

음운 변동 실현 오류의 예측 인자 분석

안성민
전주대학교
smsmsm@jj.ac.kr

Analysis of Predictors of Phonological Variation Realization

An, Sung-min
Jeonju University

요 약

본 연구에서는 음운 변동에서 나타나는 오류가 어떤 변수에 영향을 받는지 확인하여 음운 변동 연구 및 교육의 기초 자료를 제공하고자 하는 데에 목적이 있다. 이를 위해 유음화 발음 데이터를 이용하여 성별, 유음화의 방향, 품사, 단어의 빈도, 단어의 음절수와 유음화의 발음 적격 유무를 변수로 설정하였다. 유음화 적격률에 영향을 줄 수 있는 독립변수를 찾기 위해 카이제곱 검정과 다중공선성의 팽창계수를 먼저 확인하였다. 이후 다중 로지스틱 회귀분석과 오즈비를 통해 유의한 예측인자를 검토하였다. 그 결과 5개의 독립 변수 중 성별과 유음화의 방향, 품사가 결과를 오류에 영향을 주는 주요한 인자가 되는 것을 확인할 수 있었다.

주제어: 음운변동, 실현 오류, 다중로지스틱회귀분석, 예측인자

1. 서론

본 연구는 음운 변동에서 나타나는 오류가 모국어의 간섭이라는 요인 이외에 다른 어떤 변수에 영향을 받는지 밝혀 음운 변동 연구 및 교육의 기초 자료를 제공하는 데에 목적을 둔다. 이를 위해 유음화 발음 데이터를 이용하여 각각의 독립 변수와 종속 변수 간의 영향 관계를 분석하였다.

언어를 학습하는 데에 있어 음운 변동의 학습은 그 언어에서 해당 음운 변동을 매개인자로 가지고 있는지의 여부에 따라 난이도가 달라지게 된다.[1] 만약 모어에서 매개인지로 갖고 있지 않은 음운 변동이 목표어에서 나타날 경우 학습자는 더 많은 어려움을 겪게 되며 오류를 생산하게 된다. 이때 발음의 오류는 모어에 존재하지 않는 매개인자라는 요인 외에도 다양한 외부 요인에 의해 발생하게 된다. 이러한 오류의 예측 인자를 파악하기 위해서는 여러 개의 변수를 처리할 수 있는 분석 도구가 필요하다. 로지스틱 회귀 분석은 다변수 처리 통계 기법의 하나로 종속 변수와 독립 변수들 간의 상관 관계를 파악하고 예측할 수 있으며 변수의 유형에 독립적으로 사용할 수 있다는 장점이 있다.[2] 따라서 본 연구에서는 음운 변동 학습에서 나타나는 오류의 원인을 밝히기 위해 다중 로지스틱 회귀 분석 기법을 이용한다. 이를 통해 오류 예측 인자를 밝혀 향후 음운 변동 관련 연구 및 교육을 위한 기초 자료를 제공하고자 한다.

2. 음운 변동 오류율 예측 모델링

발음 오류를 예측하기 위해 음운 변동의 하나의 유음화를 선택하였다. 유음화의 데이터는 한국어 학습자를 대상으로 하였으며 유음화 조건을 갖춘 단어와 문장을 꾸려 연구자가 직접 수집하였다. 실험에 참여한 피험자는 우즈베키스탄 국적의 학습자로 남녀 각 4명씩 총 8명

이었다. 이들은 유음화에 대한 학습이 이루어진 중급 수준의 학습자였으며 총 3회에 걸쳐 발음 녹음을 실시하였다. 실험을 통해 피험자별로 각 111개의 토큰을 확보하였으며 총 888개의 토큰을 분석에 사용하였다. 전체 유음화 실현 비율을 34.4%였으며 이러한 결과에 대해 각각의 변수는 얼마나 상관관계를 갖는지, 모든 독립변수가 종속변수에 영향을 미치는지, 예측 인자로서의 변수는 무엇이 되는지를 살피고자 한다.

변수는 아래 <표 1>과 같이 성별, 유음화의 방향, 품사, 단어의 빈도, 단어의 음절수, 유음화의 적격 유무로 설정하였다. 연구에서는 유음화의 적격 유무가 종속변수가 되며 나머지 다섯 개의 변수가 독립변수가 된다. 연구에 사용된 변수는 모두 이산형 변수로 성별은 남녀, 유음화의 방향은 순행과 역행, 품사는 체언과 용언, 단어의 빈도는 고빈도와 저빈도, 단어의 음절수는 2음절과 3음절, 유음화의 적격 유무는 적격과 부적격의 값을 갖게 된다.

<표 1> 변수 설정

변수	값	
	성별	남
유음화의 방향	순행	역행
품사	체언	용언
단어의 빈도	고빈도	저빈도
단어의 음절수	2음절	3음절
유음화의 적격 유무	적격	부적격

각각의 변수가 유음화의 적격 유무와 상관 관계를 갖는지 알아보기 위해 카이제곱 검정을 사용하였으며 검정 결과는 다음과 같다.[3]

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$$

$$H_1 : \beta_j \neq 0 (j = 1,2,3)$$

$$\text{검정통계량 } \chi^2 = \sum (\text{관측값} - \text{기댓값})^2 / \text{기댓값}$$

$$\text{기각역 } \chi^2_0 > \chi^2_{1 - \alpha(r - 1)(c - 1)}$$

각각의 변수에 대한 검정통계량 기각역은 <표 2>에 제시하였다.

<표 2> 검정통계량 및 기각역

변수	검정통계량	기각역
성별	109.8044	1.433135e-24
유음화의 방향	42.70342	5.334222e-10
품사	52.99061	3.113406e-12
단어의 빈도	0.03002965	0.9850973
단어의 음절수	0.9027189	0.6367619

분석 결과 단어의 빈도와 단어의 음절수는 기각역에 포함되어 H_1 을 채택하였고 나머지 변수는 H_0 를 채택하였다.

다음으로 각 요인 간의 상관관계 여부를 확인하기 위해 다중공선성(Multicollinearity)를 확인하였다. 이는 상관관계가 높은 독립변수가 회귀모델에 포함될 경우 유의성 검정값이 크게 떨어져 검정을 신뢰할 수 없게 되기 때문에 반드시 거쳐야 한다. 다중공선성은 분산팽창계수(VIF: Variance Inflation Factor)를 이용하였으며 다음과 같다.

$$VIF_j = 1 / (1 - R^2_j), j = 1, 2, \dots, p - 1$$

각각의 변수에 대한 VIF값은 <표 3>에 제시하였다.

<표 3> 분산팽창계수

변수	VIF
성별	1.010776
유음화의 방향	1.066696
품사	1.032323
단어의 빈도	1.025898
단어의 음절수	1.081534

다중공선성 결과 다섯 개의 독립변수 사이에는 다중공선성이 존재하지 않는다는 것을 확인할 수 있었다.

3. 다중 로지스틱 회귀분석

로지스틱 회귀는 종속 변수가 범주형일 때 사용할 수 있는 회귀 분석법이다. 독립변수가 X이고 종속변수가 Y일 때 선형 회귀모형을 로짓변환하여 설정할 수 있다. 종속변수 Y_i 가 이산형 범주일 때 기댓값과 로지스틱 반응함수는 다음과 같다.

$$E(\epsilon_i) = 0, E(Y_i) = \beta_0 + \beta_1 X_i$$

$$E(Y_i) = \exp(\beta_0 + \beta_1 X_i) / (1 + \exp(\beta_0 + \beta_1 X_i))$$

이를 바탕으로 다중 로지스틱 회귀분석을 진행하였고 그 결과는 <표 4>와 같다.

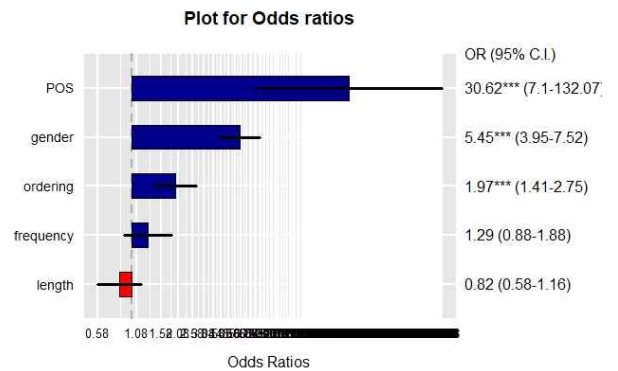
<표 4> 다중 로지스틱 회귀분석 결과

변수	Estimate	Std. Error	z-value	Pr(> z)
성별	1.7913	0.1715	10.447	< 2e-16 ***
유음화의 방향	1.3250	0.2098	6.315	2.71e-10 ***
품사	3.0147	0.7514	4.012	6.02e-05 ***
단어의 빈도	0.4486	0.1992	2.252	0.0243 *
단어의 음절수	0.3913	0.2087	1.875	0.0608 .

각각의 요인이 오류율에 미치는 영향 관계를 살피기 위해 오즈비를 구하였다.

$$\text{오즈비(Odds Ratio)} = \text{Odds}_1 / \text{Odds}_2$$

이를 직관적으로 확인하기 위해 오즈비에 대한 플롯 그림을 아래 <그림 1>을 제시하였다.



[그림 1] 오즈비 플롯

따라서 음운 변동에 대한 예측인자로는 성별과 품사, 유음화의 방향이 주요한 예측인자라는 것을 확인할 수 있다. 품사(POS)는 용언일 때보다 체언일 때에 유음화의 적격률이 상승함을 확인할 수 있으며 성별은 남성일 때보다 여성일 때 적격형이 많이 나타나는 것을 알 수 있었다. 또한 유음화의 방향은 역행일 때가 순행일 때보다 적격 확률이 상승한다는 것을 알 수 있었다. 단어의 빈도는 고빈도와 저빈도가 차이가 없었으며 음절의 수도 2음절과 3음절이 유의한 차이가 없었다.

4. 결론

이 연구는 음운 변동 실현 오류에서 주요하게 작용하는 변수가 무엇인지를 확인하고자 하였다. 이를 위해 학습자의 발음 데이터를 수집하여 적격 유무를 확인하고 적격률에 영향을 미치는 예측인자를 분석하였다. 먼저 카이제곱 검정을 통해 통계적 유의성을 파악하였고 다중 공선성 분산팽창계수로 독립변수 간의 상관관계를 확인하였다. 이후 다중 로지스틱 회귀분석과 오즈비 분석을 통하여 음운 변동 발음 적격 유무에 영향을 많이 미치는 예측 인자를 찾아내었다. 음운 변동 중 유음화 발음에 영향을 주는 예측 인자는 성별과 유음화의 방향, 품사 정보가 주요한 영향을 주는 것으로 확인할 수 있었다. 본 연구는 음운 변동에 영향을 줄 수 있는 다양한 변수들의 영향 관계를 살펴 기초적인 연구 및 교육 자료를 제시하였다는 데에 그 의의를 찾을 수 있다.

참고문헌

- [1] 허용, 한국어교육에서의 대조언어학과 보편문법의 필요성 연구, 이중언어학 36, 1-24, 2008.
- [2] Atkinson, P. M., Massari, R., Generalised Linear Modelling of Susceptibility to Landsliding in the Central Apennines, Italy. Computers & Geosciences, 24-4, 373-385, 1998.
- [3] K. N. Neter, Applied linear regression models, McGraw-Hill, New Yourk, USA, 2008.