

## 동북아시아 영어 학습자들의 서법 조동사 사용에 대한 연구\*

윤태진 · 이용훈

(성신여자대학교, 교수 · 충남대학교, 강사)

### 1. 서론

본 논문의 목적은 한중일 영어 학습자들이 토플 시험에서 작성한 영어 에세이에서 사용한 서법 조동사를 바탕으로 Bates & MacWhinney(1989)가 제안한 경쟁 모델(Competition Model)을 이용하여 중간 언어(Interlanguage) 모형을 구축하는 것이다. 서법 조동사에 대한 선행연구들은 영어학습자들이 양태동사를 올바르게 사용하는 데 어려움이 있으며, 특정 양태의 의미나 형태들을 모국어 화자에 비해 너무 많이 사용하거나 너무 적게 사용하여 중간언어의 특징을 드러낸다고 알려져 있다(Adjemian, 1976; Aijmer, 2002; Oh, 2007; Park 2012 등). 여기서 중간언어(interlanguage)는 L1에서도 L2에서도 나타나지는 않지만 체계적으로 구성된 언어를 일컫는다(Selinker, 1969; Odin, 1989; Jarvis, 2001).

Bates & MacWhinney(1989)의 경쟁모형(competition model)은 언어 학습자가 목표어를 학습할 때, 의미 단서, 형태 단서, 통사 단서, 음운 단서들을 고정된 순서로 처리하는 것이 아니라, 동시에 경쟁적으로 처리하는 모형을 일컫는다. 제2언어 습득에 경쟁모형을 적용하면, 모국어인 원천 언어(source language)의 문장을 이해하는 데 활용되던 단서들이 습득하고자 하는 목표 언어(target

---

\* 제1저자: 윤태진, 교신저자: 이용훈. 이 논문은 2016년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2016S1A5A2A03927441).

language)의 문장을 이해하는 데에도 영향을 받는 지, 영향을 받는다면 어느 정도로 영향을 받는지를 파악하는 데 사용할 수 있다.

경쟁 모델은 다음의 몇 가지 특징을 갖는다. 첫째, 경쟁 모델은 정보 처리와 관련하여 언어의 기능적인 측면(functional aspect)을 강조한다. 언어를 학습하는 사람들은 자신의 의견을 주장하거나, 부정하거나, 요구하는 등의 다양한 화용론적인 목적을 달성하기 위해 언어를 사용하며, 자신들의 목적을 달성하기에 가장 적절한 형태를 학습하게 된다고 전제하고 있다. 둘째, 경쟁 모델은 언어가 그 기능을 달성하기 위해서는 수많은 종류의 다양한 정보가 동시에 처리되고 표상되어야 한다고 가정한다. 예를 들어, 영어 학습자가 ‘coffee, please’와 같이 간단한 문장을 말하기 위해서, 영어학습자는 ‘coffee’와 ‘please’ 등의 적절한 단어를 선택해야 하며, 선택한 단어들을 어떻게 적절히 배열해야 하는가를 결정하고, 공손한 표현을 위해 ‘please’를 사용하는 등의 통사 규칙에 대한 정보 처리를 해야 한다. 셋째 경쟁 모델은 언어의 사용에는 여러 과정이 포함되며 각 과정마다 다수의 대안들 중 하나를 선택하기 위한 경쟁이 존재한다고 가정한다. 예를 들어, ‘Can I help you?’ 대신에 ‘May I help you?’가 사용될 수도 있다. 영어학습자들은 자신의 과거 경험에 의해 여러 가지 선택 가능한 대안 중에서 현재의 맥락에서 가장 활성화(activation)의 강도가 높은 것을 선택하게 된다. 언어를 말하고 듣는 경험이 증가됨에 따라 활성화 강도가 변화하고 이에 따라 발달이 이루어진다.

이런 점에서는 경쟁모형은 처리 기반(process-based) 제2언어 습득론의 이론적 기반과 언어 사용의 관찰을 통한 경험적 방법론을 결합한 모형이다. 넓은 의미에서, 경쟁모형은 제2언어 습득 및 제1언어 습득 등의 연구에서 언어진 풍부한 결실로 생긴 확률기반의 문법 처리이론이다. 즉, 경쟁모형의 관점에서는 문법 규칙이 문장을 문법적 혹은 비문법적 문장이라는 이분법적으로 나누는 기준이라기보다는, 상대적인 영향력을 지니고 있는 전략이나 단서로 여겨진다. 따라서 한 문장의 의미를 도출하기 위해서 여러 가지 규칙들이 서로 협력하거나 갈등을 일으키게 되는 현상들을 설명할 수 있는 장점을 가진 모형이다(Bates & MacWhinney, 1982, 1989; 김현진, 2014; Jun & Lee, 2015;

Yoon & Lee, 2016).

확률기반 문법처리 이론인 경쟁 모델을 경험적 방법론으로 적용하여 채택하기 위해서는, 코퍼스를 이용하여 풍부한 언어 정보를 컴퓨터에 입력한 후, 입력된 여러 정보들이 활성화되는 정도를 측정할 수 있는 통계 모형을 사용할 필요가 있다. 이를 위해 본 논문에서는 한국어, 중국어, 및 일본어를 각각 모국어로 구사하는 영어 학습자들이 치른 TOEFL 시험의 에세이를 말뭉치로 하여, 형태 단서, 통사 단서, 그리고 의미 단서들과 관련된 다양한 언어정보들을 태깅하였다. 그리고 태깅한 언어 자료를 바탕으로, 경쟁모형이 서법 조동사를 선택할 때 미치는 영향을 살펴보았다. 이는 선행 연구에서 서법 조동사의 선택에 있어서 언어를 구성하는 핵심적인 환경들이 얼마나 영향을 미치는지에 대한 연구는 거의 찾아보기 힘들다는 단점을 극복할 수 있다는 점에서 유의미한 연구 주제라고 할 수 있다(Ehrich, 2005; Deshors, 2010, 2014; Yoon & Lee, 2016).

## 2. 연구방법

### 2.1 TOEFL11 코퍼스

TOEFL11은 2014년도 미국의 공인시험을 담당하는 ETS(English Testing Service)가 LDC(Linguistic Data Consortium)를 통해 공개한 코퍼스로서, 서로 다른 모국어를 구사하는 영어학습자들이 주제별로 쓴 에세이로 구성되어 있는 코퍼스이다(Blanchard et al. 2013). 이 코퍼스는 11개의 서로 다른 모국어를 구사하는 영어학습자들이 각각 영어로 작성한 1,100개의 에세이를 구성하여, 총 12,100개의 에세이로 구성되어 있다. TOEFL11의 구성 내용과 관련하여서는, 각 텍스트 파일이 에세이를 작성한 영어학습자의 모국어(L1)에 대한 정보와 전문적인 훈련을 받은 에세이 채점자들의 채점결과인 상중하의 영어 구사능력을 제외하고는 아무런 정보가 담겨있지 않는 원시코퍼스(raw corpus)

이다.

TOEFL11은 2014년도 공개되었지만, 2006년에서 2007년 사이에 TOEFL iBT® 시험을 보는 동안 작성한 독립형 에세이들로 구성되어 있다. TOEFL 글쓰기는 두 섹션으로 이루어져 있고, 첫 섹션에서는 통합형 글쓰기(Integrated Writing) 과제가 주어지고, 두 번째 글쓰기 과업으로는 독립형 글쓰기(Independent writing)가 주어진다. 독립형 글쓰기 과업에서는 8개의 프롬프터 가운데 하나가 주어지고, 이 주어진 주제에 대한 자신의 의견을 구체적이고 논리적으로 설명해서 독자를 설득시키는 *persuasive writing*으로서 30분 동안 작성한다.

독립형 에세이는 두 명의 채점자가 정해진 기준에 의해 0에서 5점까지 점수를 매겨서 그 평균값을 산출하는 방식으로 채점이 진행된다. 만약 두 채점자의 점수 차이가 1점 이상이면 제3의 채점자가 최종 점수를 확정한다. 공개된 코퍼스에서는 6가지의 점수를 상·중·하(High, Medium, Low)의 3등급으로 묶었다. 다음의 표 2에서는 한중일 TOEFL 응시자들이 취득한 수준별 수를 보여주고 있다.

표 2 점수에 따른 각 언어별 에세이 수 (Blanchard et al. 2013)

언어	Low	Medium	High
Chinese	98	727	275
Japanese	233	679	188
Korean	169	678	253

참고로, 윤태진·이용훈(2016)에서는 TOEFL11 코퍼스에서 영어 응시자들이 받은 등급과 모국어에 따라 관찰되는 서법 조동사들의 분포를 분석하였다. 그리고 윤태진(2018) 및 윤태진(2019)에서는 영어 응시자들이 받은 등급과 에세이 내용에 담겨진 언어정보들을 사용하여 기계학습을 이용하여 주제 프롬프터를 자동으로 분류하는 연구와 비지도 학습(unsupervised learning) 방식 중 토픽모델링을 이용하여 주제 프롬프터를 유추하는 연구를 각각 수행하였다.

본 논문에서는 한·중·일의 극동 아시아 TOEFL 응시자들이 제출한 에세

이에 언어 정보를 수동으로 태깅하여 분석의 토대로 사용하고자 한다. 따라서 본 논문에서는 분석의 대상이 되는 에세이의 수는 총 3,300(=1,100×3)개이다. 그림 2는 TOEFL 11 코퍼스에서 조동사를 확인한 후, 조동사를 기준으로 프롭르터, 모국어, 취득 점수 등을 포함한 식별 정보와 연구의 초점이 되는 서법 조동사의 좌우에 공기되어 있는 단어들을 추출한 후 저장한 정보를 보여 주고 있다. 분석 결과 본 코퍼스에서 can, will, 그리고 may가 가장 빈번하게 관찰되었다(윤태진 · 이용훈, 2016). 그리고 한중일 사이의 조동사의 빈도수를 정규화하기 위해서 관찰되는 조동사의 수를 무작위로 1,000개씩 추출한 다음, 추출된 조동사와 문맥 정보를 이용하여 언어 정보를 코딩하였다.

A	B	C	D	E
1	N	Source	Right	
2	82	115526_P2_JPN_medium.txt	, young people have much time in the future; they	can draw their dream as they want and can do their b
3	863	1701182_P7_JPN_high.txt	ts because understanding ideas and concepts	can help students to improve their imagination
4	128	1373465_P2_JPN_medium.txt	an enjoy their free time more easily older people	can. Secondly, because generally young people
5	988	1953464_P8_JPN_high.txt	primary reason for my belief is that many people	can achieve their success by avoiding risks. For exam
6	1005	546758_P5_JPN_high.txt	as Los Angeles or Washington D.C. Only New Yorkers	can afford not to have cars. In most of the parts in
7	134	1340428_P1_JPN_medium.txt	c subjects: If you have many field knowledge, you	can adjust society, find what you really want to do, a
8	164	1064359_P6_JPN_low.txt	and take risks rather than only doing what they	can already know ow to do well because if people try
9	986	1425210_P1_JPN_medium.txt	once you can enjoy to learn the subject then you	can also get lots of knowledge. On the other hand, if
10	437	566002_P5_JPN_medium.txt	e fewer at the same time. Finally, it	can also reduce the rate of car crash accidents. Acco
11	772	1608155_P5_JPN_medium.txt	o not do something by all ourself I think. But we	can always go back to pass and think about how our an
12	125	1179881_P2_JPN_low.txt	pe of life. Young people almost think that they	can any thing and they can break some situations with
13	142	497082_P1_JPN_medium.txt	ledge, you can answer the questions wily and you	can appeal how you understand general today's events
14	1012	1027445_P7_JPN_medium.txt	l when they complete their education because they	can apply ideas and concepts, whereas facts can not b
15	42	1884003_P7_JPN_high.txt	ed the specific facts in addition to whether they	can apply the equations introduced. If students unde
16	606	782666_P2_JPN_low.txt	to find something fun anytime. Thus, young people	can apply to many thing and they can make their life
17	941	112715_P7_JPN_medium.txt	ackground information. After these processes, we	can approach to the stage to make answers or own opini
18	1003	102895_P6_JPN_medium.txt	y have more information that the books and people	can ask any thing which the book do not have. People
19	740	895335_P6_JPN_medium.txt	ke care of luggage. But if I travel in a group, I	can ask others or a tour guide to take a pictures. Al
20	224	1300821_P1_JPN_medium.txt	ld learn various things from him as my teacher. I	can ask the question what I can't solve, and everytim
21	403	1707244_P8_JPN_medium.txt	en risks. That's why they get successes. A chance	can be a risk. A risk can be a chance. Third, I thi
22	730	1064474_P5_JPN_high.txt	d we can use car without harming our planet. That	can be a strong element for the favor of cars. Se
23	339	1120880_P6_JPN_medium.txt	. Second, feeling and knowing different ways	can be achieved by knowing the native people. But ther
24	443	1098663_P8_JPN_medium.txt	because of following reasons. Firstly, success	can be achieved if you continue doing what you already
25	607	1021981_P8_JPN_medium.txt	fine the word success as to achieve something that	can be admired by others. So if there is someone that

그림 2 추출되어 정리된 조동사의 예

## 2.2 언어 정보 코딩

본 논문에서는 서법조동사의 유형에 따른 빈도 측정을 넘어서, 표 4에서 예시된 것과 같은 여러 언어학적인 정보들을 태깅하여, 조동사의 사용에 영향을 미치는 언어학적인 요소들이 무엇인지를 밝히고, 더 나아가 그러한 요소들을 토대로 경쟁 모델을 구축하는 것을 목적으로 하고 있다. 서법조동사를 중

심으로 언어적 정보를 태깅하는 것은 수작업으로 진행되었다. 가장 빈도수가 높은 서법조동사인 *can, may, will*을 대상으로, 선행연구와 비교를 할 수 있도록 한정된 기능태거를 정립한 후 태깅을 수행하였다.

표 4는 한중일 TOEFL 응시자들을 대상으로 추출한 자료를 바탕으로 태깅을 한 형태 단서, 통사 단서, 그리고 의미 단서를 열거하고 있다. 형태 단서와 관련한 것은 조동사의 형태, 조동사 다음에 동사구가 생략 여부, 조동사와 공기하여 동사의 상(*aspect*), 태(*voice*) 그리고 법(*mood*)의 유형들, 조동사를 포함한 문장의 주어와 관련한 형태론적인 정보들이다. 통사 단서와 관련하여서는 조동사가 부정어를 포함하고 있는지의 여부, 문장의 유형이나 절의 유형을 포함하고 있다. 의미 자질과 관련하여 조동사의 의미, 조동사를 포함한 문장의 동사구에 대한 Vendler(1957) 식의 분류, 주어의 유정성(*animacy*) 및 유정성의 종류, 그리고 문장이 속어나 은유의 의미를 가지고 있는지에 대한 정보들을 코딩하였다.

표 4 각 서법조동사의 특징을 밝히기 위한 태깅에 사용할 언어학적 변수

ID Tag Type	ID Tag	ID Tag Levels
언어	L1	Korean, Japanese, Chinese
형태 단서	Form	<i>can, may, etc.</i>
	Verb	be, have, enjoy, etc.
	Elliptic	yes, no
	Voice	active, passive
	Aspect	simple, progressive, perfect
	Mood	indicative, subjunctive
	SubjMorph	adj., adv., common noun, proper noun, relative pronoun, noun phrase, etc.
	SubjPerson	1, 2, 3
	SubjNumber	singular, plural
통사 단서	Neg	affirmative, negated
	SentType	declarative, interrogative
	CLType	main, coordinate, subordinate

의미 단서	Sense	epistemic, deontic, dynamic
	Vendler	accomplishment, achievement, process, state
	VerbSemantics	abstract, general action, action incurring transformation, action incurring movement, perception, etc.
	RefAnim	animate, inanimate
	AnimType	animate, floral, object, place/time, mental/emotional, etc.
	Use	idiomatic, literal, metaphorical

이상 모국어가 한·중·일인 영어 학습자들이 제출한 에세이에 언어 정보를 코딩하였고, 그 중 일본어에서 추출된 토큰에 대해 정보들이 태깅된 예를 그림 3에서 보여주고 있다.

A	B	M	N	O	P	Q	R	S	T	U	V	W	X	
1	N	Source	SubjPe	SubjNum	SubjRelNnt	Neg	SentType	CLTyp	Sense	Vendler	VerbSem	RefAni	AnimTy	Use
2	82	115526_P2_JPN_medium.txt	3	plural	plural	affirmative	declarative	main	dynamic	accomplishment	gen_action	animate	human	metaphorical
3	863	1701182_P7_JPN_high.txt	3	singular	singular	affirmative	declarative	subordinate	epistemic	process	gen_action	inanimate	mental	literal
4	128	1573605_P2_JPN_medium.txt	3	plural	plural	affirmative	declarative	subordinate	dynamic	process	gen_action	animate	human	literal
5	988	1953644_P8_JPN_high.txt	3	plural	plural	affirmative	declarative	subordinate	dynamic	accomplishment	gen_action	animate	human	literal
6	1005	546758_P5_JPN_high.txt	3	plural	plural	affirmative	declarative	main	dynamic	state	gen_action	animate	human	literal
7	134	1340428_P1_JPN_medium.txt	3	singular	singular	affirmative	declarative	main	epistemic	accomplishment	gen_action	animate	human	metaphorical
8	164	1064359_P8_JPN_low.txt	3	plural	plural	affirmative	declarative	subordinate	epistemic	state	perception	animate	human	literal
9	986	6425210_P1_JPN_medium.txt	2	singular	singular	affirmative	declarative	main	dynamic	accomplishment	gen_action	animate	human	literal
10	437	566002_P5_JPN_medium.txt	3	singular	singular	affirmative	declarative	main	epistemic	achievement	gen_action	inanimate	it	literal
11	772	5608155_P5_JPN_medium.txt	1	plural	plural	affirmative	declarative	main	dynamic	process	act_motion	animate	human	metaphorical
12	125	1179801_P2_JPN_low.txt	3	plural	plural	affirmative	declarative	subordinate	dynamic	achievement	gen_action	animate	human	literal
13	142	497082_P1_JPN_medium.txt	3	singular	singular	affirmative	declarative	main	dynamic	achievement	gen_action	animate	human	literal
14	1012	1027445_P7_JPN_medium.txt	3	plural	plural	affirmative	declarative	subordinate	dynamic	achievement	gen_action	animate	human	literal
15	42	1084003_P7_JPN_high.txt	3	plural	plural	affirmative	declarative	subordinate	dynamic	achievement	gen_action	animate	human	literal
16	606	762666_P2_JPN_low.txt	3	plural	plural	affirmative	declarative	main	dynamic	achievement	gen_action	animate	human	literal
17	941	1127159_P9_JPN_medium.txt	1	plural	plural	affirmative	declarative	main	dynamic	accomplishment	act_motion	animate	human	literal
18	1003	102895_P6_JPN_medium.txt	3	plural	plural	affirmative	declarative	main	dynamic	process	comm	animate	human	literal
19	740	89535_P6_JPN_medium.txt	1	singular	singular	affirmative	declarative	main	dynamic	process	comm	animate	human	literal
20	224	1300821_P1_JPN_medium.txt	3	singular	singular	affirmative	declarative	main	dynamic	process	comm	animate	human	literal
21	403	1072144_P8_JPN_medium.txt	3	singular	singular	affirmative	declarative	main	epistemic	state	copula	inanimate	abstract	literal
22	339	1066474_P5_JPN_high.txt	3	singular	singular	affirmative	declarative	main	epistemic	state	copula	inanimate	effect	literal
23	343	1120880_P1_JPN_medium.txt	3	singular	singular	affirmative	declarative	main	epistemic	accomplishment	gen_action	inanimate	mental	literal
24	443	1098663_P8_JPN_medium.txt	3	singular	singular	affirmative	declarative	main	epistemic	accomplishment	gen_action	inanimate	abstract	literal
25	607	1021981_P8_JPN_medium.txt	3	singular	singular	affirmative	declarative	subordinate	epistemic	achievement	gen_action	inanimate	abstract	literal

그림 3 태깅된 자료의 일부 예시

### 3. 결과 분석 및 논의

형태, 통사 및 의미 정보를 코딩한 자료는 경쟁 모델을 구축하기 위한 입력 자료로 활용이 되었다. 경쟁 모델을 구축하기 위해 우선 파악해야 하는 것은 빈도수를 계산하는 것이다. 윤태진 · 이용훈(2016)은 TOEFL11 코퍼스에서 서법 조동사가 범언어적으로 그리고 취득 점수에 따른 빈도수 조사를 하였다.

본 논문에서는 무작위로 추출한 1,000개의 조동사 중 가장 빈번하게 관찰되는 조동사를 살펴보았는데, 무작위 추출 자료에서도 윤태진·이용훈(2016)에 서와 같이 가장 빈번하게 나타나는 조동사는 can, will 그리고 may였다.

그림 2는 TOEFL11 코퍼스에서 발견되는 can, will, may의 형태가 한·중·일 모국어 화자들에 따라 나타내는 빈도수를 보여주고 있다. 한·중·일 모국어 화자들 모두에서 서법 조동사 can이 가장 빈도수가 높았으며, may는 전반적으로 빈도수가 can보다 현저하게 낮게 관찰되었다. 윤태진·이용훈(2016)에 의하면 will의 빈도수가 높은 것은 에세이에서 사용되는 프롬프터 8개 중 하나가 “In twenty years, there will be fewer cars in use than there are today”인데, 미래의 상황에 대해 예측을 묻는 프롬프터로 인해 will이 빈번하게 사용된 것으로 관찰되었다. 그럼에도 불구하고 can이 가장 높은 빈도수를 보인다는 점은 여러 선행연구에서 보고한 내용과 맥을 같이 한다(윤태진·이용훈 2016 참조)

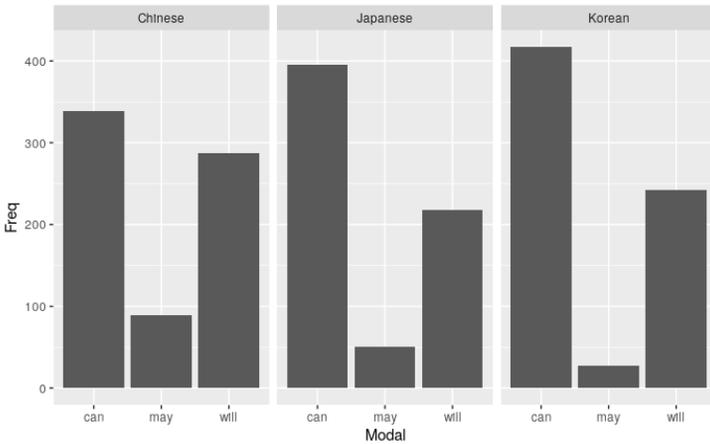


그림 4 can, may, will의 언어별 빈도수

본 논문에서는 그림 4에서 제시된 분포를 가진 자료들을 기반으로 하여, 서법 조동사 can, will, may에 영향을 미치는 형태론적, 구문론적 및 의미론적 특성들이 어떤 것이 있는지를 살펴보고, 그러한 특성들이 어떻게 언어마다 혹은 서법 조동사마다 유사성을 보이는지 살펴보고자 하겠다. 다만 지면상 태깅한

모든 언어 정보에 대한 자료를 설명을 하는 것 대신 일부 눈여겨볼 만한 언어 정보에 대한 설명을 하고, 군집 분석을 하도록 하겠다. 본 논문에서 살펴보고자 하는 개별적인 언어 정보로는 조동사의 의미(표 4의 Sense), 조동사와 연결한 동사의 종류(표 4의 VerbSemantics), 조동사와 연결한 동사의 Vendler에 기반한 유형(표 4의 Vendler), 조동사와 공기하는 동사의 유형(Verb), 그리고 주어의 유형성 유형(표 4의 AnimType)을 들 수 있다.

### 3.1 단일 요소 분석

서법조동사들을 그 의미에 따라 크게 인식론적(epistemic) 서법성, 의무적(deontic) 서법성, 그리고 동적(dynamic) 서법성의 세 가지 범주로 구분한다. 본 논문에서도 can, will 및 may 등 조동사의 의미를 인식론적(epistemic), 의무적(deontic), 그리고 동적(dynamic) 의미로 분류하여, 서법 조동사의 의미에 따른 한·중·일 언어 간 차이를 살펴보았다. 그림 5는 한중일의 TOEFL 응시자들이 서법 조동사 can, will, 그리고 may와 관련하여 사용한 의미의 빈도를 보여주고 있다.

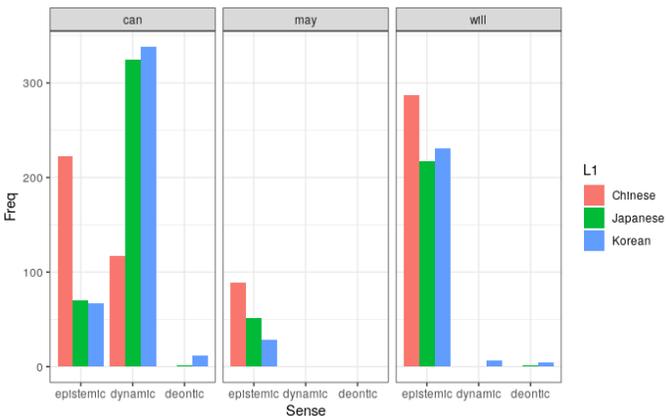


그림 5. 서법 조동사 별 및 언어별 서법 조동사의 의미 분포

그림 5에 따르면, 전반적으로 can은 동적(dynamic) 및 인식론적(epistemic) 의미를 가장 많이 사용한 것으로 나타났으며, may는 다른 조동사에 의해서 빈도수는 낮지만 인식론적(epistemic) 의미로, 그리고 will도 인식론적(epistemic) 의미로 사용되는 것을 관찰할 수 있다. 여기서 주목할 만한 사실은 한국어와 일본어를 L1으로 구사하는 영어 학습자들과 중국어를 L1으로 구사하는 학습자들 사이에 의미 사용빈도에 있어서 차이점을 보인다는 점이다. 예를 들어, can이 인식론적(epistemic) 의미로 사용된 경우, 중국어 학습자의 빈도수가 한국어와 영어 학습자의 빈도수보다 월등히 많으며, 한국어와 일본어 학습자들 사이에서는 차이가 거의 없다는 것을 알 수 있다. can이 인식론적(epistemic) 의미로 사용되는 경우도 중국어 학습자의 경우는 빈도수가 높은 반면, 한국어와 일본어 학습자들은 상대적으로 빈도수가 낮으며, 빈도수가 낮은 이 두 언어 사이에서는 별다른 차이가 없음을 관찰할 수 있다. will의 경우도 can의 인식론적(epistemic) 의미를 나타내는 경우와 마찬가지로 패턴을 보이고 있다. 이와 같은 관찰을 기반으로 하였을 때, 중국어화자의 L1에서의 문법과 한국어 및 일본어 화자의 L1에서의 문법이 영어의 조동사 사용에 영향을 미쳤을 것으로 해석할 수 있을 것이다.

Deshors(2010: 135)는 영어 화자 및 프랑스를 모국어로 하는 영어 학습자의 조동사 중 may와 can에 대한 의미 분석을 하였다. Deshors에 의하면 can은 동적(dynamic) 의미로, may는 인식론적(epistemic) 의미로 가장 많이 사용된다고 하였다. 본 논문에서도 중국어를 모국어로 하는 영어 학습자를 제외하고는 유사한 결과를 보여주고 있다. 이에 더해, 본 논문의 분석 결과 will 또한 may와 마찬가지로 인식론적인(epistemic) 의미가 가장 많이 사용된다는 것을 관찰할 수 있다. 인식론적인 의미를 will이 가장 많이 드러내고 있다는 것은 미래의 자동사 이용에 관한 프롬프터(“*In twenty years, there will be fewer cars in use than there are today*”)에 대한 에세이를 작성한 결과에 기인한다고 설명할 수 있다.

다음으로는 조동사와 공기하는 동사의 종류에 대해 살펴보기로 한다. 조동사와 공기하는 동사의 종류 중 한중일 전체 데이터 중 빈도수가 10개 이상인 것을 그림 6에 제시하였다. 두드러진 특징은 can은 다양한 동사와 동기해서

사용이 되는 반면, may는 주로 be동사와 공기하는 것으로 보이는 점과 will은 be와 have 동사와 주로 공기가 된다는 점이다. 즉, can은 be, get, do, enjoy, make, go, have, use, find, learn, see, understand 등 다양한 동사들과 공기(co-occur)하는 반면, may는 be 동사와만, 그리고 will은 be 동사와 have 동사와만 일반적으로 공기하는 현상을 보여주고 있다. Deshors(2010; 140)에서도 may와 can이 유사한 패턴을 보이고 있음을 보고하고 있다. Deshors(2010)에 의하면, may는 copula와 공기하며, can의 경우는 특별히 선호하는 동사가 있는 것 같지는 않다고 한다(p. 140). 본 논문에서도 can과 may와 관련하여 공기하는 동사들이 보이는 양상을 관찰해 보면, can은 많은 동사들이 공기하는 반면 may는 특별히 be동사와 공기하는 양상을 보여주고 있다. 이와 관련하여 Deshors는 can은 더 일반적인 어휘 항목(more generic lexical items)과 함께, may는 더 제약되고 특정한 어휘 항목(more restricted and more specific lexical items)과 함께 나타나고 있다고 설명하고 있다. Deshors(2010)에서는 조사되지는 않았지만, 본 논문의 연구 결과는 will의 경우도 be 동사와 have 동사를 선호하여 공기하는 양상임을 보여주고 있다.

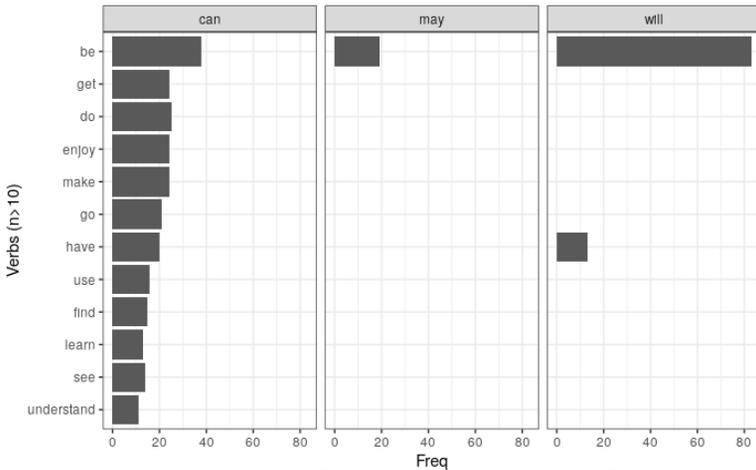


그림 6. 서법 조동사와 공기하는 동사의 종류

그림 7은 위의 그림을 언어마다 분류해서 다시 보여주고 있다. 대체적인 경향은 비슷하며 can과 공기하는 동사의 종류가 한국어에서 가장 다양하게 나타나며, 일본어와 중국어에서는 좀 더 제한적인 동시들과 공기하는 양상을 보여주고 있다. 또한 will과 공기하는 have 동사는 주로 일본어 영어 응시자에 게서 나타난다는 점을 관찰할 수 있다.

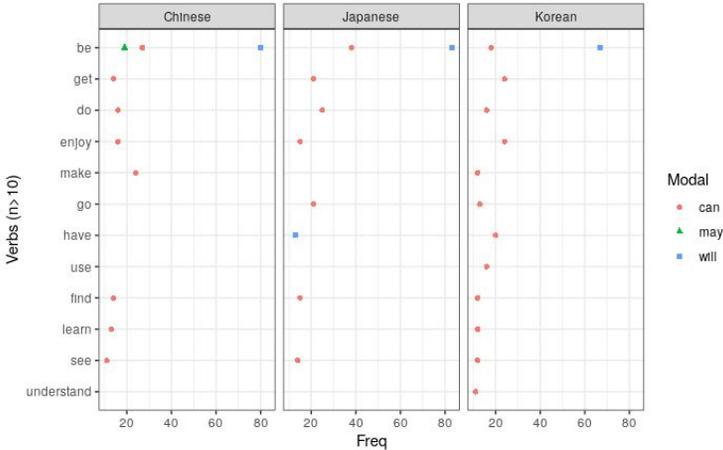


그림 7. 언어별로 구분한 서법 조동사와 공기하는 동사의 종류

조동사와 공기하는 동사를 Vendler(1957)가 제안한 유형으로 분류한 후, 조동사와 공기하는 분포를 살펴보고자 한다. Vendler(1957)는 “동사와 시간 (Verbs and times)”이라는 논문에서 동사구들이 이 동사구들이 표상하는 상태나 행위가 시간상에서 어떻게 분포되어 있는가에 따라 accomplishment(한정된 기간을 암시하는 동사구), achievement(한정된 순간을 암시하는 동사구), process(한정되지 않는 기간을 암시하는 동사구), 그리고 state(한정되지 않은 순간을 암시하는 동사구)의 4가지 종류로 구분하였다(Griffiths, 2006, Deshors 2010).

Deshors(2010: 142)에 의하면 may는 state 동사를 선호하는 것으로 can는 process, achievement, 그리고 achievement를 선호하는 경향이 있음을 보고하고

있다. 그림 8을 살펴보면, 한·중·일 언어마다 Vendler의 동사구 분류가 사용되는 양상이 달라짐을 관찰할 수 있다. can의 경우 4가지의 동사 유형 중 process와 가장 많이 공기하는 것으로 관찰된다. will의 경우 중국어 화자의 경우는 state와 가장 많이 공기하는 반면, 한국어 화자들은 process와 가장 많이 공기함을 관찰할 수 있다. 일본어 화자인 경우는 state, process, 그리고 achievement 등 차이를 많이 보이지 않은 채 사용됨을 관찰할 수 있다. 또한 주목할 만 한 점은 한국어 화자의 경우 can을 process와 공기해서 가장 빈도수가 높게 사용하는 반면, accomplishment와는 거의 공기하지 않는다는 점을 보여주는 데, 이는 다른 모국어화자들과 비교되는 점이다. 중국어 화자의 경우 can이 accomplishment와는 일반적으로 공기되지 않는 반면, achievement, state, process와 공기되는 형태를 보여주며, 일본어 화자의 경우 can이 process와도 공기가 빈번하게 되지만 achievement와 accomplishment와도 잘 공기하는 형태를 보여주고 있다. 따라서 Vendler의 동사구 유형과 관련한 서법 조동사의 공기 형태는 한·중·일 세 언어에서 차이를 보이고 있음을 관찰할 수 있다.

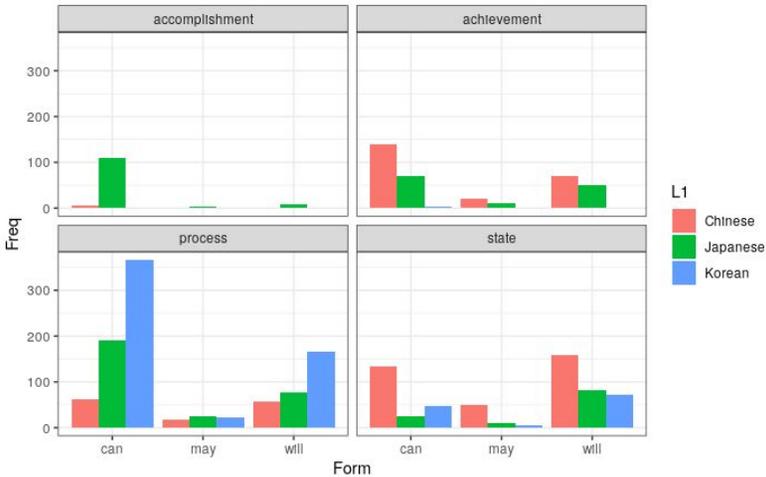


그림 8. Vendler의 동사구 유형과 서법조동사의 공기

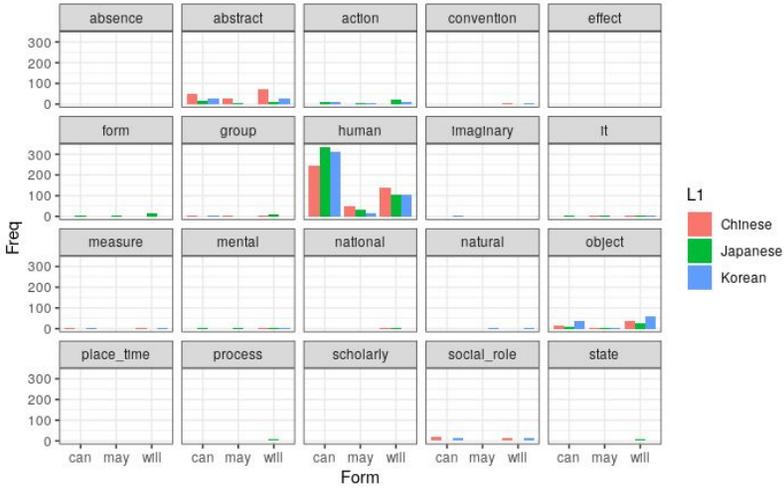


그림 9. 의미 자질 중 Animacy Type에 따른 조동사의 빈도수

마지막으로 서법 조동사를 포함한 문장의 주어의 유정성 유형(Animacy type)에 대해 살펴보고자 한다. 그림 9는 유정성 유형과 관련된 결과이다. 비록 많은 종류의 유정성 유형(Animacy Type)을 코딩하였지만, 일반적으로 인간(human)이 주어로서 현저하게 많이 사용되었음을 관찰할 수 있다.

### 3.2 다중 요인 분석

이상 언어 정보 중 조동사의 의미, 공기하는 동사의 유형, Vendler 기반 동사구 분류와의 공기, 유정성 유형(Animacy Type) 등을 통해 이러한 단일 요인들이 한·중·일을 모국어로 하는 영어학습자들의 서법조동사 사용에 있어서 보이는 유사성과 차이점을 살펴보았다. 한정된 지면상의 이유로 비록 코딩한 모든 언어 정보들을 살펴보는 못했지만, 일부 형태, 통사 및 의미 단서들을 살펴봤을 때, 언어마다 차이를 보이는 자질도 있었으며, 언어끼리 유사성을 보이지만 서법 조동사간에 차이를 보이는 자질도 있었다. 종합해 보면 서법 조동사가 언어별로 복잡한 양상을 보인다고 할 수 있다.

이러한 단일요인 분석은 백터에 기반을 해서 일차원상의 분포를 보여준다거나, 두 쌍의 백터를 비교하여 유사성을 계산할 수 있다는 장점이 있다. 하지만, 우리가 앞서 이론적인 틀로 채택할 경쟁모형을 실현시키기 위해서는, 단일요소 분석이 아니라 다중요소 분석을 채택해야할 필요성이 있다(Gries, 2003; Deshors 2010; Deshors, 2014; Deshors & Gries 2014). 행동 프로파일에서 추출한 관련 정보에 적용될 수 있는 다중요인 분석은 여러 가지가 있다. 본 논문에서는 언어자료 분석에 적합하고, 널리 사용되는 기법인 계층적 병합식 군집 분석(HAC; hierarchical agglomerative cluster analysis)이라는 탐색적인 기법을 사용하였다(Manning and Schütze 1999; Baayen, 2008; Johnson 2008).

계층적 병합식 군집 분석(HAC)은 서로 다른 항목 사이의 차이점이나 유사성을 확인하고 표상해 주는 군집 분석 방법이다. 일반적으로 군집은 항목들의 유사성 매트릭스에 기반을 두거나 항목들의 특징이 되는 변수들을 기반으로 수행된다. 유사성은 일반적으로 유클리드 거리(Euclidean distances)를 사용하여 측정하는 경우가 많다. 유사성 매트릭스가 생성된 후에는, 합병 전략(amalgamation strategy)을 선택한다. 합병 전략은 군집을 이룰 항목들이 어떻게 군집을 이룰지를 결정하는 알고리즘이다. 이를 위해서는 Ward의 규칙(Ward's rule)이라고 하는 합병 전략이 널리 사용된다(Murtagh & Legendre, 2014).

그림 10은 본 논문에서 사용한 데이터를 계층적 병합식 군집 분석의 결과로 만들어진 계층적 수형도이다. 이 계층적 수형도는 여러 군집을 가지고 있다. 이 군집들은 군집 내 유사성(within-cluster similarity)은 높게, 군집 간 유사성(between-cluster similarity)은 낮게 하는 특징을 가지고 있다. 그러한 계층적 수형도에서 뽑아낸 정보는 수형도가 주어진 다중요소적인 자료의 복잡성에서 사람이 거의 식별할 수 없는 정보를 용이하게 요약해 준다는 점에서 의의를 찾을 수 있다. 구체적으로, 그림 10에서 우리는 조동사 can의 경우 한·중·일 화자들이 하나의 군집을 이룬 것을 볼 수 있다. 또한 may와 will의 경우는 한·일이 군집을 이루고 있으며, 중국어 화자들의 will과 may는 분리되어 있음을 볼 수 있다. 이상의 결과를 요약해 보면, can의 경우는 한·중·일

이 유사성을 보여주며, may와 will의 경우는 한·일 화자들 사이에서만 유사성이 있으며 중국어 화자의 may와 will과는 차이를 보인다고 할 수 있다. 이와 같은 군집 결과는 언어 유형론적으로 한·중·일의 세 언어를 분류할 경우, 한국어와 일본어가 하나의 무리를 이루고, 중국어는 다른 유형에 속하는 언어라는 것을 고려해보면 타당한 군집이라고 할 수 있을 것이다.

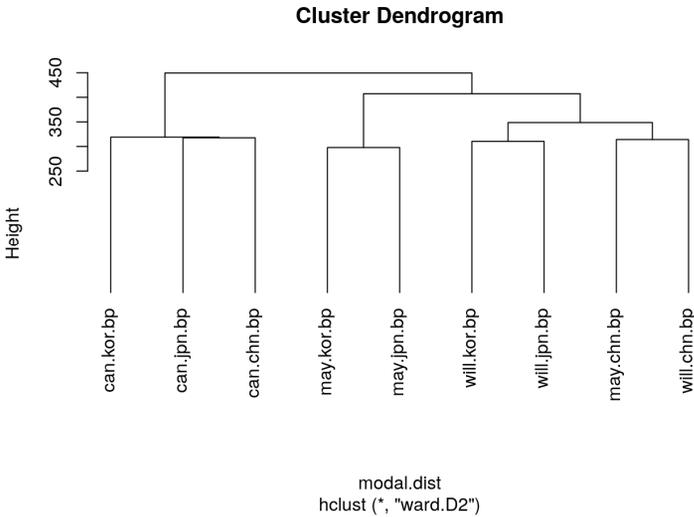


그림 10. 계층적 병합식 군집 분석(cluster analysis)

### 3. 결론

본 논문에서는 유형별로 서로 다른 모국어어를 구사하는 한·중·일 모국어 화자들을 통해 서법 조동사가 언어유형별에 따라 어떤 유사점과 차이점을 보이는지를 단일요인분석과 경쟁모델을 구현한 다중요인분석을 통해 살펴보았다. 본 논문에서는 이러한 분석을 위해 TOEFL11이라는 학습자 코퍼스에서 한·중·일의 TOEFL 응시생들이 제출한 에세이를 사용하여 서법조동사에

대한 연구를 진행하였다.

본 논문을 끝맺기 전에, 서법조동사의 의미(sense)를 통해 한중일 화자들의 L1이 L2에 영향을 미치며, 이러한 영향으로 이해 경쟁 모델에서 한중일 화자들의 군집이 언어유형별로 형성된다는 점을 살펴보도록 하겠다. 서법 조동사의 의미와 관련하여, Stephany(1995)는 영어, 독일어, 그리스어와 같은 인도-유럽어의 서법성(modality) 습득을 검토하였다. 그녀에 의하면, 어린이들의 언어 자료에서 인식론적 서법성(epistemic modality)이 의무적 혹은 역동적 서법성(deontic or dynamic modality)보다 더 나중에 나타난다고 한다. 인식론적인 서법성이 다른 종류의 서법성보다 늦게 나타나는 것은 아이들의 인지 발달, 아이들의 언어 사용의 목적, 아이들이 노출된 언어 자료의 본질 등을 포함해서 여러 가지 요인들이 복합적으로 작용하기 때문이라고 제안하였다.

한국어의 서법성의 습득에 관한 연구는 Choi(1995)가 있다. 그런데 Choi(1995)는 Stephany(1995)의 연구와 차이를 보이고 있다. 그 차이점은 바로 한국어 화자의 경우 인식론적 서법성이 의무론적 서법성보다 더 늦게 습득되는 것은 아니라는 것이었다. Choi(1995)는 그 이유가 한국어가 가지고 있는 서법을 표시하는 특성에 기인하기 때문이라고 생각하였다. 한국어는 양태(modality)를 표현하기 위해 조동사(auxiliary verbs)와 문미 접사(sentence-ending suffixes) 두 유형의 형태 부류를 사용한다. (2)에서 살펴볼 수 있듯이, 의무 양태(deontic modality)는 특정 연결사 뒤에 위치하는 조동사에 의해 표현되는 반면, 인식 양태(epistemic modality)는 조동사 혹은 문미 접사에 의해 문법화되어 있다.

(2) Mary-는 숙제-를 해야 돼.

영화-TOP homework-OBJ do-CONN become.

'Mary must do homework'

반면 중국어는 교착어인 한국어와는 달리 고립어로서 서법조동사와 어기조사를 사용하여 양태의 의미를 실현하는 방법이 있다. 서법조동사는 항상 동사

구 앞에 나오고 어기조사는 일반적으로 한 문장의 마지막에 나온다. 예를 들면, 다음의 예문 (3)과 같다.

(3) 我会说英语 [wǒ huì shuō yīngyǔ]

I can speak English

이상의 한국어와 중국어에서 양태를 표현하는 방식에 따르면, 우리는 영어 서법조동사와 유사한 중국어 모국어화자들이 영어 화자들과 유사한 형태를 보이며, 한국어 및 일본어 모국어 화자들은 중국어와는 차이점을 보인다고 추측할 수 있다. 즉 L2의 다양성을 설명하는 데 L1이 어느 정도 영향을 미치는 것으로 해석할 수 있다.

이상 본 논문에서는 여러 형태 단서, 통사 단서, 그리고 의미 단서들이 가지고 언어별 및 서법 조동사별 특징을 단일 요소 분석을 통해 알아보았고, 이러한 요소들을 조합하여 계층적 군집 분석을 하였다. 계층적 병합식 군집 분석을 수행하여 한중일 세 언어를 모국어로 하는 화자들이 L2 에세이에서 어떻게 군집을 이루는 지 살펴보았다. 한중일의 동북아시아권에 속하는 언어뿐만 아니라, TOEFL11의 나머지 L1들에 대해서도 분석을 하여 여러 언어들이 가지는 유사성 및 특이점을 살펴볼 필요가 있다.

## 인용문헌

- 김현진. (2014). 한국어 문장 이해 전략으로 조사, 어순, 생물성 단서의 사용 연구 - 중국어권 한국어 학습자를 중심으로-. 『한국어 교육』 25(3): 29-58.
- 윤태진. (2019). TOEFL11 을 이용한 비지도 토픽 모델링. 『언어과학』 26(1), 51-70.
- 윤태진. (2018). TOEFL11 코스에서 주제 프롬프터의 자동 분류. 『인문언어』 20(1), 157-177.
- 윤태진. (2018). 대응 분석을 통한 서법 조동사의 범언어적 분포 연구. 『인문언어』 20(2), 311-330.

- 윤태진 & 이용훈. (2016). 학습자 코퍼스를 이용한 서법조동사의 분포의 범언어적 연구. 『인문언어』 18, 137-154.
- Aijmer, K. (2002). Modality in advanced Swedish learners' written interlanguage. In S. Granger, J. Hung & S. Petch-tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, pp. 55-76, Amsterdam: John Benjamins.
- Adjemian, C. (1976). On the nature of Interlanguage systems. *Language Learning* 26.2:297-320.
- Baayen, R. H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics*. Cambridge: Cambridge University Press.
- Bates, E. and B. MacWhinney. (1982). Functionalist approaches to grammar. In Wanner, E. and L. Gleitman (eds.) *Language Acquisition: The State of the Art*, 73 - 218. Cambridge: Cambridge University Press.
- Bates, E. and B. MacWhinney. (1989). Functionalism and the Competition Model. In MacWhinney, B. and E. Bates (eds.) *The Cross-linguistic Study of Sentence Processing*, 3 - 73. Cambridge: Cambridge University Press.
- Blanchard, D., J. Tetreault, D. Higgins, A. Cahill, and M. Chodorow. (2013). *TOEFL11: A Corpus of Non-native English*. ETS RR - 13-24. Princeton, NJ: Educational Testing Service.
- Choi, S. (2015). Acquisition of meaning. In N. Riemer (ed.), *The Routledge Handbook of Semantics*. (pp. 457-472). Oxford, UK: Routledge.
- \_\_\_\_\_. (1991). Early acquisition of epistemic meanings in Korean: a study of sentence-ending suffixes in the spontaneous speech of three children. *First Language* 11: 93 - 119.
- Deshors, S. (2014) Constructing meaning in L2 discourse: The case of modal verbs and sequential dependencies. In Glynn, Dylan and M. Sjölin (eds.) *Subjectivity and Epistemicity: Corpus, Discourse and Literary Approaches to Stance*, 329-348. Lund: Lund University Press.
- \_\_\_\_\_. (2010). *Multifactorial Study of the Uses of May and Can in French-English Interlanguage*. Ph.D. dissertation, University of Sussex.
- Deshors, S. and S. Gries. (2014). A case for the multifactorial assessment of learner language: The Uses of May and Can in French-English Interlanguage. In Glynn, D. and J. Robinson (eds.), *Corpus Methods for Semantics: Quantitative*

- Studies in Polysemy and Synonymy*, 179-204. Amsterdam: John Benjamins.
- Gries, S. (2010a). Behavioral Profiles: A fine-grained and quantitative approach in Corpus-based lexical semantics. *The Mental Lexicon* 5.3: 323-346.
- \_\_\_\_\_. (2010b). Behavioral Profiles 1.01: A Program for R 2.7.1 and Higher.
- \_\_\_\_\_. (2003). *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement*. London: Continuum Press.
- Gries, S. and N. Otani. (2010). Behavioral Profiles: A corpus-based perspective on synonymy and antonymy. *ICAME Journal* 34: 121-150.
- Griffiths, P. (2006). *Introduction to English Semantics and Pragmatics*, Edinburgh University Press.
- Hanks, P. (2000). Do word meanings exist? *Computers and the Humanities* 34.1/2:205-215.
- Jarvis, S. (2000). Methodological rigor in the study of transfer: Identifying L1 influence in the Interlanguage lexicon. *Language Learning* 50.2:245 - 309.
- Johnson, K. (2008). *Quantitative Methods in Linguistics*. Oxford: Blackwell.
- Jun, K. and Y. Lee. (2015). A statistical analysis of can and may in British and Indian English. *English Language and Linguistics* 21.3, 63-84.
- Manning, C., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: The MIT Press.
- Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion?. *Journal of Classification* 31(3), 274-295.
- Odlin, T. (1989). *Language Transfer*. Cambridge: Cambridge University Press.
- Oh, S.-Y. (2007). A corpus-based study of epistemic modality in Korean college students' writings in English. *English Teaching*, 62(2), 147-175.
- Park, H. (2012). Modality in Korean Learner's Spoken Interlanguage. *English Language & Literature Teaching*, 18(1), 197-216.
- R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Selinker, Larry. (1969). Language transfer. *General Linguistics* 9.2:67-92.
- Stephany, U. (1995). Function and form of modality in first and second language acquisition. In *From Pragmatics to Syntax: Modality in second language acquisition*, A. Giacolone-Ramat & G. Crocco-Galeas (eds), 105-120. Tübingen:

Gunter Narr.

Vendler, Z. (1957). Verbs and times. *The Philosophical Review*, 143-160.

Yoon, T. & Lee, Y. (2016). A Multifactorial Analysis of Can and May in Korean EFL Learners Writings. *English Language and Linguistics*, 22(1), 117-136.

[Abstract]

**Examining Uses of Modal Auxiliaries by Learners of English  
from Three East Asian Language Users**

Tae-Jin Yoon · Yong-Hun Lee

(Sungshin Women's University, Professor · Chungnam National University, Lecturer)

The choice and use of modal auxiliaries depends on multiple grammatical factors in the sentential environment. The most frequently occurring auxiliary verbs in the TOEFL 11 corpus were extracted from the Chinese, Japanese, and Korean subset of the corpus, and manually coded with a number of morphological, syntactic and semantic features. The paper presents examples of linguistic features such as modal semantics, the co-occurring verb types, and semantic classes of verb phrase proposed by Vendler, as well as the animacy type of the subject shows similarities or differences among the three groups of language learners. The behavioral profile based analysis of the linguistic features revealed that the modal *can* show similarities among the Chinese, Japanese, and Korean. Korean and Japanese learners show a clustering of *May* and *Will*, while Chinese learners does not form a group with the Korean and Japanese learners regarding *may* and *will*. The paper contributes to expanding our understanding on the choice of modal auxiliaries by the learners of English in the three East-Asian countries.

Keywords: Chinese, Japanese, Korean, modal auxiliaries, behavior profile, monofactorial analysis, competition model, clustering analysis, TOEFL 11

윤태진 (Yoon, Tae-Jin), 제1저자  
성신여자대학교 영어영문학과, 교수

02844 서울특별시 성북구 보문로 34다길2

02-920-7185

tyoon@sungshin.ac.kr

이용훈 (Lee, Yonghun), 교신저자

충남대학교 영어영문학과, 강사

34134 대전광역시 유성구 대학로 99 인문대학(W7) 446호

010-7797-7844

yleeuiuc@hanmail.net

접 수 일: 2020년 05월 20일

심사기간: 2020년 05월 21일~06월 17일

게재결정: 2020년 06월 18일