

확률적 모델 기반 중세한국어 유성마찰음 /△/의 음운론적 대립에 대한 연구*

박선우
(계명대학교)

Park, Sunwoo. 2017. Voiced fricative /z/ in Middle Korean revisited from the perspective of Probabilistic Phonological Relationship Model. *Studies in Phonetics, Phonology and Morphology* 23.1. 27-54. This study investigates intermediate phonological relationships between /s/ (ㄱ) and /z/ (ㄷ) in Middle Korean from the 15th century to the 16th century. In Middle Korean, the phonological contrast between /s/ and /z/ was complex. In word-medial intervocalic position, /z/ and /s/ are considered to be contrastive. However, in word-initial or word-final position, they are allophonic, so we can predict which of the two segments will occur. I propose that putting /z/ and /s/ into a gradient and probabilistic model provides insight into the phonological relationship between them. The present paper examines the frequencies of /z/, /s/ and other consonants from the 15th to the 16th century to analyze a historical Korean corpus from the perspective of the Probabilistic Phonological Relationship Model (PPRM), measuring the quantitative analysis index of intermediate phonological relationships. Comparing the entropies between /s/ and other consonants with the entropies between /z/ and other consonants, a continuum defined in terms of entropy reveals that /z/ in Middle Korean was more allophonic than phonemic, and it was lost or merged into /s/ in the 16th century. (Keimyung University)

Keywords: complementary distribution, Information Theory, Middle Korean, minimal pair, Probabilistic Phonological Relationship Model, voiced alveolar fricative

1. 머리말

15세기 중세한국어 문헌에서 관찰되는 반치음(半齒音) /△/에 대해서는 그동안 많은 연구 성과가 축적되었다. /△/에 대한 지배적인 견해는 무성 치경

* 본 연구는 2017년 2월 24일 ‘2017년 동남어문학회 춘계 전국학술대회’에서 발표된 논문을 수정하고 보완한 것입니다. 논문의 내용에 대하여 조언해 주신 익명의 심사위원들께 깊이 감사드립니다. 본 연구는 2015년도 계명대학교 비사연구기금으로 이루어졌습니다.

마찰음 /ㅅ/와 대립되는 음소로서, 그 음가는 유성 치경마찰음 [z]라는 것이다. 이기문(1972)의 평가에 의하면 이미 이승녕(1956)에서 /△/가 중세한국어의 독립적 음소라는 견해가 확립되었다. 그러나 유성마찰음 /△/는 중세한국어의 자음체계 안에서 독특한 특성을 갖는다. ‘유기음-경음-평음’의 대립 체계를 갖는 다른 장애음들과 달리 /△/와 /ㅅ/는 유무성 자질에 의한 대립을 형성하며, 유성마찰음 /△/는 음운론적 제약이 많아서 제한적인 환경에만 분포하였다¹.

(1) 15세기 /△/의 분포 (이기문 1972: 30-31)

- a. V __ V: 으스, 너△
- b. j __ V: 새슴
- c. *r __ V: 두셔(<*돌셔), 프셔리(<*플셔리)
- d. n __ V: 한슴, 한삼, 손소
- e. m __ V: 몸소
- f. V __ ɦ: 조애, 것위
- g. V __ β: 옷브-

분절음의 음운론적 지위를 ‘음소’와 ‘변이음’으로 구분하는 이분법적 관점에서는 /△/와 같이 독특한 특성과 제한적인 분포를 갖는 음소 혹은 변이음을 분석하기가 쉽지 않다. 이분법적 관점에서는 두 음소를 음소 대립 혹은 변이음의 관계로 해석할 수밖에 없는데, 예를 들어 /△/와 /ㅅ/의 음운론적 관계와 /ㅎ/와 /ㅅ/ 혹은 /ㄷ/와 /ㅅ/의 관계에 차이가 있다는 점을 가시적으로 보여주기 어렵다.

/△/와 같이 음소와 변이음 사이에서 다른 음소들과 애매한 대립관계를 갖는 분절음들은 다른 언어에서도 관찰된다. 예를 들어 러시아어의 모음

¹ 중세한국어 시기에는 /△/(z) 이외에도 양순음 /병/(β)와 성문음 /ㅇ/(ɦ)가 유성마찰음으로 자음체계 내에 존재하였으며, /병/와 /ㅇ/ 역시 /△/와 마찬가지로 제한적인 환경에만 분포하였다. 따라서 이들 유성마찰음들을 통합적으로 분석하는 것이 바람직하겠으나, /병/와 /ㅇ/는 표기법의 문제로 인하여 현실적으로 /△/와 함께 분석하기 어렵다. /ㅇ/는 한글철자법의 특성상 음절두음의 음가 없는 ‘ㅇ’과 구분이 되지 않으므로 빈도를 측정하기가 쉽지 않으며, /병/는 능엄경언해(楞嚴經諺解 1461)까지만 관찰되므로 표본이 매우 적은 편이다. 특히 /병/는 몇몇 문헌에서 표기의 원칙이 달라서 본 연구와 같은 거시적 분석의 대상으로 적합하지 않다. 15세기 표기법과 /병/의 관계에 대해서는 조규태(1998)을 참조하기 바란다.

[y]에 대해서는 [i]의 변이음이라는 주장과 다른 모든 모음들과 대립되는 별개의 음소라는 주장이 맞서고 있으며(변군혁·홍성훈·박선우 2013), 일본어의 마찰음 [s]와 [ɕ]는 비전설모음 앞에서만 대립되고 전설모음 [i]와 [e] 앞에서는 [ɕ]로 중화된다(Hall 2009: 176).

전통적인 음소론의 관점으로는 이중적인 모습을 보이는 분절음의 대립을 설명하기 어려우므로 음소와 변이음 사이의 애매한 관계를 설명하기 위해서는 음운론적 대립을 양적으로 분석할 수 있다는 인식의 전환이 필요하다(Mitterer et al. 2016). 본 연구에서는 분절음의 대립 관계를 음소와 변이음으로 구분하는 이분법적 분석의 한계를 극복하기 위하여 확률적 모델(Hall 2009)을 통하여 /△/와 /ㅅ/의 관계를 분석하고자 한다. 만약 음운 대립의 최댓값을 ‘1.00’, 최솟값을 ‘0.00’으로 설정하고 분절음 사이의 대립 관계를 그 사이의 값으로 수량화할 수 있다면 위에서 살펴본 애매한 대립을 양적으로 표현할 수 있다.

예를 들어 어떠한 분절음들 사이의 대립이 1.00에 가까운 값을 갖는다면 두 분절음은 대립되는 별개의 음소에 가깝다고 볼 수 있다. 반면 두 분절음의 대립이 0.00에 가까운 값을 갖는다면 두 분절음은 상보적 분포를 갖는 변이음에 가깝다고 해석할 수 있다. 만약 0.50에 가까운 중간 값을 갖는다면 음소나 변이음 가운데 어느 한 범주로 보기 어려운 애매한 대립 관계를 갖는다. 본 연구에서는 확률적 모델을 적용하기 위하여 /△/가 포함된 15세기와 16세기 문헌의 역사자료 말뭉치를 가공하여 음소의 출현 확률을 측정하고, 확률을 통하여 얻은 엔트로피(entropy)를 바탕으로 자음들 사이의 대립 관계를 분석하겠다. 분석 결과를 통하여 중세한국어의 /△/와 /ㅅ/가 음소로서 대립되는 관계인지 공시적인 변이음인지 논의하겠다.

2. 중세한국어 시기 /ㅅ/과 /△/의 분포

중세한국어 자음체계에 대한 선행 연구들을 살펴보면 /△/의 음가가 [ʒ]라는 점에는 이견이 없다. /△/는 16세기 말의 문헌까지 관찰되지만 16세기 중엽을 전후하여 소실되었다(김무림 2004). 현대한국어와 마찬가지로 중세한국어의 평장애음 /ㄱ/, /ㄷ/, /ㄴ/, /ㄷ͡/, /ㅈ/가 모두 유성음 사이의 환경에서 유성음화를 겪었을 것으로 예상되지만 이러한 유성음들은 모두 변이음이므로 훈민정음의 자음 체계에는 반영되지 않았다. 반면 /ㅅ/과 /△/의 경우, ‘프서리~프서리, 한숨~한숨’과 같이 일부 유성음화의 사례가 관찰되지만, 일반적으로는 유성음 사이의 환경에서도 /ㅅ/의 유성음화가 일어나지 않으므로 기존의 연구에서는 /△/를 /ㅅ/의 공시적 변이음이 아니라 독립된 음소

로 분석하는 관점이 지배적이었다(이기문 1972, 이기문 1998).

(1)의 분포를 보면 어두에서는 /△/의 분포가 제한되며 어말에서는 /△/와 /ㅅ/의 대립이 /ㅅ/로 중화되지만, 어중의 환경에서는 /ㅅ/와 /△/의 분포가 중복되므로 두 분절음이 상보적으로 분포되는 변이음들이라고 보기는 어렵다.

(2) 중세한국어 /ㅅ/와 /△/의 분포 환경과 실현 양상 (박창원 1996: 76)

환경 음소	C__V	V__C	V__N	N__V	V__F	V__V
/ㅅ/	[s]	[s]	[s]	[s]		[s]
/△/	[s]	[s]	[s], [z]	[s], [z]	[z]	[z]

(C: 비음과 유성음을 제외한 자음, V: 모음, N: 비음, F: 유성자음/마찰음)

박창원(1996: 74-85)에서는 /ㅅ/와 /△/의 분포 환경을 위와 같이 분석하였는데, ‘V__N, N__V’와, ‘V__V’에서는 /ㅅ/와 /△/의 분포가 중복되어 상보적으로 분포되지 않는다. 기저형이 /ㅅ/인 경우에는 어떤 경우에도 [s]로만 실현되고, 기저형이 /△/인 경우에는 ‘V__N’와 ‘N__V’에서는 [s]나 [z]로, ‘V__F’와 ‘V__V’에서는 [z]로만 실현된다.

/ㅅ/와 /△/의 분포를 비교해 보면, 두 음소의 분포는 세 가지 환경에서 다르게 나타난다. 첫째, 무성자음에 접하는 ‘C__V, V__C’에서는 /ㅅ/와 /△/의 음운론적 대립이 중화되어 모두 [s]로 실현된다. 둘째, 유성음 사이 ‘V__N, N__V’에서는 /△/이 불규칙적으로 [s]나 [z]로 실현된다. 이러한 현상은 공시적 변이음 규칙으로 설명하기 어려우며, 음운론적으로 대립되는 두 음소가 소위 ‘어휘적 환산’이나 ‘음운의 통시적 변화’를 겪는 과정으로 해석할 수 있다. 셋째, 유성음 사이 ‘V__F, V__V’에서는 /ㅅ/의 분포가 제한되거나 /ㅅ/와 /△/가 상보적으로 분포한다.

요약하면 /ㅅ/와 /△/의 분포는 두 음소의 복합적인 관계를 보여준다. 어떤 환경에서는 대립이 중화되고, 어떤 환경에서는 불완전하게 대립되며, 어떤 환경에서는 상보적으로 분포한다. 따라서 공시적인 변이음의 관계와 음운론적으로 대립되는 음소의 관계를 동시에 내포하고 있다. 사실 대부분의 자음들은 음운론적, 음성학적 환경에 따라 그 분포가 어느 정도는 제한되므로 /△/와 /ㅅ/뿐만 아니라 다른 자음들의 대립에서도 이러한 복합성이 나타날 수 있다. 예를 들어 한국어의 음절말음(coda) 위치에서는 마찰음, 파찰음, 유기폐쇄음, 경폐쇄음이 제한되므로 이러한 자음들과 평폐쇄음의 대립은 음절말음 위치에서는 중화되고, 음절두음(onset)의 위치에서만 유지된다.

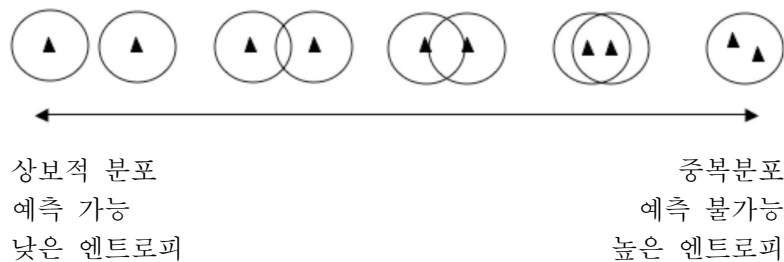
3. 정보이론과 확률적 음운대립 모델

Goldsmith (2002)와 Hall (2012)에서는 분절음 사이의 복합적인 대립 관계를 설명하기 위해 전통적 음운론의 관점에서 벗어나, ‘정보이론(Information theory)’을 도입하였다. 정보이론은 언어 체계 내에서 대응하는 두 소리의 음운론적 지위를 밝히기 위해 특정 환경에서 두 분절음이 출현할 확률을 바탕으로 ‘엔트로피(Entropy)’ 혹은 ‘정보량’이라는 개념을 이용한다. 엔트로피를 바탕으로 음운론적 지위가 불분명한 음소의 대립관계 분석하기 위하여 Hall (2009)에서는 ‘확률적 음운대립 모델(Probabilistic Phonological Relationship Model, 이하 PPRM)’을 제안하였다. 특정 환경에서 두 분절음이 나타날 확률이 동일하다면 두 분절음 가운데 어떤 음이 나타날지 예측하기 어려우므로 엔트로피는 증가하고, 반대로 특정 환경에서 두 분절음 중 한 분절음이 나타날 확률이 크다면 엔트로피는 감소한다. PPRM에서는 두 분절음이 균등한 확률을 갖는 경우로부터 한 가지 분절음으로 편중된 분포를 보이는 양 극단의 사이에서 엔트로피를 측정하여 두 분절음의 관계를 가시적인 수치로 분석한다.

3.1 확률적 음운대립 모델

정보이론에 의하면 어떤 소리가 나올지 예측이 힘든 분절음들은 엔트로피가 높고, 어떤 소리가 나올지 예측하기가 더 수월한 분절음들은 엔트로피가 낮다. 따라서 분포를 예측하기 어려운 분절음들의 엔트로피는 높고 상보적으로 분포하여 출현을 예측할 수 있는 분절음들의 엔트로피는 낮다.

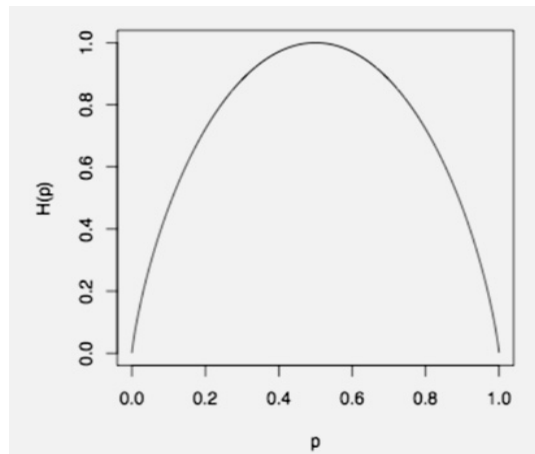
(3) 정보이론의 관점에서 본 음성의 분포와 대립 (Hall 2009: 108)



음소의 분포 환경과 규칙을 중심으로 음운의 대립을 기술하는 구조음운론 혹은 생성음운론과 달리 정보이론에서는 확률적 분석을 통해 특정 환경에서 두 분절음 중 어떤 분절음이 출현할지 예측한다. PPRM은 음운론적 지위가 불분명한 대립 쌍을 대상으로 삼아 특정 환경에서 두 분절음의 출현 가능성을 분석한다. 이 모델은 분절음 사이의 음운론적 대립관계를 밝히기 위하여 분절음의 빈도를 이용하는데, 빈도를 통하여 음운의 변화와 인지, 음소들 사이의 관계를 설명할 수 있다고 본다(Hall 2009: 82). PPRM의 빈도 분석은 확률과 엔트로피 척도에 의해 이루어진다. 엔트로피는 일반이론에서는 체계의 무질서(disorder), 무작위성(randomness)의 척도로 사용되지만 정보이론에서는 정보의 불확실성(uncertainty) 척도로 사용된다.

두 분절음이 상보적으로 분포한다고 가정해 보자. 특정 환경에서 대립되는 두 분절음 가운데 하나의 분절음만 나타나고 다른 분절음은 전혀 나타나지 않으므로 음운론적 환경이 제시되면 어떠한 분절음이 나타날 것인지 정확하게 예측할 수 있다. 따라서 불확실성이 전혀 없으므로 두 분절음의 대립에서 엔트로피는 0에 가까운 낮은 값을 가지며, 공시적인 변이음의 관계에 있다고 할 수 있다. 반면 두 분절음의 분포가 중복된다고 가정해 보자. 두 음소는 동일한 환경에서 실현되므로, 음운론적 환경에 제시되더라도 어떤 분절음이 나타날 것인지는 예측하기 어렵다. 따라서 두 분절음에 대한 불확실성이 커지고 엔트로피가 증가한다. 확률과 엔트로피의 관계를 도표로 표현하면 다음과 같다.

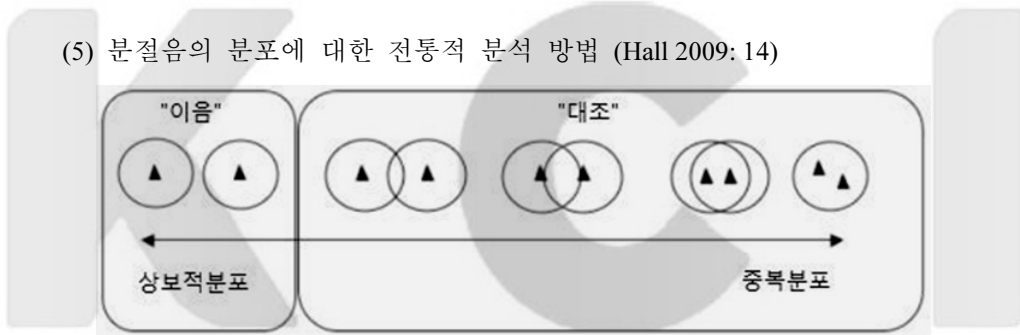
(4) 확률과 엔트로피 기여도의 상관관계 (Shanon 1948: 394)



가로 x축의 p 는 분절음이 나타날 확률을 의미하고, 세로 y축의 $H(p)$ 는 엔트로피를 의미한다. (4)에서 양 분절음 사이의 엔트로피가 가장 높은 지점은 각 분절음의 출현 확률이 0.5인 경우이다. 양 분절음의 출현 확률이 50%로서 동일한 경우 어떤 분절음이 출현할지 전혀 예측할 수 없다. 반면에 엔트로피가 가장 낮은 지점은 특정 분절음이 나타날 확률이 0이거나, 1인 경우로서 해당 분절음이 전혀 출현하지 않거나 항상 출현하는 경우이다.

Hall (2009)에서는 분절음의 분포 및 관계와 관련하여, 중복 분포를 보이면 대립 관계로, 상보적 분포를 보이면 변이음 관계로 양분하는 전통적인 음운분석 방법에 문제가 있다고 지적한다. 아래에서 보듯이, 분절음의 분포는 완전 중복분포와 상보적 분포 사이에 여러 중간 단계가 있을 수 있는데, 대립과 변이음 관계만으로 분절음들 사이의 관계를 파악하는 데에 문제가 있다는 것이다.

(5) 분절음의 분포에 대한 전통적 분석 방법 (Hall 2009: 14)



분절음의 분포에 있어서 ‘음소대립 관계’와 ‘변이음 관계’ 사이 중간 단계의 사례로는 영어의 [s]-[ʃ]와 캐나다 영어 *writer*, *rider*의 [aɪ]-[ʌɪ] 등이 있다. 영어에서는 [s]와 [ʃ]가 모음 앞 혹은 뒤에 있을 때 최소대립쌍에서 대립관계를 이루는 예를 쉽게 찾을 수 있다(e.g. *sue* [su] vs. *shoe* [ʃu], *mass* [mæʃ] vs. *mash* [mæʃ]). 하지만 [s]와 [ʃ]가 음절 첫 위치에서 자음 앞에 있을 때에는 서로 상보적인 분포를 보인다. 즉, [ʃ]는 [r] 앞에 올 수 있지만(e.g. *shrine*) [s]는 이 위치에 올 수 없고, [s]는 다른 자음([p], [t], [k], [l], [m], [n]) 앞에 올 수 있지만(e.g. *spine*, *steak*, *school*, *sleep*, *smile*, *snake*) [ʃ]는 이러한 위치에 올 수 없다. 따라서 [s]와 [ʃ]는 일부 환경에서는 상보적으로, 일부 환경에서는 대립적으로 분포한다.

캐나다 영어의 이중모음 [aɪ]와 [ʌɪ]는 일반적으로 상보적 분포를 보이는 것으로 알려져 있다(Chambers 1973). [ʌɪ]는 무성음이 후행하는 비도출환경에서 나타나는 반면(e.g. *right* [raɪt], *life* [laɪf]), [aɪ]는 후행 자음이 없거나 유

성음이 후행하는 비도출환경에서 나타난다(e.g. *tribe* [traɪb], *live* [laɪv], *rye* [raɪ]). 하지만 [ʌɪ]와 [aɪ]는 *writer* [raɪtɪ], *rider* [raɪdɪ]에서는 정확히 같은 환경에 나타나 완전 중복 분포의 예를 보인다.

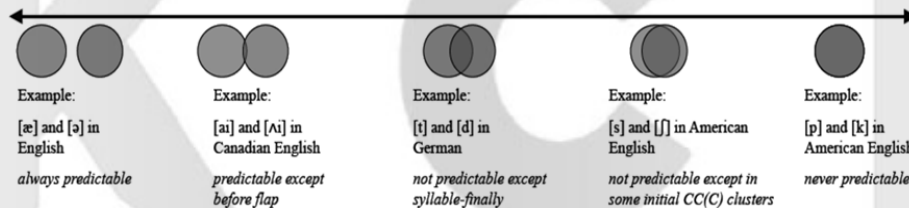
또한 독일어의 경우, 어말의 장애음들은 모두 무성음화되므로 어말에서는 무성음만 나타난다(Piroth et al. 2004). 따라서 기저형이 /ra:d/인 *Rad*(‘자전거’)와 기저형이 /ra:t/인 *Rat*(‘충고’)가 어말 위치에서 중화된다. 두 단어 모두 [rat]로 발음되므로 대립되는 두 가지 분절음 [d]와 [t] 가운데 어떤 분절음이 나타날지 예측할 수 있다. 그러나 나머지 위치에서는 서로 음소 대립을 이루고 있어서 어떤 분절음이 나타날지 예측할 수 없다.

Hall (2012)는 이와 같이 음운 대조와 변이음 관계를 동시에 가지는 분절음 쌍들을 아래와 같이 PPRM으로 분석하였다.

(6) 분절음 대립쌍 구분 (Hall 2011)

변이음 관계 (엔트로피 0.00)

음소대립 관계 (엔트로피 1.00)



위의 그림에서 영어의 [æ]와 [a]는 엔트로피 값이 0에 가까워서 두 분절음 중 특정 분절음의 출현을 거의 정확하게 예측할 수 있는 변이음 관계이고, 반대로 [p], [k]는 엔트로피 값이 1에 가까워서 두 분절음 중 어떤 분절음이 출현할지 예측이 어려운 음소 관계이다. 두 가지 극단적인 대립관계 사이에 있는 분절음의 쌍들은 0.00과 1.00의 사이의 좌표를 통하여 음운론적 대립의 성격을 확인할 수 있다.

3.2 PPRM 적용: Toy grammar

PPRM의 분석 결과는 언어 체계 내에서 두 분절음의 대립관계를 알아보기 위한 것으로서 확률과 엔트로피 값을 통하여 구할 수 있다. 분절음 X, Y의 확률은 이 두 분절음이 특정 환경에서 나타난 모든 경우의 수에 각각의 분절음이 나타난 수를 나눠주면 된다.

(7) 분절음의 출현 확률 (Hall 2009: 95)

분절음 X와 Y가 대립되는 환경 e에서 분절음 X가 나타날 확률
 $p(X/e) = N_{X/e} / (N_{X/e} + N_{Y/e})$ (N: 해당 분절음의 출현 횟수)

다양한 환경에 나타나는 분절음의 엔트로피는 다음과 같은 공식으로 구할 수 있다(Shanon 1948: 393).

(8) $H = -\sum p_i \log_2 p_i$ (H: 엔트로피, p: 확률)

언어 체계 내에서의 엔트로피 값은 개별 환경에서 두 분절음의 확률(p_i) 값에 전체 확률 중 해당 환경에서의 확률 로그값($\log_2 p_i$)을 곱한 값을 모두 합한 것이다.

Hall(2009: 103)에서는 다음과 같이 간단한 toy grammar로 엔트로피를 구하는 방법을 소개하였다.

(9) PPRM의 toy grammar (Hall 2009: 100)

환경 음성	#_a	a_#	a_a	i_i
[t]	ta, ta, ta tara, tat	at, at, at, tat	*	iti, iti
[d]	da, da, da, dara, dara, dara, dara, dara	ad, ad, ad	*	*
[r]	*	*	ara, ara, tara, dara, dara, dara, dara, dara, sara,	iri
[s]	sa, sara	As	*	*

(9)와 같은 단어들의 분포를 가진 가상의 언어가 있다고 가정하자. 분절음 [t], [d]는 환경에 따라 음소와 변이음 관계를 보이고 있다. [#_a], [a_#] 환경에서는 분절음이 모두 나타나지만, [i_i] 환경에서는 [t]만 보이고, [d]는 나타나지 않는다. 전통적인 음운론적 관점에서 보면, 앞선 [#_a], [a_#] 환경에서 [t], [d]는 서로 다른 음소이고, [i_i]에서는 두 분절음 중 한 분절음만 나타나는 상보적 분포관계를 보인다.

이 언어에서 대립되는 분절음 [t], [d]의 엔트로피 값을 구하기 위해서는 우선 확률을 구해야 한다. [#_a]라는 환경에서 분절음 [t]는 5번, [d]는 8번 사용된다. 따라서 [t]의 확률은 5/13이므로 0.38, [d]는 8/13이므로 0.62이다 ($p(t)=0.38$, $p(d)=0.62$). [t]와 [d]의 확률을 (8)의 엔트로피 공식에 대입하여 모든 환경에 대한 엔트로피를 구한 결과는 다음과 같다.

- (10) a. [#_a]: $H = -\{(0.38 \log_2 0.38) + (0.62 \log_2 0.62)\} = 0.96$
 b. [a_#]: $H = -\{(0.57 \log_2 0.57) + (0.43 \log_2 0.43)\} = 0.99$
 c. [i_i]: $H = -\{(1 \log_2 1) + (0 \log_2 0)\} = 0$

(10)을 보면 [#_a], [a_#] 환경에서는 [t], [d]가 모두 나타나고, [i_i]에서는 [t]만 나타난다. 분절음 [t]-[d]가 나타나는 세 가지 환경에 따른 엔트로피는 각각 0.96, 0.99, 0으로 엔트로피가 1에 가까운 [#_a], [a_#] 두 환경에서 이 두 분절음은 어떤 분절음이 사용될 지 예측할 수 없는 음소 관계에 있고, 엔트로피 값이 0인 [i_i] 환경에서는 확실하게 예측이 가능하므로 변이음 관계에 있다고 볼 수 있다. PPRM은 환경에 따라 서로 다른 음운론적 대립 관계를 보이는 두 분절음을 해당 언어의 전체 엔트로피 값으로 환원하여 이 분절음들이 해당 언어에서 음소 대립을 하는지, 단일 음소의 변이음으로 사용되는지 보여준다.

다만 엔트로피 값을 구할 때 고려해야 할 것이 한 가지 더 있다. 바로 가중치의 포함 여부이다. 전체 엔트로피 값은 가중치가 부여되지 않은 유형빈도(type-frequency)를 기준으로 측정하거나 가중치가 부여된 출현빈도(token-frequency)를 기준으로 측정할 수 있다(Hall 2009: 141).

- (11) [t]-[d]의 전체 엔트로피 ($p(t)=0.38$, $p(d)=0.62$)
 [#_a]: $H = 0.96$, [a_#]: $H = 0.99$, [i_i]: $H = 0$
 a. 가중치를 적용하지 않음: $(0.96 + 0.99 + 0) / 3 = 0.65$
 b. 가중치를 적용: $(0.96 \times 13/22) + (0.99 \times 7/22) + (0 \times 2/22) = 0.88$

유형빈도를 기준으로 가중치 없는 엔트로피를 구할 경우 각 분절음의 엔트로피 값을 더해서 해당되는 음운론적 환경의 종류만큼 나눠 주면 된다. (9)에서 [t], [d]가 나타난 환경은 [#_a], [a_#], [i_i] 세 가지이므로 3으로 나눈 값이 0.65이다. 이러한 분석에서는 [i_i] 환경에 나타난 분절음은 3가지 경우 밖에 없음에도 불구하고 다른 환경과 동일한 비중으로 다루어지고 있어서 실제 분절음의 대립 분포를 왜곡할 수 있다는 문제가 생긴다.

이를 보완하기 위해 Hall (2009)에서는 가중치를 적용해서 엔트로피 값을 구하였다. (11b)는 출현빈도를 고려하여 가중치를 적용한 결과이다. 세 가지 환경에 [t], [d] 가 나타난 빈도는 모두 22번이고, [#_a] 환경에서 [t], [d] 가 사용된 빈도수는 13번, [a_#]에서는 7번, [i_i]에서는 2번 사용되었다. (11b)는 해당 환경의 엔트로피 값과 전체 확률에서 차지하는 비율을 곱하여 환경에 따른 엔트로피 값들을 모두 합한 값이다. 각각의 환경에 따른 엔트로피 값을 모두 더한 값이 0.85이다.

가중치를 적용한 경우와 적용하지 않은 경우의 엔트로피 값이 달라지는데 두 분절음의 음운론적 대립관계를 살펴보기 위해서는 단순히 환경만 고려할 것이 아니라, 그 환경에서 어느 정도의 비중으로 사용되는지도 고려해야 한다. 따라서 언어 내 두 분절음의 대립을 살피기 위해서는 (11b)와 같이 가중치가 적용된 모든 엔트로피 값의 평균을 구해야 한다.

(9) toy grammar에서 [t], [d] 두 분절음의 대립관계를 PPRM으로 계산한 결과, 엔트로피 값은 $H=0.85$ 로 1에 가깝다. 따라서 이 두 분절음은 독립적인 음소로 기능한다고 볼 수 있다. 반면 [d]-[t]와 [t]-[s] 두 분절음의 대립에서는 동일한 환경에서 두 분절음 중 하나의 분절음만 쓰이는 완전한 상보적 분포관계를 보인다. 따라서 PPRM으로 분석하면 이 분절음이 나타나는 환경의 엔트로피 합은 $H=0$ 이 되어야 한다. PPRM을 적용한 결과, 아래에서 보는 것처럼 엔트로피 값은 0으로 완전한 상보적 분포관계를 보이는 변이음 관계라고 볼 수 있다.

$$\begin{aligned} (12) \text{ [#_a]: } H &= -\{(1 \log_2 0) + (0 \log_2 0)\} = 0 \\ \text{ [a_#]: } H &= -\{(1 \log_2 0) + (0 \log_2 0)\} = 0 \\ \text{ [a_a]: } H &= -\{(0 \log_2 0) + (1 \log_2 0)\} = 0 \\ \text{ [i_i]: } H &= -\{(0 \log_2 0) + (1 \log_2 0)\} = 0 \end{aligned}$$

지금까지 PPRM의 엔트로피를 통하여, 음운체계 안의 두 분절음의 대립관계를 분석하는 방법을 살펴보았다. 다음 장에서는 중세한국어의 /△/와 /ㅅ/의 대립을 PPRM을 적용하여 분석하겠다.

4. 중세한국어 /△/-/ㅅ/의 PPRM 분석

엔트로피를 분석하기 위해서는 일단 중세한국어 말뭉치를 가공하여 음소의 빈도를 구해야 한다. 본 연구에서는 홍윤표(2012)를 통하여 배포된 ‘SynKDP(감작새) 1.5.1’에 포함된 ‘21세기 세종계획 원시말뭉치’를 이용하

였다. 이 말뭉치에는 15세기부터 19세기까지의 한글문헌 자료들이 연대순으로 입력되어 있는데, 이 가운데 /△/가 포함된 15세기와 16세기의 자료를 세기별로 나누어 분석하였다. 말뭉치에 포함된 문헌자료의 목록은 다음과 같다.

(13) 원시 말뭉치에 포함된 문헌자료 목록

a. 15세기 문헌

- 1447 석보상절(釋譜詳節)
- 1447 용비어천가(龍飛御天歌)
- 1447 월인천강지곡(月印千江之曲)
- 1459 훈민정음언해본(訓民正音諺解本)
- 1459 월인석보(月印釋譜)
- 1461 능엄경언해(楞嚴經諺解)
- 1463 법화경언해(法華經諺解)
- 1464 금강경언해(金剛經諺解)
- 1464 반야심경언해(般若心經諺解)
- 1464 오대산상원사중창권선문(五臺山上院寺重創勸善文)
- 1464 선종영가집언해(禪宗永嘉集諺解)
- 1464 아미타경언해(阿彌陀經諺解)
- 1465 원각경언해(圓覺經諺解)
- 1466 구급방언해(救急方諺解)
- 1467 목우자수심결언해(牧牛子修心訣諺解)
- 1467 몽산법어약록언해(蒙山和尚法語略錄諺解)
- 1474 내훈언해(內訓諺解)
- 1481 두시언해(杜詩諺解, 초간본)
- 1481 삼강행실도(三綱行實圖, 런던대본)
- 1482 금강경삼가해(金剛經三家解)
- 1482 남명집언해(南明集諺解)
- 1485 관음경언해(觀音經諺解)
- 1485 영험약초언해(靈驗略抄諺解)
- 1489 구급간이방언해(救急簡易方諺解)
- 1496 육조법보단경언해(六祖法寶壇經諺解)
- 1496 시식권공언해(施食勸供諺解)
- 1500 개간법화경언해(改刊法華經諺解)

b. 16세기 문헌

- 1514 속삼강행실도(續三綱行實圖)
- 1517 몽산화상법어약록언해(蒙山和尚法語略錄諺解, 고운사판)
- 1517 번역노걸대(翻譯老乞大)
- 1517 번역박통사(翻譯朴通事)
- 1517 사성통해(四聲通解)
- 1517 창진방찰요(瘡疹方撮要)
- 1518 번역소학(翻譯小學)
- 1518 이륜행실도(二倫行實圖, 옥산서원본)
- 1518 정속언해(正俗諺解, 초간본)
- 1522 법집별행록절요병입사기언해(法集別行錄節要并私記諺解)
- 1525 간이벽온방(簡易辟瘟方)
- 1525 몽산화상법어약록언해(蒙山和尚法語略錄諺解, 심원사판)
- 1527 훈몽자회(訓蒙字會, 존경각본)
- 1535 몽산화상법어약록언해(蒙山和尚法語略錄諺解, 빙발암판)
- 1541 우마양저염역병치료방(牛馬羊猪染疫病治療方, 중간본)
- 1542 분문온역이해방(分門瘟疫易解方, 중간본)
- 1554 구황촬요(救荒撮要, 만력본)
- 1555 자훈언해(字訓諺解)
- 1559 훈몽자회(訓蒙字會, 내각문고본)
- 1563 부모은중경(父母恩重經)
- 1567 몽산화상육도보설언해(蒙山和尚六道普說諺解)
- 1569 선가귀감언해(禪家龜鑑諺解, 상문각)
- 1569 육자대명왕경(六字大明王經)
- 1569 칠대만법(七大萬法)
- 156x 여씨향약언해(呂氏鄉約諺解, 화산본)
- 1574 여씨향약언해(呂氏鄉約諺解, 일석본)
- 1575 광주천자문(廣州千字文)
- 1575 석봉천자문(石峰千字文, 내각문고본)
- 1575 천자문(千字文, 대동급기념문고본)
- 1576 백련초해(百聯抄解, 동경대본)
- 1576 신증유합(新增類合, 초간본)
- 1577 계초심학인문언해(誠初心學人文諺解)
- 1577 몽산화상법어약록언해(蒙山和尚法語略錄諺解, 송광사판)
- 1579 중간경민편(重刊警民編)

- 1581 속삼강행실도(續三綱行實圖, 중간본)
- 1583 석봉천자문(石峰千字文, 박찬성본)
- 1583 석봉천자문(石峰千字文, 초간본)
- 1588 소학언해(小學諺解, 도산서원본)
- 1588 주역언해(周易諺解, 초간본)
- 1590 논어언해(論語諺解, 초간본)
- 1590 대학언해(大學諺解)
- 1590 맹자언해(孟子諺解)
- 1590中庸언해(中庸諺解)
- 1590 효경언해(孝經諺解)
- 1593 선조국문교서(宣祖國文教書)
- 159x 무예제보언해(武藝諸譜諺解)
- 15xx 삼강행실도(三綱行實圖, 동경대본)
- 15xx 순천김씨묘출토간찰(順川金氏墓出土簡札)
- 15xx 장수경(長壽經)
- 15xx 훈몽자회(訓蒙字會, 동중본)

4.1 역사자료 말뭉치의 가공

말뭉치는 다음과 같은 과정에 따라 가공하였다. 음소의 빈도는 원문에서 그대로 추출하지 않고, ‘SynKDP’를 활용하여 (13)의 그림과 같이 가나다순으로 정렬한 이후에 중세한국어의 음운체계를 따르지 않는 동국정운(東國正韻)식 한자음 표기가 포함된 어절은 모두 삭제하였다. 따라서 Hall (2009)에서는 순수한 출현빈도를 기준으로 정보량을 계산하였으나, 본 연구에서는 분석의 편의를 위하여 어절단위로 정렬한 이후에 출현빈도를 측정하였다. 예를 들어 Hall (2009)의 방식에서는 부사 ‘손소’가 20번 ‘손소’가 10번 출현할 경우 /ㄴ/ 앞에서 /ㅅ/가 20번, /ㅅ/가 10번 출현한다고 계산되었으나 본 연구에서는 모두 동일한 음절이므로 1번만 출현하는 것으로 계산하였다. 따라서 본 연구에서 측정한 출현빈도는 원시말뭉치로부터 직접 측정한 것이 아니라 유형별로 정렬된 어절로부터 측정한 출현빈도이다.

(15) 가중치를 적용한 엔트로피

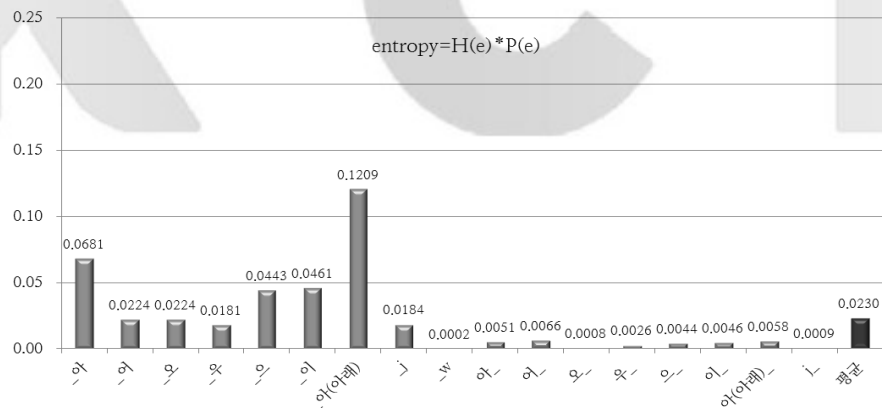
$$\sum \{H(e) \times P(e)\} \text{ (H: 엔트로피, P: 확률)}$$

두 음소의 전체 엔트로피는 0.00에서 1.00 사이의 값을 갖는데 0.00에 근접할수록 공시적인 변이음의 관계에 가깝다는 것을 의미하며, 1.00에 근접할수록 대립되는 음소의 관계와 가깝다는 것을 의미한다.

4.2 15세기 /△-/ㅅ/ 대립 분석

15세기 문헌의 발음치를 대상으로 /△/와 /ㅅ/의 대립을 분석한 결과, 가중치를 반영하여 모든 환경에서 구한 엔트로피는 0.3916이었다. 두 대립 쌍의 엔트로피가 1에 가깝다면 음소대립의 관계이고, 0에 가깝다면 변이음 관계를 가지므로 △-ㅅ은 일부 환경에서는 상보적 분포하고 일부 환경에서는 중복적으로 분포되지만, 출현빈도의 가중치를 고려할 경우, 독립적인 음소대립 관계보다는 공시적인 변이음의 관계에 가까웠다.

(16) 15세기 /△-/ㅅ/ 대립 쌍의 엔트로피 (합계: 0.3915)

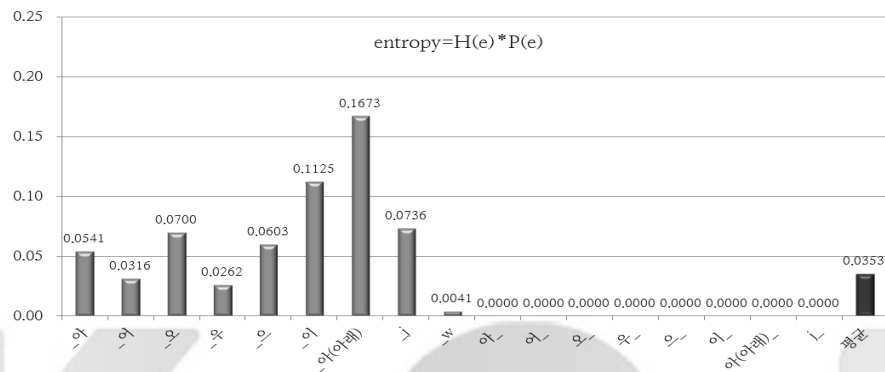


모음 /ㅛ/의 음절두음인 경우 /△/(1,421번)과 /ㅅ/(1,765번)의 출현빈도가 비슷하여 엔트로피(0.121)가 가장 높은 반면, 활음 /w/ 앞의 환경(ㅁ)과 활음 /j/ 뒤의 환경(ㅊ)에서는 /△/이 출현하는 빈도가 매우 낮아서 엔트로피도 낮은 편이었다. 결론적으로 △-ㅅ의 엔트로피는 일반적인 음소들 사이의 엔트로피보다는 상당히 낮은 편이었다. 마찰음 ㅅ-ㅎ나, 치경음 ㄷ-ㅅ, ㄴ-ㅅ, ㄹ-ㅅ 사이의 엔트로피와 △-ㅅ의 엔트로피를 비교해 보면 그 차이를

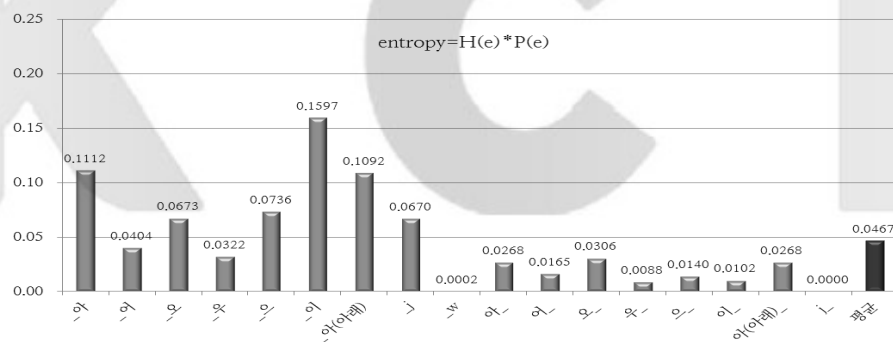
확인할 수 있다.

(17) 15세기 ㅅ-ㅎ, ㅅ-ㅈ, ㅅ-ㄴ, ㅅ-ㄹ의 엔트로피

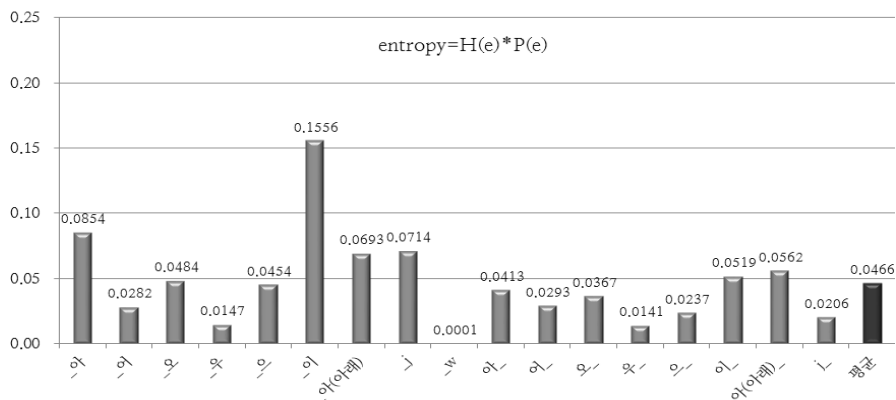
a. /ㅅ/-/ㅎ/ 대립 쌍의 엔트로피 (합계: 0.5997)

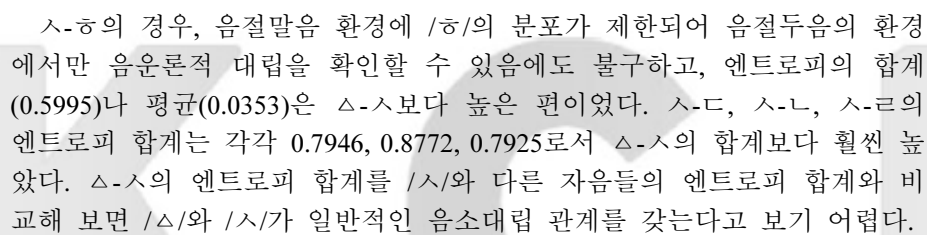


b. /ㄱ/-/ㄷ/ 대립 쌍의 엔트로피 (합계: 0.7946)



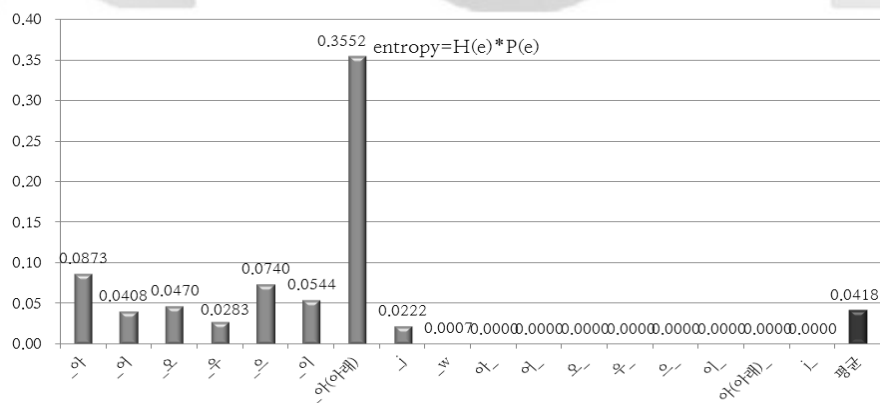
c. /ㅅ/-/ㄴ/ 대립 쌍의 엔트로피 (합계: 0.8772)



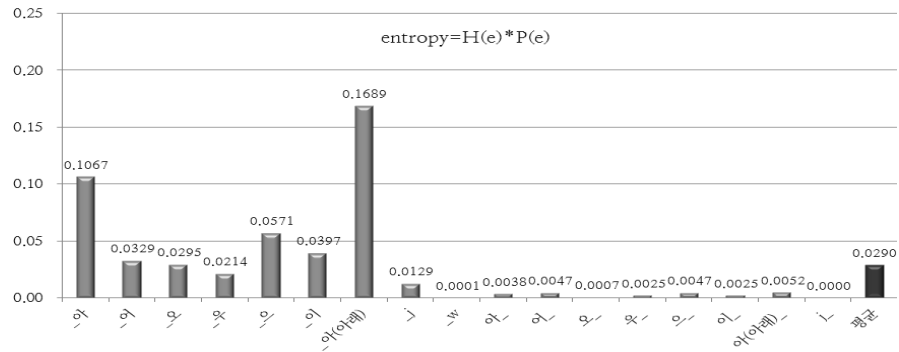


(18) 15세기 Δ -ㄷ, Δ -ㄴ, Δ -ㄹ의 엔트로피

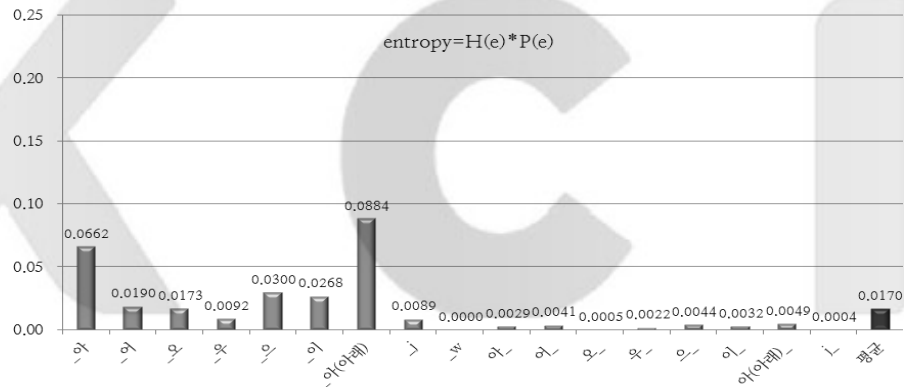
a. /△/-/ㅎ/ 대립 쌍의 엔트로피 (합계: 0.7099)



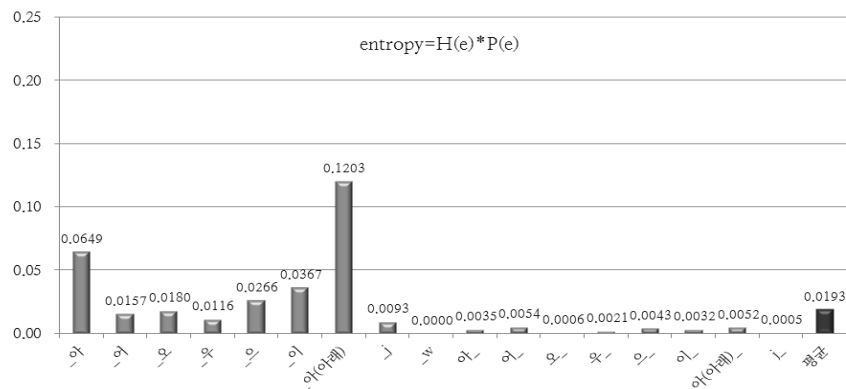
b. /△/-/ㄷ/ 대립 쌍의 엔트로피 (합계: 0.4933)



c. /△/-/ㄴ/ 대립 쌍의 엔트로피 (합계: 0.3280)

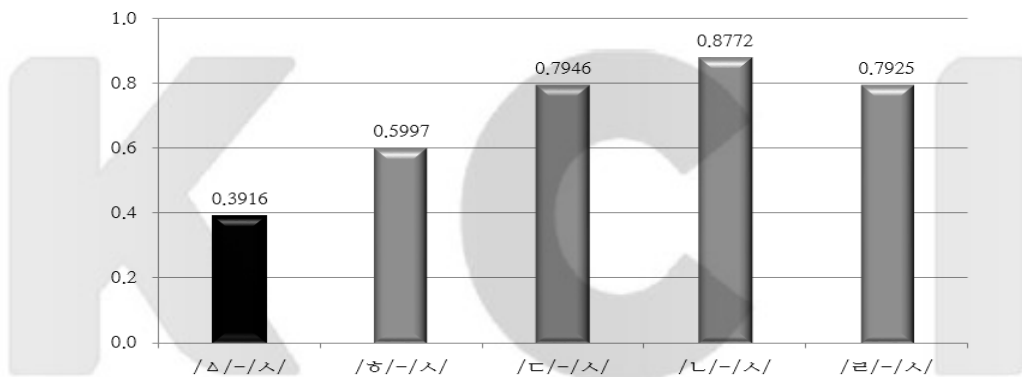


d. /△/-/ㄹ/ 대립 쌍의 엔트로피 (합계: 0.2885)

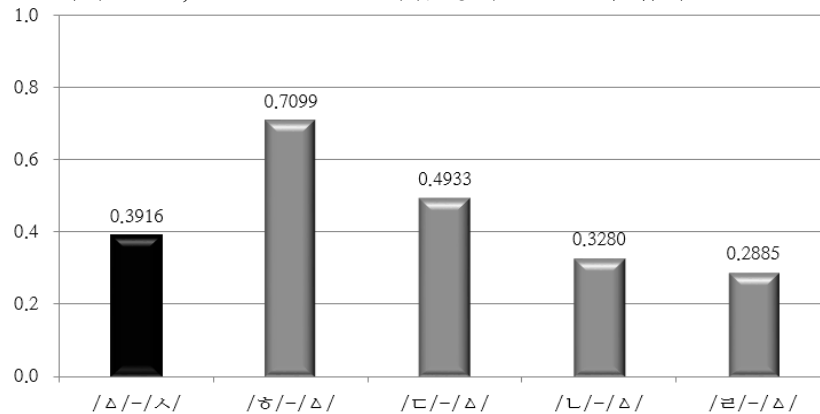


/△/과 다른 자음들 사이의 엔트로피를 분석한 결과 ㅅ-△뿐만 아니라 △-ㅎ, △-ㄷ, △-ㄴ, △-ㄹ의 엔트로피도 매우 낮은 편이었다. △-ㅎ(0.7099)만은 ㅅ-ㅎ(0.5997)보다 높은 편이었으나, △-ㄷ, △-ㄴ, △-ㄹ의 엔트로피 합계는 상당히 낮은 편이었으며 공명음인 △-ㄴ(0.3280), △-ㄹ(0.2885)의 합계는 ㅅ-△(0.3622)보다도 낮은 편이었다. 따라서 /ㅅ/와 /△/ 사이의 엔트로피, 혹은 이들과 다른 자음들 사이의 엔트로피를 분석한 결과 /△/는 다른 자음들과 일반적인 음소 대립보다 훨씬 약한 수준의 대립관계를 유지하고 있었으며, 특히 치경공명음과의 대립관계는 음소보다는 변이음에 가깝다는 점을 확인할 수 있었다.

(19) a. 15세기 △-ㅅ, ㅅ-ㅎ/ㄷ/ㄴ/ㄹ 대립 쌍의 엔트로피 합계



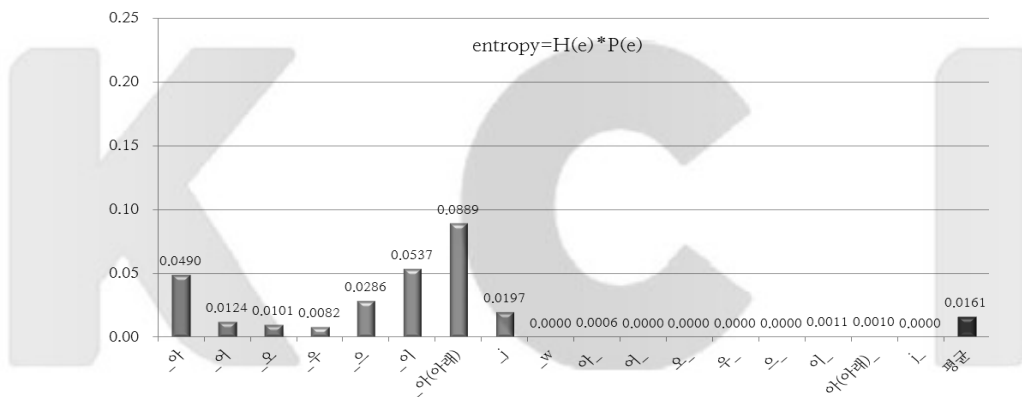
b. 15세기 △-ㅅ, △-ㅎ/ㄷ/ㄴ/ㄹ 대립 쌍의 엔트로피 합계



4.3 16세기 /△-/ㅅ/ 대립 분석

/△/는 16세기 중엽을 전후로 소멸되면서 점차 사라지므로 16세기 /△-/ㅅ/의 엔트로피는 15세기보다 낮아질 것이라고 예측할 수 있다. 16세기 말뭉치를 대상으로 /△/와 /ㅅ/의 대립을 분석한 결과, 모든 환경에서 가중치를 반영하여 구한 엔트로피의 합은 0.2734이었다. 15세기와 비교하면 전체 엔트로피는 30.17% 포인트 하락하였다. /ㄹ/ 앞의 음절두음 환경에서는 여전히 엔트로피가 높으나, /△/가 음소로서의 기능을 상실하고 공시적인 변이음으로 변화되어 점차 소멸되는 과정이 반영된 것으로 보인다².

(20) 16세기 /△-/ㅅ/ 대립 쌍의 엔트로피 (합계: 0.2734)

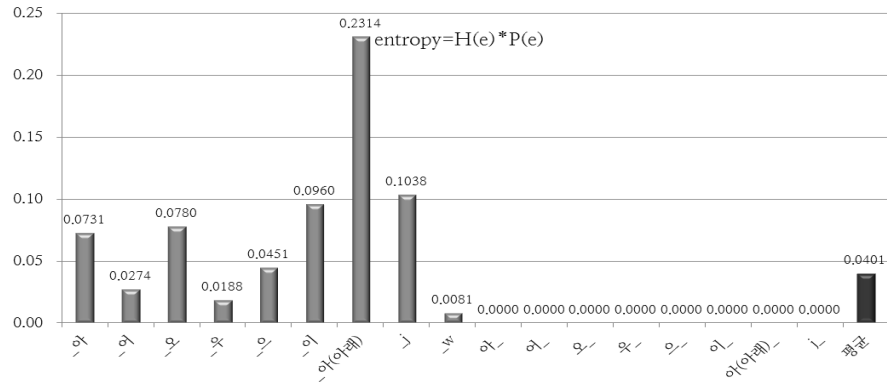


반면 16세기 ㅎ-ㅅ나 ㄷ-ㅅ의 엔트로피는 15세기보다는 조금씩 증가하였다. /ㅎ-/ㅅ/의 경우, 엔트로피의 합계가 0.5997에서 0.6816으로 15세기 기준으로 엔트로피가 13.66% 포인트 증가하였으며, /ㄷ-/ㅅ/ 역시 0.7946에서 0.8819로 10.99% 포인트 증가하였다.

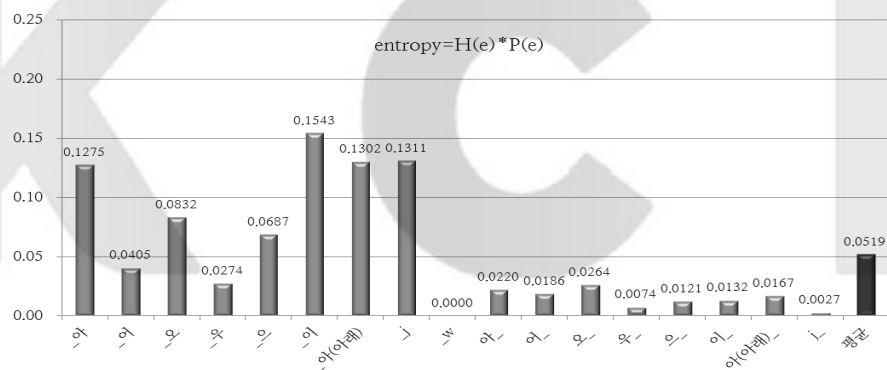
² 익명의 심사위원은 마찰음 /ㅎ(h)과 비교하였을 때, 중세한국어 /△(z)가 음소로서의 기능을 상실하고 공시적인 변이음으로 변화되어 점차 소멸된 결과는 ‘음성적 이행’(phonetic implementation) 측면에서 특이하고 이해하기 어려운 현상이라고 논평하였다. 본 연구에서 /△/의 소멸원인을 논의하기는 어려우나 ‘손소>손소>손수, 새삼>새삼(스렵다), 오소리>오소리>오소리’와 같이 무성마찰음으로 강화되어 살아남은 경우가 있으며 중앙어(중부방언) 이외의 지역어에서는 이러한 경향이 보다 강했다는 점을 밝혀 둔다. 이러한 현상에 대한 자세한 논의는 김성규(2006)을 참조하기 바란다.

a. /ㅅ/-/ㅎ/ 대립 쌍의 엔트로피 (합계: 0.6816)

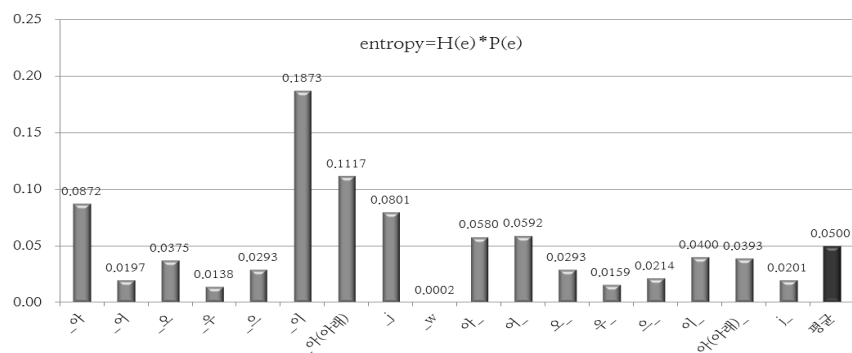
a. /ㅅ/-/ㅎ/ 대립 쌍의 엔트로피 (합계: 0.6816)



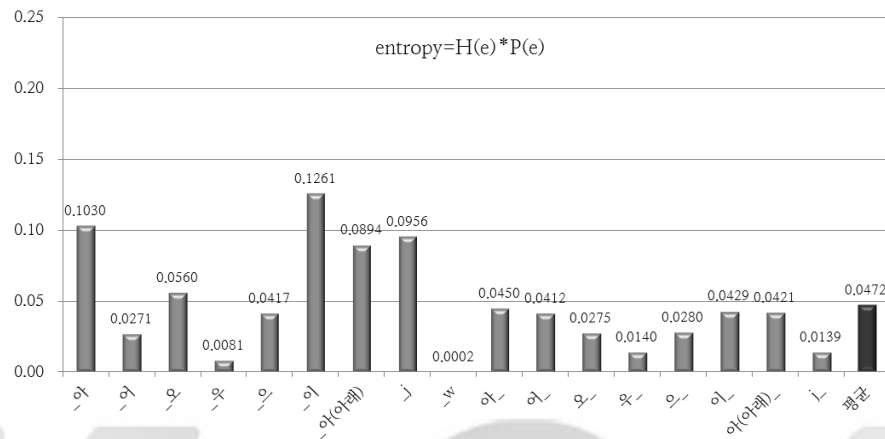
b. /ㅅ/-/ㅈ/ 대립 쌍의 엔트로피 (합계: 0.8819)



c. /ㅅ/-/ㄴ/ 대립 쌍의 엔트로피 (합계: 0.8502)



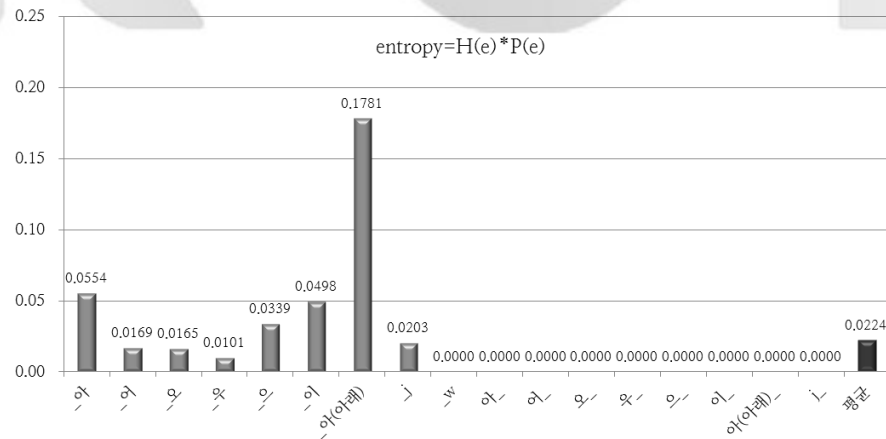
d. /ㅅ/-/ㄹ/ 대립 쌍의 엔트로피 (합계: 0.8017)



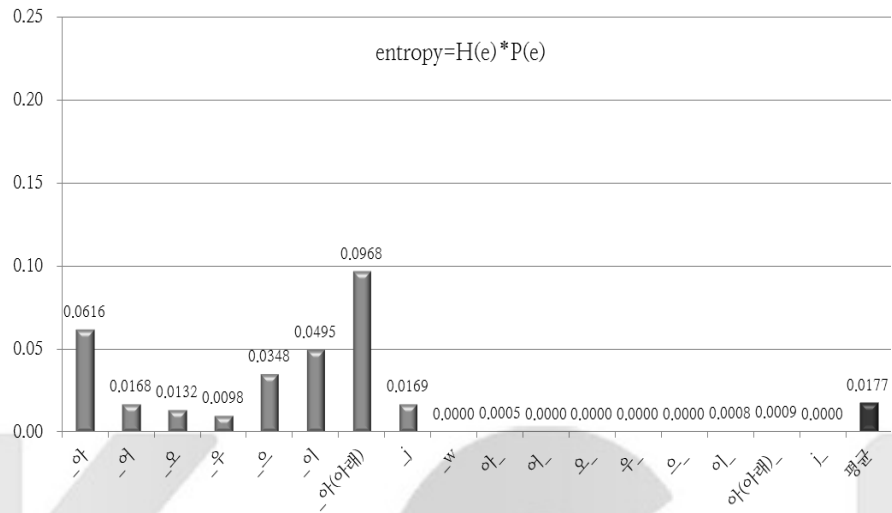
엔트로피 합계의 증가는 16세기 /ㅅ/의 빈도 증가를 반영한 것으로서 소멸 과정에서 /ㅅ/가 대부분 어중에서 탈락되었으나, 일부가 /ㅅ/로 합류된 결과를 반영하는 것으로 보인다.

(22) 16세기 △-ㅎ, △-ㄷ, △-ㄴ, △-ㄹ의 엔트로피

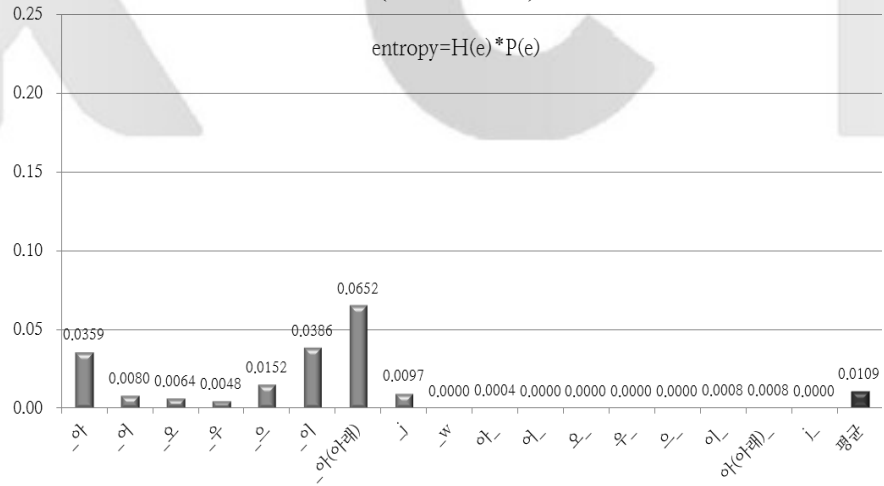
a. /ㅅ/-/ㅎ/ 대립 쌍의 엔트로피 (합계: 0.3810)



b. /△/-/ㄷ/ 대립 쌍의 엔트로피 (합계: 0.3017)

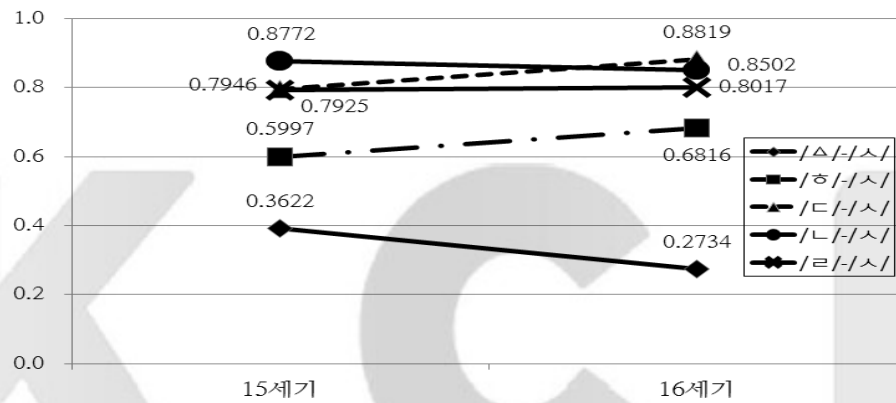


c. /△/-/ㄴ/ 대립 쌍의 엔트로피 (합계: 0.1858)

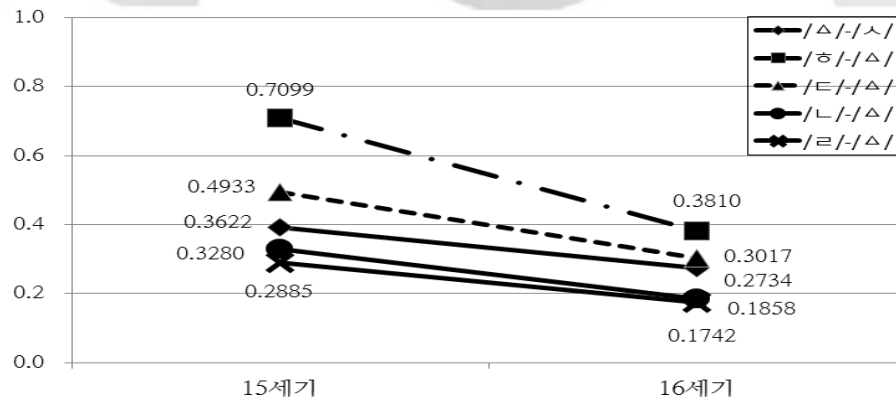


15세기와 16세기 사이 엔트로피 합계의 변화는 16세기에 /△/이 /ㅅ/으로 합류되거나 탈락되는 과정을 거쳐서 자음체계에서 사라지는 음소체계의 변화과정을 PPRM을 통하여 적절하게 분석할 수 있음을 보여준다. 다음의 도표는 15세기와 16세기 엔트로피의 변화 추이를 분석한 결과로서 무성음 /ㅅ/와 유성음 /ㅆ/이 16세기에 겪은 역사적 변화의 차이를 명확하게 보여준다.

(24) a. 15-16세기 /ㅅ/와 다른 자음 사이의 엔트로피 변화



b. 15-16세기 /△/와 다른 자음 사이의 엔트로피 변화



5. 맺음말

지금까지 정보이론과 ‘확률적 음운대립 모델’(PPRM)을 바탕으로 중세한국어의 유성마찰음 /△/와 무성마찰음 /ㅅ/의 음운론적 대립 관계를 분석하였다. 음소대립 관계와 변이음 관계라는 이분법적 범주의 틀에서 벗어나 음소의 출현빈도를 기반으로 두 음소의 관계를 구체적인 엔트로피 값으로 제시하였다. 일반적인 음운론의 관점에서는 후기 중세한국어 시기 독립된 음소로서 /△/와 /ㅅ/의 대립을 인정하고 있다. 그러나 본 연구에서 분석한 엔트로피에 의하면 15세기 /△/와 /ㅅ/의 대립은 일반적인 음소대립보다 약화되어 상보적으로 분포하는 변이음의 관계에 가까운 편이었다. 16세기에는 /△/가 음소로서의 기능을 상실하면서 △-ㅅ의 엔트로피는 더 낮아졌다. 음운의 역사적 변화에 의하여 일부의 /△/가 /ㅅ/로 합류하면서 /ㅅ/와 다른 음소들 사이의 엔트로피가 증가하는 것을 확인할 수 있었다.

PPRM (Hall 2009)에 의한 중세한국어 /△/와 /ㅅ/의 분석은 두 가지 의미를 갖는다. 첫째, 확률적 모델을 통하여 공시적으로 음소와 변이음 관계의 사이에 존재하는 복합적인 성격의 음운대립을 효과적으로 설명할 수 있다. 15-16세기 /△/와 /ㅅ/의 음운론적 대립은 ‘V__N, N__V, V__V’와 같은 유성음 사이의 환경에서만 부분적으로 나타나는데, PPRM의 분석은 이러한 부분적 음소대립을 가시적인 수치로 표시할 수 있다. 둘째, 확률적 모델을 통하여 음운체계의 역사적 변화도 설명할 수 있다. 중세한국어의 유성치경 마찰음 /△/는 16세기 중엽부터 어중에서 탈락되거나 /ㅅ/에 합류되면서 소멸을 길을 걷게 된다. △-ㅅ의 엔트로피를 포함하여 /△/와 다른 자음들의 엔트로피를 분석해 보면, 자음체계 안에서 일어난 /△/의 소멸에 의해 엔트로피의 합계가 낮아지는 것을 확인할 수 있다.

REFERENCES

- 김무림. 2004. *국어의 역사*. 한국문화사.
- 김성규. 2006. △ 변화의 예외에 대한 해석, *우리말 음운연구의 실제*, 경진문화사, 185-201.
- 박창원. 1996. *중세국어 자음 연구*. 한국문화사.
- 변군혁·홍성훈·박선우. 2013. 러시아어의 모음 [y]는 음소인가? 변이음인가? 2013 한국언어학회 봄학술대회 발표요지.
- 이기문. 1972. *국어음운사 연구*. 탑출판사.

- 이기문. 1998. *신정판 국어사개설*. 태학사.
- 이승녕. 1956. △音考, *서울대 논문집* 3, 1-235.
- 조규태. 1998. 여린 비읍에 대하여, *한글* 240 · 241, 89-124. 한글학회.
- 홍윤표. 2012. *국어정보학*. 태학사.
- CHAMBERS, JACK. 1973. Canadian raising. *The Canadian Journal of Linguistics* 18, 113-135.
- GOLDSMITH, JOHN. 2002. Probabilistic models of grammar: Phonology as information minimization. *Phonological Studies* 5, 21-46.
- HALL, KATHLEEN CURRIE. 2009. *A Probabilistic Model of Phonological Relationships from Contrast to Allophony*. PhD Dissertation. The Ohio State University, Columbus, OH.
- _____. 2011. Phonological relationships: A probabilistic model. Phonology in the Twenty-First Century. (Poster) [http://www.ling.ohio-state.edu/~kchall/Hall_2011_MOT_poster.pdf]
- _____. 2012. Phonological relationships: A probabilistic model. *McGill Working Papers in Linguistics* 22 (1), 1-14
- MITTERER, HOLGER, TAEHONG CHO and SAHYANG KIM. 2016. What are the letters of speech? Testing the role of phonological specification and phonetic similarity in perceptual learning. *Journal of Phonetics* 56, 110-123.
- PIROTH, H. G. and P. M. JANKER. 2004. Speak-dependent differences in voicing and devoicing of German obstruents. *Journal of Phonetics* 32, 81-109.
- SHANNON, CLAUDE ELWOOD. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27.3, 379-423.

Sunwoo Park
 Department of Korean Language Education
 Keimyung University
 1095 Dalgubeol-daero, Dalseo-gu, Daegu
 South Korea 42601
 e-mail: sunwoopark@kmu.ac.kr

received: March 20, 2017
 revised: April 12, 2017
 accepted: April 19, 2017