

문서번호: 한국음운론학회 2022-게예증01

## 논문 게재예정 증명서

귀하의 건승을 기원합니다.

한국음운론학회의 학회지인 『음성·음운·형태론 연구』에  
다음과 같이 귀 논문이 게재될 예정임을 확인합니다.

- 논문 제목: 이름의 성별에 나타나는 음성상징성: 동물의 숲 주민 캐릭터  
이름을 중심으로
- 논문 저자: 문 성 민 (조선대학교, 교신저자)
- 예정 호수: 28집 1호
- 발행 예정일: 2022년 4월 30일

2022년 4월 20일

한국음운론학회 회장



우) 17579 경기도 안성시 중앙로 327

한경대학교 인문융합공공인재학부 영미언어문화전공 안미연 교수실

한국음운론학회 <http://phonology.or.kr>, 한국음운론학회 JAMS <https://phonology.jams.or.kr>

연락처: 편집위원장 윤관희, E-mail: [ghyun@daegu.ac.kr](mailto:ghyun@daegu.ac.kr) (Tel:

010-2045-8598)<http://phonology.or.kr>, 한국음운론학회 JAMS <https://phonology.jams.or.kr>

## 이름의 성별에 나타나는 음성상징성: 동물의 숲 주민 캐릭터 이름을 중심으로

**Sound symbolism according to gender: Focusing on the names of residents' characters in the Animal Crossing game, *Studies in Phonetics, Phonology and Morphology* 0.0. 00-000.** Sound symbolism reflected in the segmental composition of personal names makes it possible to predict the gender associated with the names. In contrast to much research in English on this issue, not much is known about sound symbolism in the gender of personal names in languages beyond English. This study investigate the sound symbolism that appears in the names of the Korean game characters in Animal Crossing. We compare the phonological features in the two genders of the residents' names in the game by chi-square test and report what phonological features have effects on predicting the gender of game character names in Korean by fitting a logistic regression model. Results show that if more numbers of syllables and sonorants exist in the name, the gender of the name is more likely to be female. If obstruents and stops account for a large portion of the name, the gender of the name is more likely to be male. Furthermore, results of logistic regression analysis show that if obstruents account for a large portion in the name and the name ends in sonorants, consonants, or vowels, the gender of the name can be predicted as male. If the name starts with an obstruent, the gender of the name can be predicted as female.

Keywords: gender, sound symbolism, naming, Korean phonology

### 1. 머리말

음성상징이란 언어와 그 언어가 의미하는 것 사이에는 아무런 관련이 없다는 언어의 자의성과 대비되는 개념으로(Saussure 1966), 특정 음성이 특정한 이미지나 의미와 연관되는 경향이 있음을 가리키는 용어이다(Sapir 1929, Kawahara et al 2018). 음성상징성을 보여주는 예시로 의성어와 의태어가 있으며 이들은 단어의 음운론적 구조가 가리키는 소리나 모양과 자의적 관계가 아닌 연관성을 지닌다. 또한 모음이나 자음의 변화에 따라 단어가 의미하는 소리나 모양 크기가 커지거나 무거운 의미를 가지기도 한다. 예를 들어 한국어의 ‘깡충깡충’과 ‘경충경충’의 경우, ‘경충경충’이 ‘깡충깡충’보다 더 무겁고 큰 형태의 의미를 가진다(e.g., 김재원 2003, 김선희 2019, 이영석. 1994). 이러한 음성상징어의 체계는 한국어뿐만 아니라 다른 언어에

서도 존재한다. Hamano (1986)는 일본어에 나타나는 음성상징어 체계에 대한 연구와 일본어의 역사적 음운 변화에 영향을 주는 음성상징어 체계에 대해 연구를 진행했다. 일본어의 경우 /pi/는 작은 동물이 우는 소리, 혹은 옷감이나 냅싯줄 같은 얇은 물체를 잡아당기는 형태를 의미하고 /pa/는 장난감 총을 쏘는 소리, 폭죽 혹은 폭탄이 터지는 소리와 형태처럼 크고 넓은 범위에서 발생한 소리나 형태를 의미한다. /a/가 들어간 음성을 /i/가 들어간 음성보다 더 큰 소리나 형태로 인식한다는 결과는 또 다른 언어에서도 드러난다. Sapir E(1929)는 영어의 모음과 자음의 특정한 차이에 따라 상징적 크기에 대한 느낌이 달라지는 가에 대해 실증적인 실험을 시도했다. 해당 연구에서는 식탁에게 *mal* 또는 *mil* 이라는 이름을 부여하고 실험 참여자들은 대상으로 두 이름을 가진 식탁 중 어떤 것이 더 크게 느껴지는 가에 대한 실험을 진행하였다. 연구의 결과, 실험 *mal* 이라는 이름을 가진 식탁을 더 커 보이게 느낀다는 결과를 제시하였다.

최근 음성상징성에 대한 예시로 일본에서 발매된 게임인 포켓몬스터에 나오는 캐릭터들인 포켓몬의 이름들을 대상으로 음성상징성을 연구한 결과가 있다(Kawahara et al 2018, Kawahara et al 2021, Kawahara and Kumagai. 2021.) 포켓몬의 이름이 음성상징성을 연구하기에 좋은 자료인 이유는 첫째, 이름의 수가 700 여 개가 넘어 정량적 분석이 가능하고 둘째, 포켓몬스터라는 게임은 포켓몬들끼리 자신만의 기술을 사용하며 대결을 하는 형식의 게임이기 때문이다. 따라서 각 포켓몬마다 크기, 무게, 체력, 속도 등의 수치화된 변수가 자세히 기록되어 있으며 이를 통해 이름에 들어간 음운론적 요소와 포켓몬의 수치화 된 변수가 어떤 상관관계가 있는 지 살펴볼 수 있다. 포켓몬의 이름들을 대상으로 한 연구들 중 Kawahara et al(2018)는 포켓몬 이름 데이터 안에서 나타나는 음운론적 특징들과 포켓몬들의 수치적 정보 사이의 주 효과와 관계를 보기 위해 스피어만 상관 분석과 다중 선형 회귀분석을 사용하여 분석을 진행하였다. 연구의 결과, 성대의 진동을 수반하는 소리인 유성 저해음(/b/, /d/, /g/, /z/)과 운율단위인 모라의 개수가 포켓몬의 크기, 무게, 진화 정도, 강함의 수치적 정보와 양의 상관관계에 있다는 것을 보고하였다.

음성상징의 또 다른 경우로 여성의 이름과 남성의 이름이 음운론적 특징에 따라 유의미한 차이가 나타나는 경우를 들 수 있다. 이어지는 다음 장에서부터는 외국어를 대상으로 진행된 사람 이름의 음성상징에 관한 선행연구를 살펴보고, 이후 한국어의 음운론적 특징에 대해 정리한 후, 마지막으로 본 연구에서 데이터 사용한 동물의 숲 게임과 게임 안에서 사용된 주민 캐릭터 이름의 특징에 대해 소개한다.

### 1.1 이름에 나타난 음성상징어에 관한 선행연구

사람 이름에 나타난 음성상징어에 관한 연구는 영어를 중심으로 많은 연구가 행해졌다. 예를 들어 Slater and Feinman(1985)은 사람의 이름에 나타나는 성별의 잠재적인 특징을 알아보기 위해 222 명의 여성 참가자와 267 명의 남성 참가자들에게 사람의 주어진 이름과 불렸으면 하는 이름을 적게 한 후 각 이름들의 음절 수와 분절음 수(모든 자음, 모음), 이름에 들어간 음절들이 자음 또는 모음으로 끝나는 지에 대한 여부, 이름의 시작과 끝 분절음(유성 자음, 무성 자음, 공명자음, 모음)의 분포, 2 음절 이상의 이름 일 경우 강세 위치를 조사했다. 분석 과정에는 앞서 조사된 성별에 따른 이름의 음운론적 특징을 변수로 사용하였으며 성별에 따른 이름에서 나타나는 음운론적 특징의 차이를 분석하기 위해 이원 분산 분석을 수행하였다. 연구의 결과, 여성의 이름이 음절 수와 분절음의 수가 더 많으며 공명음(/m/, /n/, /l/, /r/, /ŋ/, /y/, /w/)으로 끝나는 경우가 더 많다는 것을 발견하였다. Barry and Harper(1995)도 마찬가지로 성별에 따라 이름에 나타나는 음운론적 특징의 차이를 보기 위해서 음운론적 특징에 따라 점수를 매겼다. 점수를 매기는 방식은 크게 두 가지인데 하나는 이름의 전반적인 발음으로, 강세의 위치와 음절의 길이에 따라 점수를 부여했다. 또 하나는 이름의 끝 음소로, 이름이 모음, 공명자음, 저해음, 모음으로 끝나는 경우에 따라 점수를 부여했다. 두 가지 기준의 점수를 낸 평균을 분석한 결과, 여성의 이름은 공명음으로, 남성의 이름은 조음과정에서 조음기관에 의해 소리의 흐름이 방해를 받는 소리인 저해음으로 끝나는 경우가 많았다. 이를 통해 공명음으로 끝나는 이름일수록 여성의 이름임이 부각되고 장애음으로 끝나는 이름일수록 남성의 이름임이 부각된다는 것을 알 수 있다. Perfors(2004)는 사람의 이름과 매력도 간의 상관 관계에 대한 연구를 진행하였다. 연구의 분석을 위해 사람의 이름 데이터에서 모음을 발음할 때 혀가 앞에 위치해 있는 모음을 말하는 전설모음, 혀가 뒤에 위치해 있는 모음인 후설모음, 조음과정에서 조음기관의 방해를 받지 않고 나오는 자음인 공명자음, 그리고 저해음 등의 음운론적 특징을 변수로 추출하여 사용하였다. 연구의 결과, 저해음과 후설모음이 들어간 남자 이름, 공명자음과 전설모음이 들어간 여자 이름을 다른 이름들보다 더 매력적으로 느낀다는 것을 발견하였다. 또한 Moorthy et al.(2018)는 브랜드 이름에 나타난 성별 변이가 브랜드 마케팅 전략에 어떤 영향을 끼치는가에 대한 연구를 진행하였다. 이 연구에는 여성 60984 개, 남성 36118 개의 사람의 이름 말뭉치를 이용하여 성별을 예측하는 알고리즘을 만들어 브랜드 이름의 남성성을 예측했다. 분석을 위해 사람의 이름 데이터에서 음절, 자음, 모음 수 그리고 이름이 어떤 자음 또는 모음으로 시작하거나 끝나는 지를 추출하여 변수로 생성하였으며, 로지스틱 회귀를 사용하여 브랜드 이름의 성별을 예측하는 모델을 생성하였다. 연구의 결과, 이름이 /a/로 끝나면 여성의 이름, 공

명음, 파열음(i.e., 소리의 흐름이 조음기관에 의해 완전히 닫혔다가 열리면서 나오는 소리), 마찰음(i.e., 파열음처럼 완전히 닫히지 않고 조음기관이 닫히기 직전까지 가까워지며 그 과정에 소음을 내는 소리)으로 끝나면 남성의 이름으로 예측 가능하다는 것을 보고 하였다. 하지만 Slater and Feinman.(1985)의 연구, Barry and Harper(1995)의 연구 결과에서 이름이 공명음으로 끝나는 경우 여성의 이름이 많다는 결과를 제시하였다는 점을 고려할 때, Moorthy et al.(2018)의 연구 결과는 선행 연구 결과에 상충되는 결과임을 알 수 있다. 제시된 표 1 은 선행 연구들에서 사용된 이름 데이터의 연도, 지역, 이름의 개수를 나타내고 있으며, 표 1 을 통해 Slater and Feinman.(1985)의 연구가 다른 선행 연구보다 최신의 이름 데이터를 연구에 사용하였다는 것을 알 수 있다. 따라서 선행연구들의 상충적인 연구 결과와 선행 연구에서 사용된 이름 데이터의 연도를 고려하면 2000 년대 전에 사용된 이름의 경우 공명음으로 끝나면 여성의 이름일 확률이 높았지만 2000 년대 이후에는 공명음으로 끝나는 경우 남성의 이름일 확률이 높아지는 것으로 시간이 흐름에 따라 연구 결과가 변해하였다는 것을 알 수 있다.

표 1 성별에 따른 이름의 음운론적 특징의 차이에 대한 선행연구

연구	데이터		
	연도	지역	이름의 개수
Slater and Feinman (1985)	명시되지 않음	미국 로키 산 지역	489
Barry and Harper (1995)	1960	미국 펜실베이니아 주	여성 45,280, 남성 72,463
	1990		여성 25,899, 남성 37,748
Prefor (2004)	명시되지 않음	랜덤 채팅 웹사이트 (hotornot.com)	남녀 각각 24
Moorthy et al (2018)	1880-2013	미국	여성 60984, 남성 36118
	명시되지 않음	영국 통계청, 북아일랜드 통계청, 스코틀랜드 일반 관보청 데이터	

다양한 선행 연구에도 불구하고 대부분의 연구는 한국어가 아닌 영어를 기준으로 진행되었다. 한국어의 음운론적 특징은 이와는 다른 음운론적 특징과 차이점을 가지고 있다. 다음 절에서 이에 대해 정리한다.

## 1.2 한국어의 음운론적 특징

한국어는 다음과 같은 음운론적 특징을 가진다. 첫째, 저해음을 분류할 때 위에서 언급한 유성과 무성의 체계가 한국어의 자음 내에서 음소적 특성으로는 존재하지 않는다. 이에 반해 저해음 중 파열음과 파찰음은 조음 방식이 기식이 있는 지 또는 성대 긴장을 유발하는 지에 따라 평음, 유기음, 경음으로 나뉜다. 예를 들어 /tal(달)/, /tʰal(탈)/, /tʰal(딸)/이라는 음절들의 경우, 두음 /t(ㄷ)/, /tʰ(ㅌ)/, /tʰ(ㄸ)/의 조음 위치가 치경으로 같지만 /tʰ(ㅌ)/은 /t(ㄷ)/에 비해 기식이 심한 유기음이고 /tʰ(ㄸ)/은 /t(ㄷ)/에 비해 성대의 긴장을 유발하는 경음이다. 또한 이 음들은 ‘달’, ‘탈’, ‘딸’과 같이 의미를 구별하게 해주는 최소의 소리 단위이므로 각각의 음소라고 할 수 있다. 그러나 한국어에서 유성음 /b/, /d/, /g/ 등은 공명음 사이에서만 실현되며 의미 구별을 하게 해주는 기능이 없는 이음이다. 그 외 한국어의 자음 체계는 표 2에서 제시한 것과 같다.

표 2 한국어의 자음의 음소체계(강옥미 2003.)

			양순음	치경음	구개치경음	경구개음	연구개음	성문음
장 애 음	파 열 음	평음	p(ㅍ)	t(ㄷ)			k(ㄱ)	
		경음	pʰ(ㅍᄂᆞ)	tʰ(ㄷᄂᆞ)			kʰ(ㄱᄂᆞ)	
		유기음	pʰ(ㅍᄃᆞ)	tʰ(ㄷᄃᆞ)			kʰ(ㄱᄃᆞ)	
	마 찰 음	평음		s(ㅅ)				
		경음		sʰ(ㅅᄂᆞ)				
		유기음						h(ㅎ)
	파 찰 음	평음			ʃ(ㅆ)			
		경음			ʃʰ(ㅆᄂᆞ)			
		유기음			ʃʰ(ㅆᄃᆞ)			
공 명 음		비음	m(ㅁ)	n(ㄴ)			ŋ(ㅇ)	
		유음		l(ㄹ)				
		반모음				y(ㅇ), ɥ	w, ɥ	

둘째, 한국어의 단모음체계는 혀의 위치에 따라 전설, 중설, 후설모음으로 나뉘고 혀의 높이에 따라 고모음, 중모음, 저모음으로 나뉘며 본 연구에서는 표 3에서 제시한 한국어의 7 모음체계를 따른다.

표 3 한국어의 7 모음 체계(강옥미 2003)

	전설		중설	후설
	평순	원순	평순	원순
고모음	i(ㅣ)		i(ㅡ)	u(ㅜ)
중모음	E/e(ㅓ)		ə(ㅗ)	o(ㅛ)
저모음			a(ㅏ)	

셋째, 많은 외국어의 경우 하나의 모음을 기준으로 두음과 말음에 자음이 복수로 최대 3 개까지 위치 할 수 있다. 하지만 한국어의 경우 하나의 모음을 기준으로 두음과 말음에 최대 1 개의 자음만 위치 할 수 있다. 한국어 음절에 대한 예시는 표 4 와 같다(강옥미 2003).

표 4 한국어의 음절에 대한 예시(강옥미 2003)

개음절	모음	아/a/, 오/o/, 우/u/, 어/ə/...
	자음+모음	개/ke/, 구/ku/, 테/te/, 모/mo/...
폐음절	모음+자음	악/ak'/, 음/om/, 운/un/, 억/ək'...
	자음+모음+자음	간/kan/, 곡/kok'/, 궁/kuŋ/, 명/məŋ/...

### 1.3 동물의 숲 게임에 나타는 등장인물 이름의 특징

동물의 숲 게임은 주인공 캐릭터가 동물 주민들이 살고 있는 마을에 이사 와서 주민들과 함께 낚시, 곤충채집, 인테리어 등의 유유자적한 일상을 보내는 게임이다. 동물의 숲 시리즈에 등장하는 캐릭터들은 남녀 구별이 있으며 각 성별마다 4 가지 성격으로 세분화 (남성 캐릭터는 운동광, 먹보, 무뚝뚝, 느긋함, 여성 캐릭터는 친절함, 아이돌, 단순활발, 성숙함)가 되어 있다. 동물의 숲에 등장하는 캐릭터의 이름은 일반적인 한국인의 이름과는 차이가 있다. 대부분의 한국인 이름이 두 글자로 구성되어 있고, 각각의 음절에 해당하는 한자가 있고 의미가 있음에 비해, 동물의 숲 캐릭터 이름은 캐릭터를 특징짓는 어휘나 구에서 유래한다. 따라서 한자가 아닌 순수 한국어 어휘가 많고, 두 글자로 제한되지 않고 성격을 묘사하는 형용사나 명사 혹은 구를 최대한 압축한 1~5 음절의 이름으로 구성되어 있다.

동물의 숲 캐릭터 이름 말뭉치는 Kawahara et al. (2018)에서 제시한 포켓몬 데이터와 마찬가지로 캐릭터의 이름이 많기 때문에 충분한 양적 연구가 가능하다. 이름들은 또한 실제 사용하는 한국어 또는 한국어 음성상징어에서 파생된 이름이 적지 않다. 예컨대 그림 1 과 같이 '오골계'라는 실제 사용하는 한국어에서 파생된 '오골'이라는 이름은 닭의 모습을 한 캐릭터이고 '철컹'이라는 한국어의 음성상징어에서 따온 이름인데 이 캐릭터는 로봇의 모습을 한 개구리 캐릭터이다. 한국 인명과는 달리 음성상징어에서 직접 파생된 이름을 통해 우리는 한자 어휘가 아닌 음성상징어의 음운론

적 특징에 대해 엿볼 수 있는 기회를 얻을 수 있으며, 이름이 두 음절로 제한되어 있지 않은 점 또한 한국 인명과는 다른 종류의 데이터를 제공한다고 볼 수 있다.



그림 1 동물의 숲 주민 캐릭터인 오골(1)과 철컹(2)

이러한 이유로 본 연구는 의성어와 의태어로 작성된 이름을 포함하여 총 392 명의 캐릭터 이름을 포함하는 동물의 숲 이름 명단을 언어적 자료로 사용하여 외국어와 다른 한국어의 음운론적 특징이 여성의 이름과 남성의 이름을 결정짓는데 있어 어떠한 영향을 미치는 지를 보고자 한다.

## 2. 방법론

### 2.1. 데이터 수집

본 연구에서 데이터를 수집한 과정과 데이터에서 음운론적 특징을 변수로 생성하는 과정은 그림 2 와 같다.

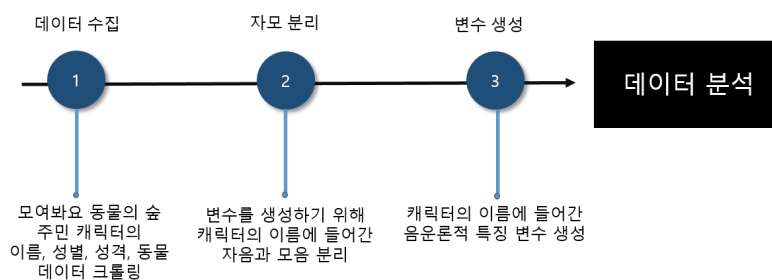


그림 2 동물의 숲 주민 이름 데이터 수집 과정



첫째, 동물의 숲 게임에 대한 공략과 주민목록이 포함된 *모여봐요 동물의 숲 백과 웹사이트*의 동물 주민 리스트에서 파이썬의 BeautifulSoup(Richardson 2007)과 Selenium<sup>1</sup> 패키지를 이용하여 주민 리스트에서 성격을 포함한 캐릭터의 정보가 나타나 있는 일반 주민들의 이름, 성별, 성격, 동물 종류에 대한 데이터를 수집하였다.<sup>2</sup>

그 다음 이름에 들어간 음운론적 특성들을 변수로 생성하였다. 변수는 이름에 포함된 음절 수, 전체 분절음의 개수, 평 파열음(ㅂ, ㅅ, ㅈ), 유기 파열음(ㅍ, ㅌ, ㅋ), 경 파열음(ㅃ, ㅆ, ㅊ), 마찰음(ㅅ, ㅆ, ㅎ), 파찰음(ㅈ, ㅊ, ㅉ), 공명자음(ㄴ, ㄷ, ㄹ, ㅇ)의 개수, 이름이 저해음(파열음, 마찰음, 파찰음), 공명자음(ㄴ, ㄷ, ㄹ, ㅇ), 전설모음(ㅣ, ㅗ), 후설모음(ㅓ, ㅕ)으로 시작하거나 끝나는 경우로 생성했다. 이름의 처음과 끝이 어떤 분절음인지를 변수로 생성한 이유는 대부분의 선행연구들이 이름의 처음과 끝이 어떤 분절음인지를 변수로 만들었기 때문에 본 연구에서도 유의미한 영향을 끼칠 것으로 예상하기 때문이다(Slater and Feinman 1985, Barry and Harper 1995, Perfors 2004, Moorthy et al. 2018). 파열음, 마찰음, 파찰음, 공명음의 개수는 각 변수들마다 전체 분절음으로 나눈 비율을 계산하여 분석에 사용되었다. 모음은 공명자음과 함께 공명음에 속하므로 모음의 개수는 위에서 측정한 공명자음의 개수에 더하여 빈도를 측정했다. 즉, 공명음의 개수는 공명자음의 개수와 모음의 개수를 합친 개수로 측정되었다.

## 2.2. 데이터 분석 방법

### 2.2.1. 카이제곱 검정

본 연구는 이름의 성별을 결정하는데 있어서 영향을 미치는 한국어의 음운론적 특징과 그 정도를 파악하는 것을 목적으로 하고 있다. 연구 목적을 위한 분석 과정으로 본 연구는 우선 이름에 성별에 따른 음운론적 특징에 차이가 있는지를 분석하기 위해 카이제곱 검정을 사용하였다. 카이제곱 검정( $\chi^2$ -test)이란 변수가 명목형일 때 주로 사용하며 두 변수가 독립적인지 검정하는 분석 방법으로 관측빈도와 기대빈도 사이에 유의미한 차이가 있는지를 확인하는 통계량이며 카이 제곱 검정의 수식은 아래와 같이 정의할 수 있다. (e.g., 홍선경 외 2014, 손원. 2020., 임채은 외 2020.).

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

<sup>1</sup> (<https://www.selenium.dev/documentation/>)

<sup>2</sup> 데이터를 수집한 웹사이트는 동물의 숲에 대한 공략이 정리되어 있는 모여봐요 동물의 숲 백과 웹사이트 (<https://animalcrossing.soopoolleaf.com/ko/acnh/animalsearch/?soopoolleaf=all>)를 이용했다.

수식에 들어가는 값들 중 ‘ $O_{ij}$ ’는 관측빈도이고 ‘ $E_{ij}$ ’는 기대빈도이다. 관측 빈도는 실제로 수집된 데이터의 빈도이고 기대빈도는 전체 빈도에 대해 열의 합을 기준으로 보았을 때 각 교차되는 칸에 몇 번의 빈도가 확인될 수 있을지를 예상하는 기댓값을 말한다. 예를 들어 표 5 와 같이 이름의 끝 분절음이 공명음으로 끝나는 지 여부에 따른 성별 간 차이를 보기 위해 교차표를 만들었다고 했을 때 관측빈도는 표 5 에서 제시한 수치와 같다.

표 5 끝 분절음이 공명음으로 끝나는 지 여부에 따른 성별 간 차이

	공명자음으로 끝난다.	공명자음으로 끝나지 않는다.
남성	28	162
여성	52	150

기대빈도는 관측빈도가 속한 행의 크기와 열의 크기를 곱한 다음 전체 데이터의 크기로 나눈 값이다. 예를 들어 공명자음으로 끝나는 남성의 이름의 관측빈도는 약 39 이다.  $((28+52)*(28+162)/392 = 38.77551)$  이처럼 모든 관측빈도에 대한 기대빈도를 구한 후 앞서 언급했던 수식에 대입하면 카이 제곱값이 도출된다.

본 연구에서는 이름의 성별에 영향을 미치는 음운론적 특징으로 음절 수, 전체 분절음 수, 공명음의 개수, 저해음의 개수, 파열음의 개수, 마찰음의 개수, 파찰음의 개수, 이름의 첫 분절음이 저해음, 공명음, 전설모음, 후설모음인 이름, 이름의 끝 분절음이 저해음, 공명음, 전설모음, 후설모음인 이름을 독립변수로 사용하고 성별을 종속 변수로 사용하여 분석을 진행하였다.

## 2.2.2. 로지스틱 회귀 분석

앞서 언급했던 카이제곱 검정에서는 두 변수 사이의 독립성에 대해서만 검정한다. 그래서 독립변수 변수들을 종합하였을 때 어떤 변수가 이름의 성을 설명하는 데에 유효한지 보기 위해 종속변수를 남성과 여성으로 하는 로지스틱 회귀분석 모델을 구축하였다. 로지스틱 회귀는 종속변수가 성별과 같이 1 과 0 으로 표현되는 이항 분포일 때 사용하는 분석 방법으로 독립변수의 유의한 효과는 승산비(odds)를 통해 검증된다. 승산비는 종속변수가 1 일 확률과 0 일 확률사이의 비율로 회귀분석에서는 이 승산비의 정중앙인 0.5 를 기준으로 0.5 보다 크면 1, 작으면 0 으로 판단하여 결과를 도출한다.(e.g., 이창환·정미나 2016, 노경섭 2020, 권선아 외 2020) 승산비는 아래와 같이 정의할 수 있다.

$$\text{odds} = \frac{p}{1-p}$$

연구에서 사용한 로지스틱 회귀분석의 절차는 첫째, 단계적 회귀분석 방법 중 하나인 전진 선택법 (i.e., forward selection)을 사용하여 종속변수인 성별에 영향을 미치는 요인들만을 모형 생성에 활용하였다. 전진선택법이란 가장 종속변수에 영향을 주는 독립변수부터 모형에 하나씩 포함시키면서 종속변수를 가장 잘 설명하는 모형을 만드는 방법이다(노경섭 2020). 본 연구는 회귀 모형에 15 개의 독립변수들(이름에 들어간 음절 수, 전체 분절음 수/저해음, 공명음, 파열음, 마찰음, 파찰음의 비율/이름이 저해음, 공명자음, 전설모음, 후설모음으로 시작하거나 끝나는 지)을 모두 입력한 후 둘째, 회귀 모델이 사용되기에 적합한 지를 알아보기 위해 변수간의 공선성, 잔차의 독립성 등을 확인하였다(Winter 2013). 셋째, 예측 모형 평가를 위해 생성된 회귀 모형식을 실제 테스트 데이터에 대입함으로써 회귀 모형의 성별 예측 정확도를 측정하였다. 위의 모든 분석 절차는 통계 프로그램인 R(R Core Team, 2021)를 사용하여 진행하였다.

### 3. 결과

#### 3.1. 카이제곱 검정 결과

위에서 제시된 독립변수들의 특징을 선행 연구를 참고하여 4 개의 그룹으로 분류하였으며, 4 개의 그룹별로 각각 카이제곱 검증을 진행하였다.

##### 3.1.1. 음절 수와 분절음 수

첫 번째로 각 이름에서 포함하는 전체 자음과 모음의 개수를 나타내는 분절음과 홀로 발음할 수 있는 최소의 소리 단위는 나타내는 음절에 대한 분석 결과, 성별 간에 음절 수는 차이가 있으나, 성별 간의 분절음 수는 차이가 없는 것으로 나타났다.

표 6 이름에 들어간 음절 수와 분절음 수에 따른 성별 간의 차이의 카이제곱 검정 결과

	$\chi^2$	N	df	p
음절 수	14.862	392	4	.004997
분절음 수	2.6666	392	8	.9535

표 7 은 성별에 따른 음절 수와 분절 수의 평균 값을 나타내고 있으며 이를 통해 여성 캐릭터의 음절 수와 분절음 수의 평균이 남성 캐릭터 보다 더 높기 때문에 결과에 차이가 난 것을 확인 할 수 있다.

표 7 성별에 따른 음절 수와 분절음 수

	남성	여성
음절 수	2.50	2.71
전체 분절음 수	5.53	5.58

### 3.1.2. 공명음, 저해음, 파열음, 마찰음, 파찰음의 개수

다음으로 성별간에 공명음의 개수, 파열음의 개수, 마찰음의 개수, 파찰음의 개수, 그리고 파열음, 마찰음, 파찰음의 개수를 합친 값을 나타내는 저해음 사이에 차이가 있는지를 분석하기 위한 카이제곱 검증 결과이다. 분석 결과, 표 8 에서 제시된 것과 같이 95% 신뢰구간인  $\alpha=0.05$  을 기준으로 파열음과 저해음, 공명음은 성별에 따라 차이가 있지만 마찰음, 파찰음에서는 성별 간의 차이가 없는 것을 확인 할 수 있었다.

표 8 이름에 들어간 각 분절음의 개수에 따른 성별 간의 차이의 카이제곱 검증 결과

	$\chi^2$	N	df	p
파열음	27.376	392	4	1.668e-05
저해음	16.951	392	3	.0007233
공명음	15.54	392	7	.02967
마찰음	4.6685	392	2	.09688
파찰음	0.99461	392	2	.6082

그림 3 은 남성의 이름에는 저해음과 파열음의 비율이 높고 여성의 이름에는 공명음의 비율이 더 높게 나타나는 것을 보여준다.

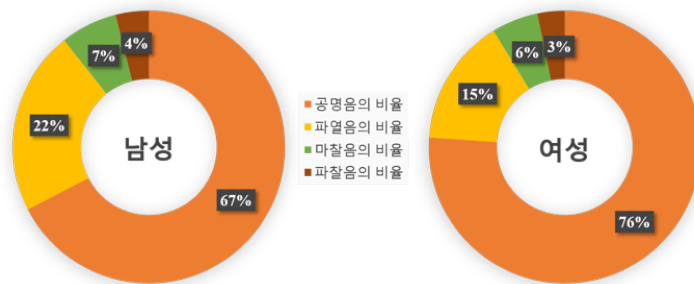


그림 3 남성과 여성 캐릭터 이름에 들어간 분절음의 분포

### 3.1.3. 이름의 첫 분절음

표 9 는, 이름의 첫 분절음으로 사용된 저해음, 공명자음, 전설모음, 후설모음이 성별 간에 차이가 있는지에 대한 카이제곱 검정결과이다. 분석 결과, 첫 분절음이 저해음인 경우 성별 간에 차이가 있다는 것을 확인 할 수 있었으나, 공명자음, 전설모음, 후설모음인 경우에는 95% 신뢰구간 안에서 유의미한 차이가 없다는 것을 발견 할 수 있었다

표 9 이름의 첫 분절음에 따른 성별 간의 차이의 카이제곱 검정 결과

	$\chi^2$	N	df	p
저해음	6.016	392	1	.01418
공명음	3.206	392	1	.07337
전설모음	3.3232	392	1	.06831
후설모음	0.049453	392	1	.824

그림 4 는 이름의 첫 분절음의 분포를 제시하고 있으며, 첫 분절음이 저해음인 남성의 이름이 더 많다는 것을 확인 할 수 있다.

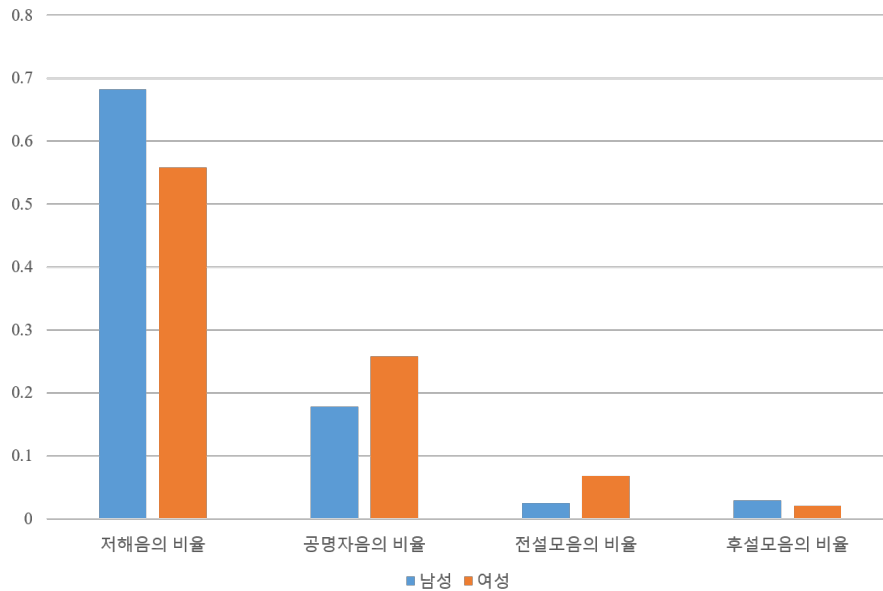


그림 4 남성과 여성 캐릭터 이름에 들어간 첫 분절음의 분포

#### 3.1.4. 이름의 끝 분절음

표 10 은 이름의 끝 분절음으로 사용된 저해음, 공명자음, 전설모음, 후설모음이 성별 간에 차이가 있는지에 대한 카이제곱 검정결과이다. 분석 결과, 이름의 끝 분절음의 분포의 경우, 첫 분절음의 분포와 다르게 저해음이 아닌 공명자음으로 끝날 때 성별에 따라 95% 신뢰구간 안에서 유의미한 차이가 나타났다 하지만 첫 분절음이 저해음, 전설모음, 후설모음일 때는 유의미한 차이가 없는 것을 확인 할 수 있었다

표 10 이름의 끝 분절음에 따른 성별 간의 차이의 카이제곱 검정 결과

	$\chi^2$	N	df	p
저해음	2.4609	392	1	.1167
공명음	6.6392	392	1	.009976
전설모음	0.68847	392	1	.4067
후설모음	1.5075	392	1	.2195

그림 5 는 이름의 끝 분절음의 분포를 제시하고 있으며 끝 분절음이 공명음인 남성의 이름이 여성에 비해 더 많다는 것을 확인 할 수 있다.

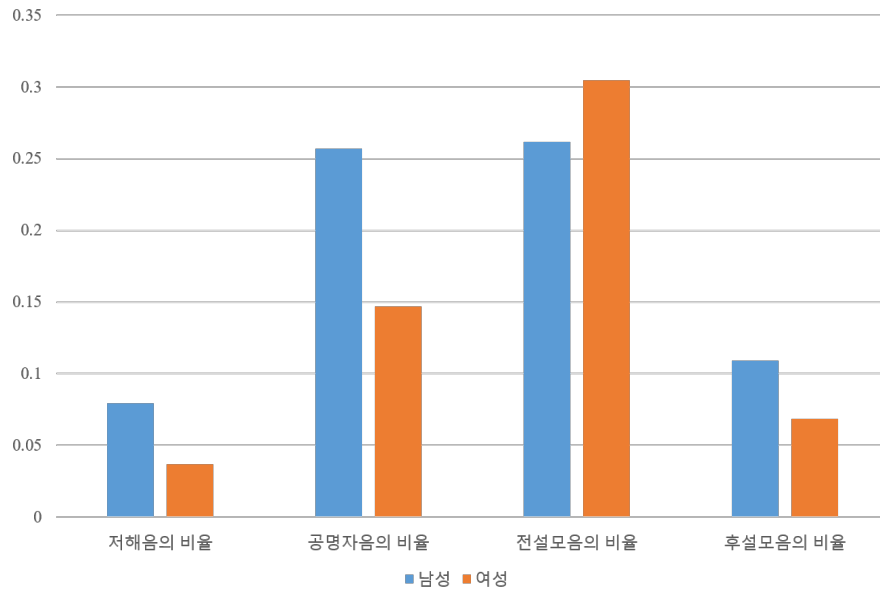


그림 5 남성과 여성 캐릭터 이름에 들어간 끝 분절음의 분포

### 3.2. 로지스틱 회귀분석 결과

앞서 기술된 내용과 같이 본 연구에서는 총 세 단계에 걸쳐 로지스틱 회귀분석을 진행하였다. 첫째로 전진 선택법 (i.e., forward stepwise)을 수행한 결과, 종속변수에 영향을 주는 독립변수는 저해음의 비율, 저해음, 전설모음으로 시작하는 이름, 공명자음, 모음으로 끝나는 이름이 유의미한 변수로서 회귀 모형 생성을 위한 분석에 사용되었다.

둘째로, 변수간의 공선성, 잔차의 독립성 등을 확인한 결과, 공선성의 경우, 독립 변수들 사이에 공선성이 존재하지 않는다는 조건을 만족하였다.

또한 잔차의 독립성을 알아보기 위해 더빈-왓슨 검정을 시행한 결과, 더빈-왓슨 값이 2 에 가까우므로 잔차의 독립성을 만족하였다 ( $DW = 1.8841$ ,  $p\text{-value} = 0.1058$ ). 마지막으로 전진 선택법 (i.e., forward stepwise)을 통해 선택된 변수들을 활용하여 로지스틱 회귀 분석을 수행한 결과는 제시된 표 8 와 같다.

표 11 로지스틱 회귀분석 결과

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	1.5973	0.3067	5.209	1.90e-07 ***
저해음의 비율	-5.6499	0.9664	-5.846	5.03e-09 ***
공명자음으로 끝나는 이름	-1.3105	0.3120	-4.201	2.66e-05 ***
모음으로 끝나는 이름	-0.5059	0.2474	-2.045	0.0409 *
저해음으로 시작하는 이름	0.5569	0.2825	1.971	0.0487 *
전설모음으로 시작하는 이름	1.0253	0.5873	1.746	0.0808 .

로지스틱에서 종속 변수로 사용된 성별의 경우 여성의 경우 1, 남성의 경우 0 으로 정하여 분석을 진행하였다. 회귀분석 결과, 이름의 성별에 영향을 주는 음운론적 특징은 표 11 에서 제시된 바와 같이 캐릭터의 이름에 저해음의 비율, 분절음이 공명자음 및 모음으로 끝나는 경우, 분절음이 저해음으로 끝나는 경우가 신뢰구간 95%로 이름의 성별에 영향을 주는 변수로 확인되었다. 분석 결과의 회귀식은 다음과 같다.

$$y = 1.5973(\text{상수항}) - 5.6499x_1(\text{저해음의 비율}) \\ - 1.3105x_2(\text{공명자음으로 끝나는 이름}) - 0.5059x_3(\text{모음으로 끝나는 이름}) \\ + 0.5569x_4(\text{저해음으로 시작하는 이름})$$

회귀식을 통한 이름의 성별과 음운론적 특징 간의 관계를 해석해보면 회귀식의 상수항이 유의하고 1.5973 으로 양수이므로 동물의 숲에서 사용된 이름의 경우, 기본적으로 여성의 이름일 확률이 높으며, 저해음으로 이름이 시작할 경우 여성일 확률이 더 높아진다. 하지만 저해음의 비율이 높고, 공명자음으로 끝나는 이름, 모음으로 끝나는 이름이면 남성의 이름일 확률이 높아진다.

생성된 로지스틱 회귀 모델의 평가를 위해서 동물의 숲 이름 데이터의 10%를 추출하여 실제 이름의 성별 정보와 모델을 통해 예측된 이름의 성별이 일치하는지에 대한 정확도를 측정하였다(e.g., 김규완 외 2017, 홍종선, 오세현 2021, 김태영 외 2021). 표준화된 로지스틱 회귀 모델의 정확도 측정을 위해 앞서 언급된 모델 평가 과정을 100 번 수행 하였으며, 수행된 결과의 정밀도, 재현율, F 점수, 그리고 정확도에 대한 평균을 계산한 값은 표 12 과 같다.

표 12 로지스틱 회귀 모델 평가 결과

정밀도	재현율	F1-score	정확도
0.6942365	0.6587479	0.6760268	0.6585



분석 결과, 음운론적 특징을 변수로 사용하여 이름의 성별을 예측하는 로지스틱 회귀 모델의 정밀도는 약 69%, 재현율은 약 66%, F 점수는 약 68%, 그리고 정확도는 약 66%로 측정되었다. 선행연구에서 저해음과 공명자음이 데이터의 성별을 예측하는데 유의미한 변수였다는 점과 본 연구의 회귀모형식에서 공명자음과 저해음이 선행연구의 결과와 일치하는 점을 감안할 때 저해음과 공명자음이 이름의 성별에 미치는 유의미한 영향으로 인해 모델의 성능이 높게 나왔다는 것으로 해석 할 수 있다.

#### 4. 논의 및 결론

본 연구는 한국어에 존재하는 음운론적 특징이 여성의 이름과 남성의 이름을 결정짓는데 있어 어떠한 영향을 미치는 지에 대해 연구했다. 연구의 결과는 다음과 같다. 우선 이름에서 나타나는 음운론적 특징이 성별에 따라 차이가 나는지를 확인하기 위한 카이제곱 검정 결과, 여성 이름의 음절 수가 남성 이름의 음절 수보다 더 길고, 공명음의 개수가 남성 이름보다 여성 이름에서 더 높은 것을 확인하였다. 반면에 저해음과 파열음의 개수는 남성 이름인 경우 여성 이름보다 더 높은 것을 확인하였다. 이러한 결과들은 Slater and Feinman(1985), Perfors(2004)의 분석 결과 중 여성 캐릭터의 이름이 음절 수와 공명음의 개수가 더 많고 남성의 이름일수록 저해음과 파열음의 개수가 더 많다는 결과와 일치했다. 즉, 공명음의 개수와 음절 수가 많은 이름일수록 이름에 여성의 이미지가 강해지고 저해음 또는 파열음의 개수가 많은 이름일수록 남성의 이미지가 강해진다는 것을 알 수 있다. 그리고 남성의 이름이 저해음으로 시작하거나 공명음으로 끝나는 경우가 많았는데 이 결과는 Slater and Feinman(1985), Barry and Harper(1995)의 분석 결과에서 여성의 이름이 공명음으로 끝나는 경우가 많다는 결과와 상반되는 결과이며 Moorthy et al.(2018)의 연구결과에서 공명음으로 끝나는 이름이면 남성의 이름이라고 예측한 부분과는 일치하는 결과이다. Slater and Feinman(1985), Barry and Harper(1995)의 연구에서 사용한 이름 데이터가 1960-90 년의 데이터이고, Moorthy et al.(2018)의 연구에서 사용된 이름 데이터의 대부분이 2000 년대 이후의 이름이라는 점을 감안해보면 본 연구의 분석 결과는 2000 년대 이후 이름의 트렌드를 반영한다고 볼 수 있다. 즉, 2000 년대 전에는 공명음으로 끝나는 여성의 이름이 많았지만 2000 년대 이후가 되면서 공명음으로 끝나는 남성의 이름이 더 많아졌다는 것을 알 수 있다.

다음으로 로지스틱 회귀를 사용하여 이름의 성별에 영향을 미치는 음운론적 특징에 대한 분석 결과, 이름에서 저해음의 비율이 높고 이름이 공명자음 또는 모음으로 끝나는 이름일 경우 남성의 이름일 확률이 높으며 이름이 저해음으로 시작하는 이름일 경우 여성의 이름일 확률이 높을 것으로 예측되었다.

본 연구의 결과에도 불구하고 본 연구가 가지는 한계점은 다음과 같다. 첫째, 본 연구에서는 성별과 음운론적 특징 간의 관계를 확인하기 위해 로지스틱 회귀 만을 사용하였다. 성별처럼 종속 변수가 이항 변수인 경우에는 로지스틱 외에도 의사 결정 나무 분석 혹은 랜덤 포레스트 등의 다른 알고리즘도 사용이 가능하다. 따라서 본 연구는 향후 언급된 모델들을 적용하여 연구를 확장하고자 한다. 둘째, 본 연구는 분석을 위해 2021년 7월 기준의 동물의 숲 캐릭터 이름 데이터를 언어적 자료원으로 사용하였다. 분석 결과 및 로지스틱 회귀를 통해 생성된 모델의 성능을 일반화 하기 위해서는 더 많은 데이터와 실제 이름 데이터를 분석에 사용 할 필요가 있다. 그 이유는 실제 사람의 이름이 아닌 게임 캐릭터의 이름을 가지고 분석을 시도했기 때문이다. 물론 서론에서 언급했듯이 실제 한국어에서 차용된 이름이나 음성상징어로 된 이름들이 적지 않게 등장하지만 이들이 한국어 이름의 음운론적 특징 전체를 반영했다고 보긴 어렵다.

그러나 한국어의 음운론적 특징과 외국어의 음운론적 특징이 서로 다른 상황에서 한국어의 이름에 나타나는 음운론적 특징과 이름의 성별에 대한 연구 결과가 미미한 것을 감안 할 때, 연구의 간극을 채웠다는 면에서 의의가 있다고 본다. 본 연구에서 한국어의 사람 이름에 나타나는 음운론적 분포와 성별 간의 관계를 통계적으로 분석하고 탐구하여 외국어의 이름에 나타나는 음성상징어의 특징과 일맥상통하는 결과를 얻은 것은 음성상징어의 보편성을 어느 정도 반영한다고 할 수 있다.

### 참고문헌

- 강옥미. 2003. *한국어 음운론*. 태학사.
- 권선아·김명진·서희정·김민정. 2020. 원격고등교육에서 성인학습자의 중도탈락에 관한 로지스틱 회귀분석. *평생학습사회*, 16, 149-169.
- 김규완·신현주·김선진·문경득·이현아. 2017. 문단 간 유사도와 로지스틱 회귀 분류를 통한 기사 내 부적합 문단 검출 시스템. *한국정보과학회 학술 발표논문집*, 1873-1875.
- 김재원. 2003. 한국어와 영어의 유음 음성상징. *영어영문학* 10, 21-36.
- 김선희. 2019. 한국어 음성상징어의 자음 분포 특성. *음성음운형태론연구*, 25, 387-414.
- 김태영·김태성·이미영. 2021. 기계학습을 통한 잉여현금흐름 예측 : 서포트 벡터머신과 로지스틱 회귀 모형을 중심으로. *전산회계연구*, 19, 25-42.
- 노경섭. 2020. *제대로 알고 쓰는 논문 통계분석 : SPSS & AMOS*. 한빛아카데미.
- 손원. 2020. 불균형 텍스트 데이터에서 카이제곱 통계량을 이용한 변수 선택의 편향성. *한국데이터 정보과학회지*, 31, 807-821.
- 이영석. 1994. 한국어 상징음의 모음조화. *언어학* 16, 121-138.

- 이창환·정미나. 2016. 속성값 기반의 정규화된 로지스틱 회귀분석 모델. *정보과학회논문지*, 43, 1270-1274.
- 임채은·박원주·장광호·백명선. 2020. 범죄통계데이터를 활용한 카이제곱 검증 기반 범죄특징분석. *한국통신학회 학술대회논문집*, 911-912.
- 홍선경·이상훈·문양세. 2014. 카이제곱 검정 통계량과 유관계수의 안전한 다자간 계산. *한국정보과학회 학술발표논문집*, 225-227.
- 홍종선·오세현. 2021. 다양한 정확도 측도들에 기반한 진단오즈비의 특성. *한국데이터정보과학 회지*, 32, 1281-1294.
- HERBERT BARRY and AYLENE S HARPER. 1995. Increased choice of female phonetic attributes in first names. *Sex Roles*, 32, 809-819.
- Hamano S. 1986. *The Sound-Symbolic System of Japanese*; doctoral dissertation, University of Florida
- KAWAHARA, S., GODOY, M. C., & KUMAGAI, G. 2021. English Speakers Can Infer Pokémon Types Based on Sound Symbolism. *Frontiers in psychology*, 12, 648-948.
- KAWAHARA, S. and KUMAGAI, G. 2021. What voiced obstruents symbolically represent in Japanese: evidence from the Pokémon universe. *Journal of Japanese Linguistics*, 37, 3-24.
- KAWAHARA, SHIGETO, ATSUSHI NOTO and GAKUJI KUMAGAI. 2018. Sound symbolic patterns in Pokémon names. *Phonetica* 75, 219-244.
- MOORTHY S, POGACAR R, KHAN S and XU Y. 2018. Is Nike female? Exploring the role of sound symbolism in predicting brand name gender. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 1128-1132.
- PERFORS A. 2004. What's in a name? The effect of sound symbolism on perception of facial attractiveness. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 26.
- SAPIR E. 1929. A study in phonetic symbolism. *Journal of Experimental Psychology* 12, 225-239.
- SLATER A S and FEINMAN S. 1985. Gender and the phonology of North American first names. *Sex Roles* 13, 429-440.
- SAUSSURE, F. 1966. *Course in General Linguistics*. McGraw-Hill Book Co.
- R CORE TEAM. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing (Version 4.1.2). [Computer program]. <https://www.R-project.org/>
- RICHARDSON, L. 2007. Beautiful soup documentation. *April*. [Computer program]. <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
- WINTER B. 2013. Linear models and linear mixed effects models in R with linguistic applications. *CoRR*.