룬다는 점은 상호 공통적이다. 이 측정법을 이용하여 통사적 복잡성을 산출한 기존 연구에는 다음과 같은 것들이 있다.

- (3) 이영근(2005), 박지순·서세정(2009), 황소진(2010), 이재은(2012), 김영주 외 3인(2013), 백준오·김영주(2015), 남주연(2015), 남주연·김영주(2016)

이영근(2005)에서는 학습자의 구어 복잡성을 연구하였는데 형태-통사적 분석을 통해 정확성과 복잡성을 함께 측정하기 위해 C-unit(의사소통 단위)당 절 수, C-unit당 굴절접사 수, 대등 접속과 종속 접속의 비율(대등지수), 절당 종속절 수를 측정하였다. 중급 및 고급 학습자를 대상으로 발화를 분석한 결과, 고급 학습자의 C-unit당 절 수는 중급 학습자의 두 배에 가까웠으며 고급 학습자의 발화는 중급 학습자보다 C-unit당 굴절접사가 69% 정도 많았다는 것을 밝혔다. 아울러 대등접속에 대한 종속 접속의 비율 역시 2.4배 정도 높았음을 밝혔다.

박지순·서세정(2009)에서도 작문 분석의 숙달도를 측정하였는데, 문어에서도 T-unit보다는 문장 단위를 통한 측정 방법이 더 효과적임을 말하였다. 또한 이 연구에서는 정확한 숙달도를 측정하기 위하여 작문 당 전체 단어, 절,

17 Chaudron(1988: 45)에서는 C-unit에 대해 “하나의 독립된 술어로 구어에서 질문과 동일한 서술어를 생략한 대답을 제외하고는 T-unit과 같다.”고 하여 C와 T-unit의 상관성을 말하였다.

문장 수 이외에 ‘무오류’ 단어, 절, 문장의 수도 분석하여, 초급, 중급, 고급 학습자들의 쓰기 텍스트의 통사적 숙달도 지표 결과를 정량화하였다. 황소진(2010)에서는 학령기 다문화 가정 아동과 일반 가정 아동 간의 설명 담화 표현의 능력을 상호 비교하는데, 두 집단 간의 C-unit당 절의 수와 종속절의 수 그리고 종속절 유형 간 사용률을 분석하였다. 그 결과 두 집단 간에 C-unit당 전체 종속절과 종속절 중 부사절이, 유의미한 결과를 나타낸 반면, 명사절, 관형절은 유의한 차이가 없음을 밝혔다. 이재은(2012)에서는 외국인 학습자를 대상으로 ‘과제의 복잡성과 친숙성’과 학습자 발화와의 상호 관계를 밝힌 논문으로, 여기서 외국인 학습자들의 발화에 대하여 정확성과 통사적 복잡성을 산출하는 가운데 C-unit 측정법을 활용하였다. 즉 실험에 참여한 학습자들의 발화에서 정확성을 측정하기 위해서 ‘오류 없는 의사소통 단위(communication unit, C-unit 또는 CU) 비율’(오류 없는 의사소통 단위/총 의사소통 단위의 수 $\times 100$)과 ‘오류 없는 절(error-free clauses, EFC)의 비율’(오류 없는 절/총 절의 수 $\times 100$)을 계산하였다. 여기에서 말하는 의사소통 단위(C-unit)를 사용한 이유는 구어에서 자주 나타나는 주어와 서술어의 생략, 질문에 대한 답이나 미완성된 문장, 그리고 감탄사를 분석에 포함시키기 위함이다. 통사적 복잡성과 다양성을 측정하였는데 이를 위하여 ‘총 의사소통 단위에 나타나는 총 절의 수’(=총 절의 수/총 의사소통 단위의 수)와 ‘총 의사소통 단위에 나타나는 총 어미의 수’(=총 활용 어미 수/총 의사소통 단위의 수)를 계산하였다. 김영주 외(2013)에서는 언어 산출과 언어 숙달도와의 관계를 연구하기 위해 중급 수준의 미국인과 프랑스인, 중국인, 일본인 학습자를 대상으로 구어와 문어 간의 통사적 복잡도를 산출하는 데 C-unit을 활용하였다. 구어와 문어를 분석할 수 있는 정확성, 유창성, 복잡성의 복합적인 기준을 제시하였는데¹⁸, 이 중에서 통사적 복잡성은 문어에서는 ‘문장 당 종속절 수’와 ‘T-unit당 종속절 수’를 계산하였으며 구어의 경우 ‘C-unit당 종속절 수’를 측정하여,¹⁹ 구어 복잡성

18 김영주 외(2013)에서는 정확성의 경우 총 절 당 무오류절을 기준으로, 유창성의 경우 분 당 음절수(구어는 유의미한 음절수)를 계산하였다.

19 여기에서 밝힌 내용은 숙달도 단계별로 문어의 정확성이 유창성이나 복잡성보다 지체되는 굴곡이

보다는 문어 복잡성이 숙달도와 더 큰 상관성을 있음을 밝혔다. 백준오·김영주(2015)에서는 외국인 학습자들의 문어에 대한 종속절 산출과 통사적 복잡도를 살펴보았는데, 여기에서는 종속절을 관형절, 명사절, 부사절로 그리고 관형절을 일반관형절, 동격관형절(보문절), 특수 관형절(이 글에서는 이를 ‘연계절’이라 지칭하였다)²⁰을 세분화하고, 문장 수, 문장 당 오류 수 종속절 수, 문장 당 종속절 등을 산출하였다. 그 결과 ‘부사절>관형절>명사절, 보문절>연계관형절>일반관형절’, 그리고 ‘목적어와 서술어로 구성된 절>단일 서술어 절>주어와 서술어로 구성된 절>주어 목적어 서술어를 모두 갖춘 절’ 순으로 외국인 학습자들의 문어 쓰임이 나타났음을 밝혔다. 남주연·김영주(2015)에서는 중급 수준의 한국어 과정에 참여하고 있는 외국인 학습자의 구어 자료를 바탕으로 하여 AS-unit를 적용한 단계별(2등급, 4등급, 6등급, 8등급) 적용 양상을 살펴보았다. 여기에서는 AS-unit을 활용한 ‘내포절/절, 대등절/AS-unit, 내포절/절, 내포절/AS-unit, 그리고 대등지수를 측정하였는데, 이러한 측정법들이 단계별 학습자의 한국어 숙달도와 대부분 유의미한 상관관계를 보인다는 점을 밝혔다. 남주연(2015)에서는 외국인 학습자들의 구어 발화에 대하여 통사와 어휘 복잡성으로 나누어 살펴보았는데, 이 중 통사적 복잡성에 대하여는 접속절과 내포절의 유형을 나눈 후 각각을 AS-unit 측정법을 활용하여 산출하였다. ‘접속절’의 경우 대등하게 연결된 문장만을 ‘내포’는 명사절, 관형사절, 부사절, 인용절²¹과 같은 성분절로 세분화하여 문장 확장의 복잡성을 산출

더 많다는 점, 문어의 정확성은 중급 초반과 후반에 크게 발달하였고 구어의 정확성은 중반과 후반에 급격히 발달한다는 점, 문어의 유창성은 쓰기라는 특질로 인해 구어만큼 큰 폭의 상승률이 나타나지는 않았다. 또한 숙달도와 언어 산출 간의 상관성을 비교한 결과, 구어와 문어 능력 모두가 숙달도와 유의미한 상관성이 있었다.

20 ‘피아노 치는 소리, 개선될 기미, 내가 안가는 대신, 차나 한 잔 든 다음에’와 같이 표제 명사(소리, 기미, 대신, 다음에)가 주절의 성분이 아니고, 또한 안긴 문장과 표제 명사가 대등하게 연결되지 않은 문장을 따로 분류하였다. 이는 ‘고영근·구본관, 2008:309’의 견해를 따른 것이다.

21 가. 명사절: -(으)로, -기, -(는) 것, -(으)냐/느냐, -(으)는가/는가, -(으)는지/는지.

나. 관형절: -(으)나, -는, -(으)르, -던.

다. 부사절: -어서, -고(서), -자(마자), -다가, -(으)니까, -(으)므로, -기에, -어도, -더라도, -다시피, -(으)르수록, -(으)르들, -(으)르망정, -(으)면, -거든, -어야, -(었)던들, -(으)러, -(으)려고, -고자, -게, -도록.

하였다. 이 연구에서는 기존의 종속절을 주절 안에서의 이동이 가능하다는 점과 학교 문법, 그리고 이선웅(2012)의 견해에 따라 부사절(내포절)으로 처리하였다. 또한 관형사절은 관계절‘관계 관형사절, 보문절-동격관형사절, 연계절 또는 특수 관계절’로 구분하였다. 그런데 하나의 단어만으로 체언을 꾸며 주는 경우에는 통사적으로 복잡한 구조를 형성하였다고 보기 어렵다고 하여 추가적인 부가어가 없는 관형절(빨간 차를 보았다, 철수는 문짜이 부서진 차를 보았다.)을 측정에서 제외시킨 반면 부가어가 있는 관형절은(‘완전히 부서진 차를 보았다.’)를 측정에 포함시켰다. 부사절의 경우 마찬가지로 ‘-게, -도록’ 등으로 구성되는 부사절 중에서도 부가적인 절 요소가 있거나 주절의 주어와 다른 주어가 내포절에서 사용되었을 때, 또는 후행절의 주어가 생략되었더라도 다른 주어로 복원될 수 있을 때만 부사절로 보고, ‘철수는 빨리/빠르게 달렸다.’와 같은 문장은 제외시켰다. 이러한 일련의 연구는 대부분 학습자들의 구어를 대상으로 한 성과물이라는 점에 공통점이 있다. 이는 T-unit의 단위가 주어 목적어와 서술어 등이 명시적으로 나타나고 정형화된 구조에 통사적 복잡성을 측정하는 데는 정확하지만 ‘생략, 다시 말하기, 미완성 문장’ 등 구어적 특성을 측정하는 데는 부정확하다는 점을 보완하기 위하여 나온 측정법인 것이다. 하지만 C-unit이 구어적 특성을 인정한다고 하지만 다음과 같은 문제점을 여전히 안고 있다.

첫째, 복문에서 후행절 문장 성분(주어나 목적어 또는 서술어)이 생략되어 있는 구조나 주절과 동일한 기능을 갖는 구조를 절로 인정하지 않는다는 점이다.

(4) 가. 산이 높고 푸르다.

나. 나는 밥을 먹었는데 너는 어머니?

다. 학교에? 그러지마 그냥 집에 가.

(5) 가. 그 산은 높기도 하지만 또 그 산은 푸르기도 하다.

라. 인용절: (조사 ‘-고’) -다고, -(으)라고, -(으)냐고/느냐고, -자고, 하고, 라고.

나. 나는 밥을 먹었는데 너도 밥을 먹었니?
다. 학교에 가려고, 그러지마 그냥 집에 가.

즉 C-unit에서는 (4)와 같은 문장에 있어 문장 성분(산어, 밥을, 가다)이 생략되어 있기 때문에 발화 사이에 휴지(0.5초 이상)이 없는 한 두 개의 절로 보지 않는다. 하지만 비록 (4)는 문장 성분이 없다고 하더라도 구조적으로 복원이 가능하며, 그 구조는 (5)와 같은 구조인 것이다. 무엇보다도 절은 휴지가 있느냐 아니냐 하는 것으로 구분하게 되는 것이 아니라 두 개 이상의 서술어가 출현하느냐 그렇지 않느냐 하는 ‘구조’에 따라 결정된다는 점을 볼 때 (2)를 절로 본다면 (1)도 마땅히 절로 보아야 한다.

둘째, 아래 (6)과 같이 후행절이 동일 주어가 생략되었다고 하여 하나의 절로 보고 (6나)는 비동일 주어이기 때문에 두 개의 절로 처리하는 것도, 역시 동일한 연결 어미‘-데’를 문법적으로 다른 연결어미로 처리해야 하는 모순에 빠진다.

(6) 가. 철수가 그 방에 왔는데 피곤해 보여.
나. 철수가 그 방에 왔는데 영희가 없다.

셋째, 남주연(2015)에서 제안되고 이후 일련의 논문에서 이 기준을 바탕으로 산출된 관형절의 처리 방법도 문제가 있다. 즉 부가어가 없는 관형절을 통사적으로 복잡한 구조를 형성하기 어렵다고 하여 관형절에 제외시키고 부가어가 있는 관형절과 부사절만을 절에 포함시킨 것은 C-unit 측정법의 정밀성을 약화시킨다. 즉 남주연(2015)에서는 (7가)의 ‘부서진 차’는 (7나)와 같이 부가어가 없기 때문에 관형절을 제외시켰다. 하지만 이는 C-unit를 제안한 Foster et al.(2000: 365)에 의하면, 독립절은 최소한 하나의 정형 동사(finite verb)를 포함하는 절인데 이 정의에 의하면, 위 (7가)는 부가어가 없다고 하더라도 (8)과 정형 동사의 특성을 지니고 있기 때문이다.

(7) 가. 철수는 부서진 차를 보았다.

나. 철수는 완전히 부서진 차를 보았다.

(8) 철수는 부순 차를 보았다. / 철수는 부서지는 차를 보았다. / 철수는 부술 차를 보았다.

넷째 통사적으로 복잡한 구조를 형성하기 어렵다고 하여 제외시킨다는 것은 역으로 학습자들의 ‘통사적 복잡성’을 파악하지 못하게 한다는 점에서도 문제가 있다. 왜냐하면, ‘지표’ 측정의 원래 취지는 학습자들의 언어 발달 모습을 종적으로 추구하는 것인데, 절 발달의 가장 기본이 되는 단위를 제외시켰기 때문에 단계별 숙달도 추이를 가늠할 수 없게끔 만들기 때문이다.

다섯째, 부사절 중에서도 부가적인 절 요소가 있거나 주절의 주어와 다른 주어가 내포절에서 사용되었을 때, 또는 후행절의 주어가 생략되었더라도 다른 주어로 복원될 수 있을 때만 부사절로 본다는 점도 통사적 복잡성 측정의 정밀성을 떨어뜨린다. 즉 ‘철수는 빨리/빠르게 달렸다.’와 같은 부사절을 통사적 복잡성을 측정하는 데에서 제외시켰다는 것은 주어를 복원할 수 있다는 점, 발달 추이를 알아볼 수 있는 기본 구조 단계를 제외시킨다는 문제점을 본유적으로 낳게 된다는 것이다.

4. CU

문장 성분 측정법(CS: Constituents unit)이란 문장 구조의 유형 즉 통사적 문장 성분의 유형과 길이를 중심으로 통사적 복잡성을 산출하는 방법으로 ‘해석문법(Analytical Grammar)’에 기반한다. 여기서 해석문법이란 김의수(2008)에서 제안된 것으로 문장 구조에 대한 분석틀을 선형적으로 기호화하여 말뭉치들의 통사적 복잡성과 다양성에 대한 객관적 산출을 용이하게 하기 위하여 만든 개념이다. 이 개념은 자연언어처리가 어려운 생성문법의 단점을 보완하고자 나온 일종의 기술 문법으로 여겨진다. 김의수·김태성(2009: 43)에서는

생성문법이 문장의 계층 구조를 입체적으로 이분 표시로 나타내 주지만 문장이 복잡해 질 경우 분석 결과 표시가 복잡해져 가독성이 떨어짐을 지적하였다. 또한 생성 문법이 문장의 두 가지 측면인 명제적 차원과 양태적 차원 증명제에 치우쳐서 문장 구조 파악이나 통사적 복잡성을 측정하기 어렵다고 하였다. 이에 따라 문장의 통사정보를 선형적(linear)으로 표시하고, 문법기능(Grammatical Function)을 중심으로 내적인 특성과 외적인 기능을 간략히 표시해 줄 수 있는 체계, 그리고 명제적 특성과 양태적 특성 모두를 표시할 수 있는 간결하고 효과적인 표시체계를 구축할 필요가 있음을 지적하였다. 그리하여 김의수(2008) 이하 논문에서 문장 성분을 기호화하고 각 개별 문장 성분의 기능을 숫자화하여 말뭉치 전산 처리를 용이하게 만들었는데 이를 김의수(2008)에서는 해석 문법이라고 명명하고 이를 체계화하였다²².

이 체계는 우선 문장성분을 ‘독립성분, 명제성분, 양태성분’ 등으로 나눈 후 각각을 아래와 같이 기호화한다.

(9) 가. 독립성분: 독립어→J.

나. 명제성분: 주어→A, 목적어→B, 보어→C, 부사어→D, 서술어→E.

다. 양태성분: 시제→T, 상→S, 높임→H, 태도→M, 종결→N, 부호→P.

라. 기 타: 인용→Q, 단독 명사구→X. (김태성 2009:44)

다음에 각 문장 성분의 세부정보를 숫자화 한다. 이중에 서술어와, 양태 정보 중 시제 정보를 예를 들어보면 (13), (14)와 같다.

(10) 가. 서술어의 종류

1: 자동사, 2: 타동사, 3: 형용사, 4: 비용언(체언), 5: 비용언(기타),

22 김의수(2008; 2009)의 해석문법에 대한 기반은 강범모·김의수(2004)에서 21세기 세종계획의 문어 말뭉치 구축 계획의 일환으로 대규모의 구문분석 말뭉치의 구축을 위한 한국어 문장 분석 방법론을 구축한 것에서 비롯된 것으로 보인다(김의수, 2009: 43). 더불어 김의수(2008; 2009)의 해석문법 체계는 이 후 관련 논문에서 그 내용이 조금씩 보완되었다.

6: 보조용언, 7: '이다', 8: 경동사 '하다', 9: 피동의 '되다, 지다'.

나. 서술어의 자릿수

1: 한 자리 서술어, 2: 두 자리 서술어, 3: 세 자리 서술어, ...

다. 서술어가 취한 어말어미의 종류:

1: 종결어미, 2: 관형사형어미, 3: 부사형어미, 4: 명사형어미, 5: 보조적 연결어미.

(11) 시제(T)

1: -었-, 2: -더-, 3: -ㄴ/는-, 4: -겠/리-, 5: 관형사형어미 -ㄴ/
은, 6: 관형사형어미 -는, 7: 관형사형어미 -ㄴ/을.

위를 실제 문장에 대입시키면 (12)와 같다.

(12) 가. 나는 그 음식을 안 먹고 있었다.

A_{0-4-5} B_{1-1-2} DE_{2-2-5} $E_{6-1}T_1S_2H_6M_7N_1P_1$

나. 주어A관형어 없음0-인칭대명사4-는5 목적어B관형사1-보통명사1-를2 부사어D서술어

E타동사2-두 자리 서술어2-보조적 연결어미5 서술어E보조용언6-한 자리 서술어1시제T있1상S

진행2높임H청자-해라체6태도M부정7종결N평서문1부호P마침표1)²³

위와 같은 목적은 문장의 다양성과 복잡성을 추출하기 위한 것으로 기호화된 문장 유형은 8가지로 나누어진다.

(13) 문장의 다양성과 복잡성의 측정 기준

가. [AE]: 1유형- AE, AEE, ADE, ADEE, ADDE, ADDEE

나. [ACE]: 2유형- ACE, ACEE, ACDE, ACDEE, ACDDE, ACDDEE

다. [ABE]: 3유형- ABE, ABEE, ABDE, ABDEE, ABDDE, ABDDEE

라. [ACCE]: 4유형- ACCE, ACCEE, ACCDE ACCDEE, ACCDDEE

23 (12나)는 해석문법의 기호화에 대한 이해를 용이하게 하기 위하여 필자가 해석한 것임을 밝혀둔다.

마. [ABCE]: 5유형- ABCE, ABCCE, ABCDE, ACCDDE, ABCDEE,
 ABCDDE, ABCDDEE

사. [ACCCE]: 6유형- ACCCE, ACCCEE, ACCCDE, ACCCDEE,
 ACCCDDE, ACCCDDEE

아. [ABCCE]: 7유형- ABCCE, ABCCEE, ABCCDE, ABCCDEE,
 ABCCDDE, ABCCDDEE.

자. [ABCCCE]: 8유형- ABCCCE, ABCCCEE, ABCCCDEE,
 ABCCCDDE, ABCCCDDEE

유형 분류 기준에 대한 구체적인 기술은 없지만, 여러 가지 맥락 상 ‘Type’ 과 ‘필수 성분’의 유무로 추측된다. 예컨대 AEE는 Token이 3(A, E, E)이지만 Type은 2(A, E)이므로 제1유형에 속하며 ADE는 Type이 비록 3(A, D, E)이지만 D는 필수 성분이 아니기 때문에 제외시키면 역시 AE 유형이 되기 때문이다. 김의수(2009)에서 보면 유형 자체가 복잡성을 보장하지는 못한다고 보았다.

(14) AEE(1유형)은 ADE(1유형)보다 덜 복잡한 유형이 된다. 왜냐하면, 비록 양자가 세 가지 token으로 구성되어 있다는 점에서는 동일하더라도, 전자는 2가지 type(A, E)으로 되어 있지만 후자는 3가지 type(A, D, E)으로 되어 있기 때문이다. ...생략... token은 type을 이길 수 없기 때문에...생략... (김의수, 2009: 48).

위 내용을 보면 문장의 복잡성은 ‘문장 성분 Type의 수’임을 알 수 있다. 예컨대 ‘AEE’는 ‘ADE’보다 덜 복잡한데, 그 이유는 전자는 Type이 2(A, E)인 반면 후자는 3(A, D, E)이기 때문이다. 아울러 김의수(2008; 2009)에서는 Type의 수가 같으면 그 복잡성은 동일하다고 가정한다. 예컨대 2유형 ‘ACE’와 3유형 ‘ABE’가 모두 Type이 3이므로 복잡성 수치는 동일하다. 더 나아가 복잡성 정도를 좀 더 정밀하게 측정하기 위하여 내포절도 (18)과 같이 숫자(7→명사

절)나 기호(()와 Q→인용절)들로 절 정보 표시를 한 후,

(15) 내포절의 정보

가. 명사절: 물리학을 공부하기가 어렵다

$$A_{0-7-1}(aB_{0-1-2}E_{4(0-1-0)}E_{8-4})$$

나. 관형사절: 철수는 [아름다운 꽃들]을 바라보았다

$$A_{6-1-2}(aE_{3-2}T_5)$$

다. 부사절: 철수는 멋있게 옷을 차려입었다.

$$D(aE_{3-3})$$

라. 인용절: 철수는 “누가 해?”라고 물었다.

$$C_{0-777-4}(Q_{10}A_{0-4-1}E_{1-1}N_2P_2)^{24}$$

(16) 내포절의 정보 해석

가. 명사절: 물리학을 공부하기가 어렵다

$$A_{0-7-1}(aB_{0-1-2}E_{4(0-1-0)}E_{8-4})$$

나. 주어_A관형어 없음₀-절₇-보통명사₁(주어생략_a목적어_B관형어없음₀-보통명사₁-를₂서술어_E비용언체언₄(관형어 없음₀-보통명사₁-조사없음₀)서술어_E하다₈-명사형 전성어미₄)

‘개별 문장들의 복잡성 점수를 합산한 후 그것을 총 문장 수로 나누는(복잡성 점수/총 문장 수)’ 측정 방법을 제안하였다. 이는 ‘모문의 복잡성 점수/총 문장 수(The Complexity scores of Main clause/Total Sentences→CM/s), 내포문의 복잡성 점수/총 문장 수(The Complexity scores of Subordinate clause→CS/s), 모문과 내포문 복잡성 점수 합/총 문장 수(CM+CS/s)²⁵’ 등을 통사적 복잡성 지표로 세분화된다. 이러한 지표 측정 방법을 통하여 통사적 복잡성을

24 이 예문은 김의수(2008: 22-24)의 예문 중 일부를 든 것이며, 인용절의 경우 논의의 편의를 위해 변용하여 제시하였다. 또한 (19나)는 기호화된 내포절에 대한 이해를 용이하게 하기 위하여 필자가 해석한 것임을 밝혀둔다.

25 김의수(2008; 2009)에서는 측정법에 대한 영문표기가 없어서 논의의 편의상 필자가 임의로 만든 것임을 밝혀둔다.

산출한 기존 연구에는 다음과 같은 것들이 있다.

- (17) 김의수·김태성(2009), 김의수·이로사(2009), 김의수·정한네(2009), 정은주(2010), 김의수·정은주(2009; 2011), 이나래(2010), 강주영(2010), 김의수·김은실(2012), 김의수·양지연(2012), 김의수·채문숙(2010), 김의수·김지혜(2010a; 2010b), 김의수·손현화(2011), 김의수·김혜림(2012), 박헌일 외(2014).

김의수·김태성(2009)에서는 해석문법의 방법론(CM/s, CS/s, CM+CS/s)을 바탕으로 하여 초·중·고급 외국인 학습자의 학습 교재(연세한국어 1급부터 4급까지의 교재)에 수록되어 있는 문장들을 분석하여 통사적 복잡성을 분석한 후 통사적 복잡성의 낙폭의 변화와 난이도 폭의 추이를 살폈다. 낙폭이 크다는 것은 교재의 균질성이 떨어진다는 것을 의미하는데, 통사적 복잡성의 낙폭이 3급 교재에서 가장 작아지고 4급에서 가장 커졌으므로 4급 교재가 교재의 난이도가 균질하지 않다고 해석하였다. 또한 등급별 교재들의 난이도 폭은 1~2급 구간에서는 상승하고, 2~3급 구간에서는 정체되거나 오히려 후퇴하였으며, 3~4급 구간에서는 큰 폭으로 상승하였음을 밝혔다. 이와 같이 해석문법의 측정 방법을 정형화 된 기존 교재나 지문(다시 말하면 오류가 없는 인쇄물들)에 적용한 연구 성과물에는 김의수·이로사(2009)(중학교 1학년 국어 교과서), 이나래(2010), 강주영(2010)(수능언어영역 지문), 김의수·정한네(2009)(외국인 학습자 한국어 교재), 김의수·정은주(2009) TOPIK 한국어능력시험 제13~15회 지문, 정은주(2010)(TOPIK 제12~16회 지문), 김의수·정은주(2011)(TOPIK 제17~20회 지문), 김의수·김은실(2012)(이주노동자용 교재), 김의수·양지연(2012)(국어와 한국어 교재) 등이 있다.

정형화된 인쇄물에서 벗어나 학습자들의 문어를 대상으로 해석문법의 측정법을 적용한 연구도 있다. 여기에는 우선 외국인 학습자들을 대상으로 통사적 복잡성을 연구한 김의수·채문숙(2010)에서는 SVO 언어권(중국어, 영어, 스페인어)의 한국어 학습자를 대상으로 작문을 분석한 것이 그 일례이다. 이 논문

에서는 우선 말뭉치들에 대한 문장의 복잡성과 다양성을 추출하는 과정을 비교적 상세하게 다루었는데, 그 과정이 “원시말뭉치 구축→구문분석말뭉치로 가공→구문분석말뭉치 문장의 모문과 내포문 분리→명제성분 추출→정보의 단순화 생략된 성분 복원→복잡성/다양성 정보 추출(CM/s, CS/s, CM+CS/s)”라고 기술하였다. 또한 이 논문은 정형화되지 않은 문어 자료이기 때문에 “완벽한 문장, 철자 정도의 오류 문장”만을 문장 분석에 포함시켰으며, 의미 파악은 되지만 분명한 문법적 오류를 지닌 문장 그리고 문장도 잘못되었고 도무지 의미도 파악되지 않는 문장은 분석에서 제외시켰다.(김의수·채문숙, 2010: 96). 이러한 과정과 분석 원칙을 바탕으로 SVO 언어권의 작문을 살펴 보면서 세 언어권 모두 통사적 복잡성에서는 초급·중급·고급 모두 모문의 사용이 비슷했으며, 숙달도가 높아질수록 내포문 사용이 비례적으로 높아졌다는 사실과 문장 유형의 다양성에서도 세 언어권 모두 중·고급으로 갈수록 내포문의 비중이 높아졌음을 밝혔다. 역시 동일한 방법으로 연구한 김의수·김지혜(2010a)에서도 중급의 중국어권 학생과 한국인 대학생의 문장을 비교하였는데, 중국인 학습자들보다 한국인 대학생들의 작문에서 내포문의 복잡성과 다양성이 더 높았음을 결과로 제시하였다.

학습자들의 구어를 바탕으로 해석문법 차원의 통사적 복잡성을 연구한 결과물도 있다. 김의수·김지혜(2010b)에서는 중국어권 학문 목적 학습자와 한국인 대학생들의 구어 발화의 모문과 내포문의 비중 통사적 복잡성, 다양성 등을 비교한 것이 그 일례이다. 여기에서는 문어와 다른 구어의 특성을 감안한 분석 원칙을 몇 가지 산정하였는데, 구어의 특징을 보여주는 종결어미로 문장이 끝나지 않는 경우, 연결 어미나 전성어미가 종결 어미화 된 문장을 포함시키고, 구어에서 반복되는 성분은 하나만 계수하고, 조사가 생략되어 있을 경우 그 생략이 자연스러운 경우 정문으로 처리하여 분석하였다(김의수·김지혜 2010: 252). 이러한 원칙 하에 해석문법의 측정법을 활용하여 분석한 결과 내포문 서술어의 총 개수에서는 한국인 대학생들이 중국인 유학생들에 비해서 5.2배 앞섰고, 서술어의 유형에서는 3.3배 앞섰지만 모문은 이와 반대로 한국인 대학생들보다 중국인 유학생이 1.2배 높음을 밝혔다. 이에 대하여 중국인

유학생이 구어 발화의 형식을 모문에 의지하는 반면 한국이 유학생은 통사적 복잡성이 높은 내포문에 의지하고 있다고 해석하였다. 김의수·김혜림(2012)은 초등학교 1, 3, 5학년에 재학 중인 다문화가정 아동과 일반가정 아동의 구어 발달 양상을 비교했다. 그 결과, 통사적 복잡성과 다양성 모두에서 학년이 올라갈수록 두 집단 간의 격차가 벌어졌다. 즉 일반가정 아동의 경우 학년이 올라갈수록 통사적 복잡성과 통사적 다양성이 증가되었지만, 다문화가정 아동의 경우 발달이 더디게 나타났음을 밝혔다. 김의수·손현화(2011)에서는 베트남 이주 노동자를 인터뷰하여 구어 산출에 나타난 통사 복잡성을 연구하였다. 문장 성분의 복잡성을 분석한 결과, 스리랑카 이주 노동자들은 장기 체류를 할수록 통사 복잡성이 향상되었지만, 베트남 이주 노동자들은 장기 체류를 하더라도 단기 체류자와 유사한 통사 복잡성 수준에 멈추어 있는 것을 발견하였다. 이 연구에서는 베트남어가 한국어와 유형적으로 크게 다르고, 문장의 기본 어순이 달라 스리랑카 이주 노동자들보다 통사 발달 과정에 큰 어려움을 겪고 있는 것으로 보였다. 김의수·김혜림(2012)의 연구는 구문분석말뭉치를 구축하여 통사적 복잡성과 다양성을 추출하여 복합적으로 구어 문장을 살핀 데에 그 의의가 있다. 박현일 외(2014)에서는 해외 체류 경험이 있는 화자(해외파)와 그렇지 않은 화자(국내파)가 영어를 한국어로 통역할 때 나타나는 통사적 복잡성의 양상을 분석하였다. 여기에서 국내파가 해외파에 비해 더 다양한 종류의 문장을 사용하며, 통사적으로 더 복잡하고 다양한 종류의 문장을 사용한다는 점과, 원문의 난도(난이도)가 높을 경우 국내파는 내포문에서 더 복잡한 문장을 많이 사용하지만 해외파는 다양한 모문 및 내포문을 사용한다는 점을 지적하였다. 해석문법의 측정 방법이 문장 성분의 수와 유형을 중심으로 통사적 복잡성을 측정하고, 생략된 성분에 대한 측정, 그리고 문형에 대한 난이도를 고려했다는 점은 기존 연구와 차별화되며 나름대로 장점을 가지고 있으나, 다음과 같은 몇 가지 문제점은 통사적 복잡성의 정밀화를 위하여 보완되어야 된다고 본다. 첫째, 해석 문법을 추구한 일련의 논문들이 거시적인 내포질의 복잡성은 측정하였지만, 명사절, 관형사절, 부사절, 인용절과 같은 세부적인 복잡성은 측정하지 않음으로 해서 구체적인 언어 발달 모습을

파악할 수 없다. 비록 김의수(2008), 김의수·김태성(2009)에서 내포절 정보를 기호화하였지만 이후 후속되는 논문에서는 세부 유형별 통사적 복잡성을 측정하지 않은 채 상위 개념인 내포절만 측정하였기 때문에 학습자들의 개별적인 통사적 숙달도를 파악하기 위한 구체적 기준의 역할이 되지 못한다는 것이다. 더 나아가 해석 문법에서의 종속절 처리 자체에도 보완될 것이 있다. 해석 문법이 선형적 기호화에 치중하였기 때문에 관형사절의 표제 명사와 같이 이중적 기능을 나타내는 것을 표시하는 데 상대적으로 약점을 가진다. 즉 관형사절의 표제 명사는 주절에서의 기능뿐만 아니라 종속절 내부(해석문법에서는 내포절)에서의 기능도 갖는데, 해석 문법이 구문 분석 처리에 용이하게 하기 위하여 선형성을 강조하기 때문에 이를 포착하고 전산 처리를 하기는 쉽지 않다는 것이다. 예컨대 “아름다운 꽃을 보았다(aB(aE)E)”의 “꽃”은 주절에서는 목적어(B)로 종속절에서는 주어(A)라는 이중적 기능을 가지지만 이를 선형적으로 표시할 때는 “B”와 같이 주절의 기능으로만 나타나기 때문이다. 이러한 이중 기능은 언어 발달에서 매우 중요한 표지로 작동하는 데 그것은 관형사절의 표제 명사가 주절의 기능과 동일 하나 그렇지 않느냐 하는 것이 학습자들의 언어 숙달도와 밀접하게 연결되어 있기 때문이다(이영옥, 1982; 김명광 2008a, b). 혹시 해석 문법에서는 생략 성분 표지(여기서는 가로 안의 소문자 a)로 알 수 있다고 볼 수 있으나, 이 생략된 정보가 이중 기능의 한 요소라는 의미를 갖기 위해서는 B와 동 지시 기능을 갖는다는 결속 표시가 별도로 마련되어야 하나 해석문법에는 이를 위한 기호 장치가 없다.

둘째, 문장의 필수성분을 ‘주어, 목적어, 보어’에만 한정하여 문장의 난이도를 측정하였다는 점이다. 잘 알다시피 부사어의 경우 문형에서 생략이 가능한 임의적인 부사어도 있지만 생략이 가능하지 않은 필수적인 부사어도 있다. 예컨대 아래와 같이 밑줄 친 부분의 성분은 부사어이기는 하나 문장 안에서 필수적인 성분으로 생략되면 비문이 된다.

(18) 가. 이 그림이 저 그림과 똑같아요.

나. 이 편지를 우체통에 넣어라.

다. 할아버지는 조카를 양자로 삼으셨다.

해석 문법에서 Type의 경우 서술어의 필수 성분이 아닌 부사어는 제외되기 때문에, (21)과 같은 구조는 다양성에 대한 측정의 정확성을 떨어뜨리게 된다. 더 나아가 (21다)와 같은 경우는 흔히 5형식으로 말하는 문형으로 대개의 문법서가 독립적인 문형으로 처리하지만 해석 문법에서는 서술어의 종류(13가)에 포함되지 않기 때문에 역시 문제가 될 수 있다.

셋째, 문어 문장과 구어 발화에서 오류 문장을 제외한 내포절만을 계산하였다는 것이다. 앞서 살펴보았듯이, 문법적인 오류를 지닌 문장이나 의미 파악이 되지 않는 문장은 분석에서 제외시켰는데, 정형화된 인쇄물의 경우 오류가 거의 없기 때문에 이것이 별 문제가 되지 않으나, 문어 문장과 구어 발화의 경우 학생들의 오류는 언어 발달 단계에서 매우 중요한 역할을 한 다는 점에서 문제가 있다는 것이다. 즉 학생들이 생성해낸 문장이나 발화에서 오류문이 얼마만큼 되느냐, 그리고 그 오류의 유형에는 어떤 것이 있느냐는 또 하나의 언어 발달 지표이기 때문이다. 이는 오류가 실수와 달리 중간 언어 체계를 가지고 있다는 사실을 염두 해 둔다면 이에 대한 측정법이 해석 문법에서도 반드시 필요한 것이다. 넷째, 해석 문법의 체계에서는 양태적 정보를 중시한다고 하였으나 후속 논문에서 양태적 정보에 대한 통사적 복잡성을 측정하는 것들이 거의 나타나지 않는다. 양태적 정보가 해석 문법의 장점이라면 이에 대한 논의 결과가 있어야 됴에도 불구하고 많은 논문들에서 통사적 복잡성과 다양성의 측정에만 치우쳐 있다. 이 이외에도 해석 문법에서 내포문에서 보문과 종속절을 구별하지 않았다는 점, 두 개 이상의 동사(E)가 연이어 나타난 구조를 단순하게 하나로 처리하여 복잡성 지수를 동일하게 처리한 점, 구어에서 흔히 나타나는 동사가 생략된 조각 문장의 경우 그 문형 처리 방법이 모호하다는 점, 상대적으로 통사적 복잡성이 높은 내포문을 주절과 동일하게 난이도 점수를 부여하였다는 점, 그리고 유형별 난이도 점수에 대한 보다 더 객관적인 검증이 필요하다는 점 등은 해석 문법의 통사적 복잡성을 측정하는 데 있어, 보다 더 정밀화를 꾀하기 위하여 향후 보완되어야 될 사항이라고 여겨진다.

5. 결론

이 글을 통하여 ‘언어 숙달도’의 하나인 ‘통사적 복잡성’을 계량적으로 측정하는 방법론에 대한 최근의 연구 현황과 앞으로의 과제를 살펴보았다. 서론에서 ‘통사적 복잡성’이란 ‘문장으로 자신의 복잡한 생각을 명확하게 표현하고 이해할 수 있는 정도’라고 정의하였으며 이를 계량적으로 측정할 수 있는 방법에 ‘최소종결단위 측정법’(T-unit: Minimal terminable unit), ‘의사소통단위 측정법’(C-unit: Communication unit), ‘문장성분단위 측정법’(CS: Constituents unit)가 있음을 말하였고, 각 절로 이 세 유형의 측정법의 간략한 소개와 이를 활용한 최근의 연구 성과, 그리고 앞으로 보완되어야 할 과제를 밝혔다. 각 절에서 밝힌 내용을 적어보면 아래와 같다.

2절에서는 T-unit 측정법에 대한 논의를 하였다. T-unit 측정법은 보통 절 이상의 단위를 측정하여 문장의 통사적 복잡성을 측정하는 방법으로 T-unit 당 절의 수, T-unit 당 종속절의 수, 그리고 대등접속과 종속접속의 비율을 측정하는 ‘대등 지수’가 있음을 말하였다. 국내 연구에는 진대연(2004), 유재임(2005), 정희모·김성희(2008), 서세정(2009), 박지순·서세정(2009), 최지영(2012), 곽미영·권도하(2012), 정주리(2014), 남주연 외(2016) 등이 있다. 이러한 일련의 연구 성과들은 T-unit 측정법이 통사적 복잡성을 측정하는데 유의미함을 말하였지만, 구어에 적용하는 데는 한계가 있음을 말하였다. 또한 문장 성분이 생략된 문장, 연결 어미나 명사형으로 종결되어 서술어가 없는 문장, 절이 도치되거나 문장 성분으로만 종결된 문장, 반복어(반복절과 문장 포함) 문장 그리고 문법적으로 오류가 빈번한 문장 등에 대한 정밀한 측정 방법의 개발이나 점검이 필요하다는 점을 말하였다.

3절에서는 C-unit 측정법에 대한 논의를 하였다. C-unit 측정법이 T-unit 측정 방법과 유사하지만 구어에서 흔히 나타나는 문장 성분이 생략된 문장 재구성된 문장, 미완성 문장, 감탄사 등과 같은 발화 단위를 인정한다는 점에서 T-unit의 개념과 다름을 말하였다. 국내 연구에는 이영근(2005), 박지순·서세정(2009), 황소진(2010), 이재은(2012), 김영주 외 3인(2013), 백준오·김영주

(2015), 남주연(2015), 남주연·김영주(2016) 등이 있다. 이러한 일련의 연구 성과들은 구어에 적용하는 방법론적 문제와, 이 측정법의 타당성, 그리고 T-unit 보다 절 차원의 유형을 좀 더 세분화하여 측정하였다는 것이다. 하지만 이 측정법의 문제로는 문장 성분이 생략된 절과 그렇지 않은 절을 달리 보았다는 점, 부가어가 없는 관형절을 제외시켰다는 점, 부사절을 매우 한정되게 설정하였다는 점이 이 측정법의 타당성을 약화시킴을 말하였다.

4절에서는 CU 측정법에 대한 논의를 하였다. 이 측정법은 주로 문장성분의 유형과 길이를 중심으로 통사적 복잡성을 산출하는 방법으로 ‘해석문법(Analytical Grammar)’에 기반함을 언급하였다. 이 측정법은 우선 문장성분을 ‘독립성분, 명제성분, 양태성분’ 등으로 나누고 서술어의 종류, 자릿수, 시제, 양태 정보에 대한 숫자화를 부여하여 통사적 복잡성을 측정한 것이다. 이를 기반으로 한 연구에는 김의수·김태성(2009), 김의수·이로사(2009), 김의수·정한네(2009), 정은주(2010), 김의수·정은주(2009, 2011), 이나래(2010), 강주영(2010), 김의수·김은실(2012), 김의수·양지연(2012), 김의수·채문숙(2010), 김의수·김지혜(2010a, 2010b), 김의수·손현화(2011), 김의수·김혜림(2012), 박현일 외 (2014) 등이 있다. 이러한 일련의 논문들은 다양한 대상과 다양한 자료에 적용한 후 학습자들의 수준에 따라 복잡성이 증가한다는 점을 밝힌 내용으로 기존에 간과하였던 문장 성분의 복잡성을 수치화하였다는 데 큰 의의를 갖는다. 하지만 해석 문법을 추구한 일련의 논문들이 거시적인 내포절의 복잡성은 측정하였지만, 명사절, 관형사절, 부사절, 인용절과 같은 세부적인 복잡성은 측정하지 않았다는 점, 선형적 기호화에 치중하였기 때문에 관형사절의 표제 명사와 같이 이중적 기능을 나타내는 것을 표시하는 데 상대적으로 약점을 가진다는 점, 문장의 필수성분을 ‘주어, 목적어, 보어’에만 한정하여 문장의 난이도를 측정하였다는 점, 문어 문장과 구어 발화에서 오류 문장을 제외한 내포절만을 계산하였다는 점 등 여러 문제가 있으며 앞으로 통사적 복잡성을 보다 더 정밀화하기 위해서는 이와 같은 문제 해결을 위한 방법론적인 정교성이 필요함을 말하였다.

참고문헌

- 곽미영·권도하, 2012, 이야기 회상산출에서 주어 생략 특성, 한국특수교육문제연구소, 특수교육 저널: 이론과 실천, 13(3), 59-77.
- 김명광, 2008a, 해외 귀국 학생들의 비동일 기능 관계절 사용 실태와 진단, 한국초등국어교육, 36, 75-113.
- 김명광, 2008b, 해외 귀국 학생들의 동일 기능 관계절 사용 실태와 진단, 새국어교육, 78, 5-34.
- 김명광, 2016, 언어 숙달도 평가에서 MLU 측정법의 해석 타당성에 대한 일고, 새국어교육, 109, 한국국어교육학회.
- 김명광, 2016, 외국인 학습자들의 언어 발달 지표에 대한 최근 연구 경향, 언어와 문화, 12(2).
- 김영주·남주연·이선진·이선희, 2013, 한국어 숙달도와 산출 능력 간의 상관관계연구, 국어국문학, (164), 209-244.
- 김영태, 1997, 한국 2-4세 아동의 발화길이에 관한 기초연구, 언어청각장애연구 2, 5-26.
- 김의수, 2008, 문장의 구조와 해석문법, 한국언어문학, 67, 5-35.
- 김의수·김은실, 2012, 이주노동자용 한국어교재들의 통사적 수준 비교 연구: 통사적 복잡성과 다양성을 중심으로, 한민족어문학, 62, 401-426.
- 김의수·김지혜, 2010, 중국어권 학문 목적 학습자 구어 문장의 통사적 특징과 한계, 우리어문연구, 36, 243-291.
- 김의수·김태성, 2009, 외국어로서의 한국어 학습 교재의 문장 분석, 언어학연구, (14), 39-67.
- 김의수·손현화, 2011, 이주노동자의 한국어 구어 문장이 지닌 통사적 복잡성과 그 한계, 한국어학, 50, 111-140.
- 김의수·이로사, 2009, 국어교육 교재와 한국어교육 교재 간의 접점 연구 -문장의 통사론적 난도를 중심으로, 문법 교육, 12, 139-171.
- 김의수·정은주, 2011, 통사적 복잡성에 기반한 TOPIK 읽기 지문의 난이도 양상 연구: 명제 차원과 양태 차원의 종합적 고찰을 중심으로, 한국문법교육학회 학술발표논문집, (2), 64-71.
- 김의수·정한네, 2009, 한국어 교재를 구성하는 텍스트의 통사론적 난도와 균질성 연구, 어문론총, 41, 1-33.
- 김의수·채문숙, 2010, SVO 언어권 한국어 학습자들의 작문에 나타난 문장의 복잡성과 다양성 연구, 한국언어문학, 75, 85-112.
- 김의수·김혜림, 2012, 다문화가정 아동과 일반가정 아동의 구어 문장 비교 연구, 우리어문연구, 42, 31-62.

- 남주연, 2015, 한국어 학습자의 구어 복잡성 연구: 통사 및 어휘 복잡성을 중심으로, 경희대학교 박사학위 논문.
- 남주연·김양희·김영주, 2016, L2 한국어 문어 산출에서의 통사 복잡성 측정, 한국어 의미학, 51, 21-56.
- 남주연·김영주, 2015, 한국어 학습자의 구어 산출에 나타난 통사 복잡성 발달, 한국어 교육 26(1), 59-87.
- 박지순·서세정, 2009, 쓰기 텍스트 분석을 통한 한국어 학습자의 통사적 숙달도 측정 연구, 언어와 문화, 5(2), 151-173.
- 백준오·김영주, 2015, L2 한국어 학습자의 종속절 산출과 통사적 복잡도, 담화와 인지, 22(2), 53-74.
- 서세정, 2009, 통사적 숙달도 진단을 통한 한국어 학습자의 중간언어 발달 연구: 학습자 작문 텍스트의 분석을 바탕으로, 연세대학교 석사학위 논문.
- 유재임, 2005, T-unit 분석방법과 문장합성연습, 신영어영문학, 31, 259-287.
- 이선웅, 2012, 한국어 문법론의 개념어 연구, 월인.
- 이영근, 2005, 한국어 학습자 중간언어 발화의 정확성과 복잡성을 측정하기 위한 분석도구에 관한 소고, 국제한국어교육학회 학술대회논문집, 71-175.
- 이재은, 2012, 과제의 복잡성과 친숙성이 학습자 언어 산출에 미치는 영향, 경희대학교 석사학위 논문.
- 이지현, 2008, 학령기 아동의 구문 발달: 대화 및 이야기과제를 중심으로, 나사렛대학교 석사학위 논문.
- 정주리, 2014, 문법연구: 대학 글쓰기 수업개발 방안: 전문대학 학생들의 글쓰기에 나타난 문법적 오류 유형을 중심으로, 문법교육, 22, 169-200.
- 정희모·김성희, 2008, 대학생 글쓰기의 텍스트 비교 분석 연구: 능숙한 필자와 미숙한 필자의 텍스트에 나타난 특징을 중심으로, 국어교육학연구, 32, 393-426.
- 진대연, 2004, 한국어 쓰기 능력 평가에 대한 연구: 텍스트 생산 능력 평가를 중심으로, 국어교육학연구, 19, 국어교육학회.
- 최지영, 2012, 다문화가정 아동의 설명담화 구문발달 양상: 필리핀 이주여성가정의 초등학교 저학년 아동을 대상으로, 한국어 교육, 23(2), 445-465.
- 황소진, 2010, 학령기 다문화가정 아동과 일반가정 아동 간의 구문표현능력 비교: 설명담화를 중심으로, 단국대학교 석사학위 논문.
- Foster, P. & Skehan, P., 1996, The Influence of Planning and Task Type on Second Language Performance. *Studies in Second Language Acquisition*, 18(3), 299-324.
- Foster, P., Tonkyn, A. & Wigglesworth, G., 2000, Measuring spoken language: a unit for all reasons, *Applied Linguistics*, 21(3), 354-375.

- Housen, A. & Kuiken, F., 2009, Complexity, Accuracy, and Fluency in Second Language Acquisition, *Applied Linguistics*, 30(4), 461-473.
- Larsen-Freeman, D., 2006, The Emergence of Complexity, Fluency, and Accuracy in the Oral and Written Production of Five Chinese Learners of English, *Applied Linguistics*, 27(4), 590-619.
- Robinson, P., 2003, The Cognition Hypothesis, task design and adult task-based language learning, *Second Language Studies*, 21(2), 45-105.
- Skehan, P., 1998, *A cognitive approach to language learning*. Oxford University Press.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y., 1998, *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity*, University of Hawai'i, Second Language Teaching and Curriculum Center.

2016.12.7 투고일, 2016.12.8 심사일, 2016.12.30 게재확정

The Current Situation and Future Tasks on Evaluation Methodologies of Syntactic Complexity for Multi-Cultural Learners

Myoung-Kwang Kim*

Abstract This study aims at clarifying the current situation problems of measurement of Syntactic Complexity among language proficiency. Syntactic Complexity is a kind of ability who he/his can express or understand his/her intricate thought clearly to others or himself/herself. The measurement of it has three type, Minimal terminable unit(T-unit), Communication unit(C-unit), Constituents unit(CU), In Chapter 2, I discuss T-unit. T-unit calculates the number of the size more than clause unit which has sub-measurements, of clause or subclause and equal one. It is appropriate for measuring the written materials not spoken ones. In Chapter 3, I discuss C-unit. The methodological approach of C-unit is similar to T-unit. but It has different instruments specific for spoken materials from T-unit. C-unit is appropriate for measuring spoken ones. but the view of clause should be supplemented. In Chapter 4, I discuss CU. CU mainly measures the number and type of constituents of sentence on the base of analytical Grammar. CU has strong point in that it includes constituents to syntactic complexity overlooking past methodological approaches. But it has also problem of connotation structure and double functional clause.

Keywords Language Proficiency, Syntactic Complexity, T-unit, C-unit, CU, Subclause, Equal Index, Analytical Grammar

* Daegu University, Professor, International Korea Language Education, kimmk0857@daegu.ac.kr