論文91-28B-9-1

한국어 多音節 單語의 초성, 중성, 종성단위의 音節間 조건부 확률

(Conditional Probability of a 'Choseong', a 'Jungseong', and a 'Jongseong' Between Syllables in Multi-Syllable Korean Words)

李 在 弘*, 李 在 鶴**

(Jae Hong Lee and Jae Hak Lee)

要 約

한국어 단어는 음절이 모여서 이루어지고 단어를 이루는 음절은 그 발생의 확률적 성질에 따라 확률변수로 간주된다. 음절은 확률변수로 간주되는 초성, 중성, 종성의 세 단위로 이루어진다. 한국어 단어 안에서의 음절들의 발생의 상관관계는 음절간 조건부 확률로 표시된다. 그런데 음절간 조건부 확률의 표본공간이 매우 커서 나열할 수 없어서 대신에 초성, 중성, 종성 단위의 음절간 조건부 확률을 고려한다. 한국어 다음절 단어에 있어서 다음절 단어의 발생 빈도수와 종류에 따른 음절의 길이 분포를 계산한다. 다음절 단어의 발생 빈도수를 고려한 음절별 누적 빈도수 통계로부터 음절의 초성, 중성, 종성 단위의 단위별 발생확률과 단위간 조건부 확률을 계산한다. 다음절 단어의 길이와 단어내에서의 음절의 위치에 따른 초성, 중성, 종성 단위의 발생확률과 조건부 확률을 계산한다. 그리고 한국어 다음절 단어내에서 인접 음절사이의 초성-초성, 중성-중성, 종성-종성, 종성-초성 단위의 음절간 조건부 확률을 계산한다.

Abstract

A Korean word is composed of syllables. A Korean syllable is regarded as a random variable according to its probabilistic property in occurrence. A Korean syllable is divided into 'choseong,' jungseong,' and 'jongseong' which are regarded as random variables. We can consider the conditional probability of syllable as an index which represents the occurrence correlation between syllables in Korean words. Since the number of syllables is enormous, we use the conditional probability of a 'choseong,' a 'jungseong,' and a 'jongseong' between syllables as an index which represents the occurrence correlation between syllables in Korean words . The length distribution of Korean words is computed according to frequency and to kind. From the cumulative frequency of a Korean syllable computed from multi-syllable Korean words, all probabilities and conditional probabilities are computed for the three random variables. The conditional probabilities of 'choseong' - 'choseong,' 'jungseong' - 'jungseong,' 'jongseong' - 'jongseong,' 'jongseong' - 'choseong' between adjacent syllables in multi-syllable Korean words are computed.

*正會員, **學生會員 서울大學校 電子工學科
(Dept. of Elec. Eng., Seoul Nat'l Univ.)

接受日字 : 1991年 7月 3日
(※ 본 論文은 峨山社會福祉事業財團의 1989年度 研究費 支援에 의하여 研究되었음.)

I. 서 론

언어는 인간의 의사소통의 가장 중요한 수단으로서 언어가 문자화 된 것이 글이다. 글은 기호들이 연속적으로 나열되어 형성된 기호열로 볼 수 있다.

이렇게 형성된 기호열의 발생의 상관관계를 표시하는 지수로서 자소(字素)의 조건부 확률이 사용된다.[1][2]

한글은 다른 언어와는 달리 초성, 중성, 종성의 각 자소가 모여서 한 음절을 구성한다. 이러한 특성을 고려한 한글 음절의 초성, 중성, 종성 단위의 발생확률과 단위간의 조건부 확률에 관한 연구는 발표된 바 있다.[2] 그리고 한글의 24자모를 영어의 26자의 알파벳에 대응되는 것으로 보고 자소단위의 엔트로피에 관한 연구도 몇차례 있었다.[3-5] 한국어 단어는 음절들로 구성되는 특징을 가지고 있다. 이러한 특성에 근거하여 음절간의 상관관계를 표시하는 지수로써 다음절 단어에 있어서 초성, 중성, 종성을 단위로 한 음절간의 조건부 확률을 사용한다. 이러한 연구는 우리 민족의 고유언어이자 세계의 유일무이한 창조문자인 한글의 철저한 확률론적 분석이라는 관점에서 의미있는 일이다. 그리고 연구결과는 한국어 음성인식 및 합성, 자연언어처리, 암호법(cryptography), 언어학, 음성학, 한글부호 표준화 연구[6-8]등에 이용될 것으로 기대된다.

한국어 다음절 단어에 있어서 초성, 중성, 종성 단위간의 조건부 확률을 계산하는데는 한국어 단어의 빈도분포가 필요하다. 1945년 이래 한국어 단어 또는 음절 단위의 빈도분포의 체계적 조사는 빈약하였다.[9][10] 조사의 내용에 있어서도 다른 언어에 대한 체계적인 분포 조사와 통계적 분석에 비하여 빈약하였다.[11] 해방후 40여년간 한국어 단어별 발생빈도분포도의 포괄적인 조사는 1956년에 문교부(지금은 교육부)에 의해 단 한번 수행되었을 뿐이고 그 이후에는 전무한 실정이다. 이 논문에서는 조건부 확률의 계산에 필요한 단어의 빈도분포로써 1956년 문교부가 조사한 "우리말 말수 사용의 잦기 조사"[9]를 사용하였다. 위의 자료는 한국어 단어가 사용되는 빈도를 조사하여 과학적인 국어의 기본형태를 파악하고 우리말의 합리적인 사용을 꾀하는 것을 목적으로 만들어졌다. 조사의 대상이 된 목적물들은 초중등교과서가 50%, 일반 간행물 중 문학 예술류가 30%, 신문, 잡지, 방송원고, 국회의사록 등이 20%의 비율로 구성하였다. 모두 56,077개의 단어가 수록되어 있는데 여기에는 조사와 외래어를 포함하며 동사와 형용사는 기본형의 형태로 수록되어 있다. 외래어 표기에 있어서는 현재와 차이를 보인다. 조사시기 이후에 사회적 변화, 인구의 이동 등에 의한 한국어 표현 방법의 변화에 따라 한국어 단어의 빈도분포도 변하였을 것으로 추정되는 바 본 연구는 1956년에 사용된 한국어에 대한 조건부 확률에 관한 연구라는 한계점을 가진다. 그러나 언어의 특성상 근본적인 변화는 없었을 것으로 추정된다.

먼저 한글을 확률론적으로 해석하고 이에 근거하여 한국어 다음절 단어에 있어서 빈도수와 종류에 따른 음절 길이의 분포와 음절의 발생확률을 계산하고, 전체 음절의 초성, 중성, 종성 단위간의 조건부 확률을 계산하고, 다음절 단어의 길이와 단어내에서 음절의 위치에 따른 초성, 중성, 종성 단위의 발생확률과 조건부 확률을 계산한다. 한국어 다음절 단어에서 인접음절의 초성-초성, 중성-중성, 종성-종성, 종성-초성 사이의 음절간 조건부 확률을 계산한다.

Ⅱ. 한글의 확률론적 해석

한글의 도형적 合字(이하 '음절')를 구성하는 초성, 중성, 종성은 각각 가능한 자소집합으로부터 어떤 확률로 발생하는 확률변수(random variable)로 간주할 수 있다. 초성, 중성, 종성을 표시하는 확률변수를 각각 X, Y, Z라 정의하고, 그 가능한 자소의 집합 즉, 표본공간(sample space)을 각각 A_X, A_Y, A_Z라 하고, 초성 X의 발생확률, 중성Y의 발생확률, 종성Z의 발생확률을 각각 p(x), p(y), p(z)라 하자. 그러면 A_X, A_Y, A_Z는 다음과 같이 주어진다.

A_X={ㄱ, ㄲ, ㄴ, ㄷ, ㄸ, ㄹ, ㅁ, ㅂ, ㅃ, ㅅ, ㅆ, ㅇ, ㅈ, ㅉ, ㅊ, ㅋ, ㅌ, ㅍ, ㅎ}

A_Y={ㅏ, ㅐ, ㅑ, ㅒ, ㅓ, ㅔ, ㅕ, ㅖ, ㅗ, ㅘ, ㅙ, ㅚ, ㅛ, ㅜ, ㅝ, ㅞ, ㅟ, ㅠ, ㅡ, ㅢ, ㅣ}

A_Z={공백소, ㄱ, ㄲ, ㄱㅅ, ㄴ, ㄴㅈ, ㄴㅎ, ㄷ, ㄹ, ㄹㄱ, ㄹㅁ, ㄹㅂ, ㄹㅅ, ㄹㅌ, ㄹㅍ, ㄹㅎ, ㅁ, ㅂ, ㅂㅅ, ㅅ, ㅆ, ㅇ, ㅈ, ㅊ, ㅋ, ㅌ, ㅍ, ㅎ}

공백소(blank)는 자소가 결여된 상태를 표시하며 종성에만 발생한다. 확률변수 X, Y, Z의 표본공간의 크기는 각각 $|A_X|=19$, $|A_Y|=21$, $A_Z|=28$이다.

하나의 음절은 초성, 중성, 종성으로 구성되는데 확률변수 X, Y, Z의 결합확률 p(x, y, z)는 특정 음절의 발생확률을 나타낸다. 음절은 그 발생의 확률적 성질에 따라 확률변수로 간주될 수 있다. 음절을 표시하는 확률변수를 S라 정의할 때 확률변수 S의 표본공간 A_S는 다음과 같이 주어진다.

$$A_S = A_X \times A_Y \times A_Z$$

따라서 확률변수 S의 표본공간의 크기는 $|A_S|=|A_X| \times |A_Y| \times |A_Z|$이다. 음절들이 모여서 이루어지는 단어도 확률변수로 간주할 수 있다.

Ⅲ. 계산 및 결과

"우리말 말수 사용의 잦기 조사"[9]에 수록된 56,077

개의 단어에 대한 빈도순서와 찾기로부터 다음절 단어에 있어서 빈도수와 종류에 따른 음절 길이의 분포와 음절의 발생확률을 계산하고, 전체 음절의 초성, 중성, 종성 단위간의 조건부 확률을 계산하고 다음절 단어의 길이와 단어내에서 음절의 위치에 따른 초성, 중성, 종성 단위의 발생확률과 조건부 확률을 계산하며 다음절 단어의 초성, 중성, 종성 단위의 음절간 조건부 확률을 계산한다.

1. 한국어 다음절 단어의 음절의 길이 분포와 음절별 발생확률

한국어 단어의 빈도수와 종류에 따른 음절의 길이 분포를 표1, 2에 보인다. 음절의 길이를 나타내는 확률변수를 L이라 하자. 여기서 L=10은 길이가 10인 단어와 그 이상인 단어를 나타낸다. 한국어 단어는 1음절 단어의 빈도가 가장 높고 2음절 단어의 종류가 가장 많음을 표1과 2에서 볼 수 있다. 1음절 단어의 발생빈도가 가장 높은 것은 조사가 있기 때문이다. 한국어 단어는 약 40.3%가 2음절 단어이고 5음절 이상의 단어는 약 0.5%만이 발생됨을 표2에서 볼 수 있다.

표 1. 한국어 다음절 단어의 빈도수에 따라 음절의 길이 분포

Table 1. Word length distribution of multi-syllable korean words according to frequency.

L	1	2	3	4	5	6	7	8	9	10
p(l)	.421480	.413039	.099342	.061079	.004240	.000730	.000071	.000011	.000004	.000003

표 2. 한국어 다음절 단어의 종류에 따른 음절의 길이 분포도

Table 2. Word length distribution of multi-syllable korean words according to kind.

L	1	2	3	4	5	6	7	8	9	10
p(l)	.034310	.402982	.285375	.218236	.043458	.013802	.001230	.000392	.000143	.000071

확률변수 X, Y, Z의 표본공간의 크기는 각각 $|A_X|=19$, $|A_Y|=21$, $|A_Z|=28$이므로 가능한 초성, 중성, 종성의 결합가지수는 $19\times21\times28=11,172$개이다. 그런데 "우리말 말수 사용의 찾기 조사"에서 조사된 단어에는 모두 1,535개의 음절만이 사용되었다. 이는 모든 가능한 초성, 중성, 종성의 결합의 약 13.7%만이 발생됨을 뜻한다. 한국어 단어에 있어서 발생되는 음절의 종류는 모든 가능한 초성, 중성, 종성의 결합에 비하여 매우 적음을 알 수 있다. 발생되는 음절을 발생확률이 높은 순서로 나열했을 때 가장 많의 발생되는 음절10개와 그것의 발생확률을 표 3에 보인다. 위의 자료에서 동사와 형용사는 기본형으로 수록되어 있으므로 음절 '다'와 '하'가 다른 음절에 비하여 발생빈도가 높다. 실제로 발생되는 1,535개의 음절중 가장 많이 발생되는 10개의 음절이 약 36.0%를 차지한다.

표 3. 음절의 발생확률

Table 3. Syllable probability.

S	다	이	하	에	가	의	을	는	리	그
p(s)	.152224	.045519	.044541	.023764	.018903	.018675	.018591	.013200	.012473	.012175

2. 다음절 단어의 길이와 단어내에서의 음절의 위치에 따른 초성, 중성, 종성 단위의 발생확률과 조건부 확률

전체 단어에서 발생 빈도를 고려하여 음절별 누적 빈도수를 구하였다. 이로부터 음절의 초성, 중성, 종성 단위의 발생확률과 초성, 중성, 종성 단위의 조건부 확률을 계산하여 각각 표4와 표5에 보인다. 초성에서는 'ㅇ'의 발생빈도가 가장 높고 그 다음으로 'ㄷ', 'ㄱ'의 순서로 발생빈도가 높음을 표4(a)에서 볼 수 있고 중성에서는 'ㅏ'의 발생빈도가 가장 높고 그 다음으로 'ㅣ', 'ㅡ'의 순서로 발생빈도가 높음을 표4(b)에서 볼 수 있고 종성에서는 '공백소'의 발생빈도가 가장 높고 그 다음으로 'ㄴ', 'ㄱ'의 순서로 발생빈도가 높음을 표4(c)에서 볼 수 있다. 음절에서 초성, 중성, 종성 단위간의 조건부 확률에는 중성이 Y일 때 초성이 X일 조건부 확률$p(x|y)$, 종성이 Z일 때 초성이 X일 조건부 확률$p(x|z)$, 초성이 X일 때 중성이 Y일 조건부 확률$p(y|x)$, 종성이 Z일때 중성이 Y일 조건부 확률$p(y|z)$, 초성이 X일 때 종성이 Z일 조건부 확률$p(z|x)$, 중성이 Y일 때 종성이 Z일 조건부 확률$p(z|y)$등이 있다. 이 여섯 종류의 조건부 확률을 모두 보이려면 그 양이 방대하므로 중성이 Y일 때 초성이 X일 조건부 확률$p(x|y)$만을 표5에 보인다.

단어의 길이에 따른 음절의 초성, 중성, 종성 단위의 발생확률과 초성, 중성, 종성 단위의 조건부 확률을 계산하여 각각 표6과 표7에 보인다. 단어의 길이가 5 이상인 단어는 0.05% 뿐이므로 단어의 길이가 4 이하인 경우만 고려한다. 여섯 종류의 조건부 확률을 모두 보이려면 그 양이 방대하므로 중성이 Y

표 4. 초성, 중성, 종성의 발생확률
Table 4. Probabilities of a choseong, a jungseong, and a jongsung.

X	ㄱ	ㄲ	ㄴ	ㄷ	ㄸ	ㄹ	ㅁ	ㅂ	ㅃ	ㅅ	ㅆ	ㅇ	ㅈ	ㅉ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
p(x)	.108966	.007222	.046754	.204413	.010847	.063960	.043656	.039998	.002746	.070430	.004787	.218130	.065485	.002821	.018410	.003228	.007847	.009545	.071134

(a) 초성의 발생확률P(x)

Y	ㅏ	ㅐ	ㅑ	ㅒ	ㅓ	ㅔ	ㅕ	ㅖ	ㅗ	ㅘ	ㅙ	ㅚ	ㅛ	ㅜ	ㅝ	ㅞ	ㅟ	ㅠ	ㅡ	ㅢ	ㅣ
p(y)	.343468	.033228	.005281	.000031	.086130	.037158	.025704	.003657	.083563	.014245	.000653	.009454	.006087	.064284	.004204	.000150	.005366	.004130	.108993	.019186	.144921

(b) 중성의 발생확률 P(y)

Z	공백소	ㄱ	ㄲ	ㄱㅅ	ㄴ	ㄴㅈ	ㄴㅎ	ㄷ	ㄹ	ㄹㄱ	ㄹㅁ	ㄹㅂ	ㄹㅅ	ㄹㅌ	ㄹㅍ	ㄹㅎ	ㅁ	ㅂ	ㅂㅅ	ㅅ	ㅆ	ㅇ	ㅈ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
p(z)	.632565	.039594	.000979	.000047	.088915	.000511	.002872	.003469	.085398	.001278	.000420	.000427	.000004	.000011	.000013	.000551	.028646	.014203	.003217	.018675	.007525	.057127	.001739	.001545	.000060	.004498	.002175	.003675

(c) 종성의 발생확률 P(z)

표 5. 중성이 Y일 때 초성이 X일 조건부 확률 p(x|y)
Table 5. Conditional probability p(x|y).

Y \ X	ㄱ	ㄲ	ㄴ	ㄷ	ㄸ	ㄹ	ㅁ	ㅂ	ㅃ	ㅅ	ㅆ	ㅇ	ㅈ	ㅉ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
ㅏ	.086096	.005982	.041038	.461020	.005939	.019542	.044693	.030692	.001905	.04849	.002403	.043387	.037735	.001183	.010776	.000602	.005430	.005589	.147489
ㅐ	.057222	.012160	.109948	.182007	.083674	.043064	.037964	.060741	.006725	.16274	.000380	.028764	.048215	.011601	.033282	.001491	.019549	.005942	.094522
ㅑ	.013604	-	.029366	-	-	.119388	-	.000188	.007037	.00394	-	.704367	.000281	.000469	-	-	-	-	.121359
ㅒ	-	-	-	-	-	-	-	-	-	-	-	.937008	.062992	-	-	-	-	-	-
ㅓ	.175550	.005313	.020068	.023264	.026364	.078167	.028582	.050842	.002272	.17179	.006141	.161004	.174868	.001775	.040562	.002318	.017540	.002456	.011123
ㅔ	.055011	.012214	.024782	.027789	.006000	.012841	.006360	.004714	.001107	.07310	.001800	.643154	.096634	.000380	.025742	.000200	.003820	.001060	.003287
ㅕ	.157363	.000578	.038331	-	-	.093153	.116728	.094493	.004491	.00040	-	.340295	.002063	.000010	.004877	.002660	-	.072961	.071592
ㅖ	.571806	.000406	-	-	-	.059342	-	-	-	-	-	.302398	-	-	-	-	-	.047284	.018764
ㅗ	.108370	.012363	.036617	.177631	.024882	.152766	.074612	.092554	.002369	.07983	.001432	.085925	.069115	.009265	.013652	.004148	.025567	.009807	.019089
ㅘ	.472800	.001826	-	-	-	-	-	.000191	-	.00013	.000157	.250209	.016504	.000226	.000261	.000661	-	-	.257026
ㅙ	.092944	.026935	-	.119499	.003794	-	-	-	-	.11722	.031108	.419196	-	-	-	.149090	.002656	-	.037557
ㅚ	.024972	.005791	.004874	.604434	.000052	.009590	.000786	.002647	-	.02001	.000105	.084139	.018211	.001834	.030920	.000026	.009669	.000105	.181825
ㅛ	.285389	-	.000204	-	-	.059626	.014245	-	.001709	.00065	-	.441473	.000407	.000122	.000041	-	-	.169394	.026740
ㅜ	.128357	.009866	.031778	.049752	.003815	.013912	.167793	.144235	.013662	.11422	.000948	.114171	.122781	.001407	.032795	.000197	.005588	.025042	.019677
ㅝ	.115505	.003123	-	.000589	-	-	.033119	-	-	-	-	.830456	-	-	.000177	-	-	-	.017031
ㅞ	.236755	.041391	-	.004967	-	-	-	-	-	-	-	.655629	-	-	.004967	.006623	.023179	-	.026490
ㅟ	.083606	.003185	.004986	.158118	.035963	-	-	.000092	-	.06200	-	.507825	.015281	.000462	.085084	.011772	.003001	-	.028623
ㅠ	.057941	.000120	.006658	.000120	.000060	.247241	.000300	.000360	-	.00120	.000660	.653491	.000480	-	-	-	-	.001679	.029690
ㅡ	.151300	.012383	.140943	.105760	.008358	.148109	.001659	.000564	.002798	.02314	.014635	.341973	.007912	.002123	.003705	.011424	.009342	.001968	.011901
ㅢ	-	-	.001433	.000026	.002776	-	-	-	-	-	.000878	.973388	-	-	-	-	.000142	-	.021357
ㅣ	.071394	.004722	.031887	.005308	.002003	.095543	.026972	.023161	.000605	.09436	.010590	.442329	.118612	.005106	.031880	.006356	.000344	.009978	.020529

표 6. 단어의 길이에 따른 초성, 중성, 종성의 발생확률
Table 6. Probabilites of a choseong, a jungseong, and a jongseong as a function of word length.

X	ㄱ	ㄲ	ㄴ	ㄷ	ㄸ	ㄹ	ㅁ	ㅂ	ㅃ	ㅅ	ㅆ	ㅇ	ㅈ	ㅉ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
p(x|l=1)	.155514	.004440	.082815	.074793	.020417	.061756	.035809	.027873	.002509	.041360	.005591	.414627	.030326	.003468	.009059	.001146	.002854	.006236	.020432
p(x|l=2)	.098298	.007215	.029753	.235997	.004974	.054783	.047089	.044052	.001805	.090125	.005287	.188273	.082001	.002432	.021779	.003317	.009892	.010189	.062920
p(x|l=3)	.099974	.011274	.057855	.244112	.010859	.089460	.057316	.046638	.005605	.065366	.003524	.113600	.073683	.002610	.021828	.003832	.007988	.011022	.073816
p(x|l=4)	.077869	.006348	.030872	.273117	.013094	.063834	.028036	.038027	.002497	.056844	.003278	.117529	.059879	.002716	.019506	.003492	.009392	.011191	.182481

(a) 단어의 길이에 따른 초성의 발생확률p(x|l)

Y	ㅏ	ㅐ	ㅑ	ㅒ	ㅓ	ㅔ	ㅕ	ㅖ	ㅗ	ㅘ	ㅙ	ㅚ	ㅛ	ㅜ	ㅝ	ㅞ	ㅟ	ㅠ	ㅡ	ㅢ	ㅣ
p(y|l=1)	.137194	.026618	.003614	.000051	.070894	.086069	.015562	.001609	.093891	.028737	.001263	.001984	.001441	.046612	.005050	.000415	.006844	.001276	.277591	.071413	.120850
p(y|l=2)	.379317	.035231	.006326	.000037	.088372	.030206	.030804	.005194	.089305	.011158	.000402	.013856	.007199	.076087	.004551	.000072	.004368	.005740	.051307	.003636	.156989
p(y|l=3)	.393523	.041370	.005694	-	.089699	.012342	.024834	.003407	.077197	.007763	.000438	.004685	.005340	.067368	.004120	.000079	.007748	.002814	.077532	.002689	.171723
p(y|l=4)	.517721	.029658	.003735	.000017	.096953	.007283	.027220	.002569	.055593	.007803	.000751	.013664	.011521	.049582	.001714	.000052	.003420	.005291	.054118	.003447	.107886

(b) 단어의 길이에 따른 중성의 발생확률p(y|l)

Z	공백소	ㄱ	ㄲ	ㄱㅅ	ㄴ	ㄴㅈ	ㄴㅎ	ㄷ	ㄹ	ㄹㄱ	ㄹㅁ	ㄹㅂ	ㄹㅅ	ㄹㅌ	ㄹㅍ	ㄹㅎ	ㅁ	ㅂ	ㅂㅅ	ㅅ	ㅆ	ㅇ	ㅈ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
p(z|l=1)	.494691	.026808	.001866	.000165	.165096	-	-	.001635	.188518	.001205	.000090	-	.000015	-	-	-	.022206	.010565	.000604	.049589	-	.025211	.000596	.004325	.000022	.002974	.003819	-
p(z|l=2)	.638863	.046997	.001033	.000011	.072544	.000914	.006090	.005635	.051572	.001796	.000510	.000825	.000002	.000023	.000019	.000988	.036634	.014674	.006153	.009641	.016465	.071125	.002337	.000810	.000109	.007067	.002249	.005061
p(z|l=3)	.713924	.037943	.000368	.000017	.062116	.000177	.000414	.002260	.054279	.000861	.000922	.000160	-	-	.000023	.000453	.024525	.017385	.000993	.012325	.000023	.057259	.002252	.000830	.000024	.002880	.000984	.006966
p(z|l=4)	.740978	.039096	.000107	-	.048925	.000460	.000239	.000948	.061042	.000249	.000116	.000160	-	.000006	.000007	.000157	.018368	.014012	.000834	.005018	.000225	.064534	.001129	.000254	.000009	.000845	.000624	.001654

(c) 단어의 길이에 따른 종성의 발생확률p(z|l)

일 때 초성이 X일 조건부 확률p(x|y)만을 표7에 보인다.

단어내에서의 음절의 위치에 따른 음절의 초성, 중성, 종성 단위의 발생확률과 초성, 중성, 종성 단위의 조건부 확률을 계산하여 각각 표8과 표9에 보인다. 각 단어에서 첫번째 음절과 두번째 음절 그리고 끝 음절에 대해서 계산하였다. 여섯 종류의 조건부 확률을 모두 보이려면 그 양이 방대하므로 중성이 Y일 때 초성이 X일 조건부 확률p(x|y)만을 표9에 보인다.

3. 초성, 중성, 종성 단위의 음절간 조건부 확률

음절간의 상호 발생의 상관관계를 표시하는 척도로 조건부 확률$p(s_n|s_p)$이 사용된다. 조건부 확률$p(s_n|s_p)$는 앞 음절이 S_P일때 그 다음 음절이 S_n일 확률이다. 그런데 위의 자료에서 실제로 발생되는 음절수는 1,535이므로 앞 음절 S_p와 그 다음 음절 S_n 간의 조건부 확률$p(s_n|s_p)$을 구하면 1,535×1,535개의 확률이 구해지므로 표시가 힘들어 진다. 따라서 음절간 상호 발생의 상관관계를 표시하는 척도로 음절간 초성, 중성, 종성 단위의 조건부 확률을 고려한다.

표 7. 음절의 길이에 따른 조건부 확률p(x|y)
Table 7. Conditional probability p(x|y) as a function of word length.

X / Y	ㄱ	ㄲ	ㄴ	ㄷ	ㄸ	ㄹ	ㅁ	ㅂ	ㅃ	ㅅ	ㅆ	ㅇ	ㅈ	ㅉ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
ㅏ	.350441	.000016	.102512	.045603	.018651	.000218	.138888	.067491	.000599	.059329	.007018	.052995	.044856	.000833	.016371	.002949	.003439	.014815	.072976
ㅐ	.071706	.000682	.140164	.072910	.345779	.000040	.005855	.087427	.000762	.059916	.000040	.036936	.005173	.042430	.048807	.000080	.000561	.004692	.076038
ㅑ	-	-	.011223	-	-	.001181	-	-	.041642	-	-	.941524	-	-	-	-	-	-	.004430
ㅒ	-	-	-	-	-	-	-	-	-	-	-	.833333	.166667	-	-	-	-	-	-
ㅓ	.584819	.000105	.016157	.034813	.002063	-	.001280	.069611	.000105	.082907	.012151	.007318	.121228	.000256	.038291	.000934	.023430	.003388	.001144
ㅔ	.003609	.005048	.028092	.026926	.001860	.000012	.000223	.000719	.000149	.023367	-	.889939	.017302	-	.002381	-	.000211	.000124	.000037
ㅕ	.047263	-	.019001	-	-	.005488	.192482	.160173	.017081	-	-	.427013	-	-	-	-	-	.092742	.038757
ㅖ	.208361	-	-	-	-	-	-	-	-	-	-	.764433	-	-	-	-	-	.027206	-
ㅗ	.132271	.021488	.004354	.336020	.079506	.173246	.066999	.013620	.000148	.071900	.000216	.034972	.026763	.013177	.004718	.004912	.006083	.001546	.008061
ㅘ	.540007	.002117	-	-	-	-	-	-	-	.000037	.000223	.444205	.002117	.000371	-	.000520	-	-	.010401
ㅙ	.007608	.054100	-	-	-	-	-	-	-	-	.000845	.883347	-	-	-	.015216	-	-	.038884
ㅚ	.001614	.025282	.008069	.040882	-	-	-	-	-	.121571	-	.289403	.179129	-	-	-	.003228	.002152	.328671
ㅛ	.146667	-	-	-	-	.005926	.015556	-	-	-	-	.214815	-	-	-	-	-	.605185	.011852
ㅜ	.132463	.012962	.072644	.093095	.001214	.000389	.100263	.118814	.041704	.243857	.001626	.005061	.110729	.001260	.006687	.000206	.000870	.016375	.039780
ㅝ	.043754	.009512	-	-	-	-	.118791	-	-	-	-	.827732	-	-	-	-	-	-	.000211
ㅞ	.048843	-	-	-	-	-	-	-	-	-	-	.938303	-	-	-	.002571	-	-	.010283
ㅟ	.063017	-	.006863	.411948	.000312	-	-	-	-	.009671	-	.489315	.009827	-	.002340	-	-	-	.006707
ㅠ	.047699	-	-	-	-	.016736	-	-	-	.001674	-	.917155	-	-	-	-	-	-	.016736
ㅡ	.091308	.003934	.202824	.082932	.002896	.157036	-	-	-	.000146	.000035	.443673	.003930	.003223	.002434	.000004	.000173	-	.005453
ㅢ	-	-	-	-	-	-	-	-	-	-	-	.999955	-	-	-	-	-	-	.000045
ㅣ	.018788	.000300	.007950	-	.008639	.005830	.015776	.016385	.000097	.043592	.030227	.794338	.038839	.000044	.009496	.001378	.000185	.004602	.012004

(a) 음절의 길이 L= 1

X / Y	ㄱ	ㄲ	ㄴ	ㄷ	ㄸ	ㄹ	ㅁ	ㅂ	ㅃ	ㅅ	ㅆ	ㅇ	ㅈ	ㅉ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
ㅏ	.066082	.009004	.028210	.519349	.002174	.025458	.037087	.030081	.001535	.063442	.001808	.041445	.043910	.001588	.011526	.000182	.005718	.005270	.106131
ㅐ	.059627	.007745	.081255	.196970	.035665	.060756	.050630	.055252	.006246	.171060	.000077	.025338	.070001	.002752	.037087	.001917	.028415	.004993	.104213
ㅑ	.019029	-	.042621	-	-	.156966	-	-	.000431	.001292	-	.627605	-	.000172	-	-	-	-	.151886
ㅒ	-	-	-	-	-	-	-	-	-	-	-	1.000000	-	-	-	-	-	-	-
ㅓ	.088158	.003877	.025257	.018533	.010064	.038982	.038298	.036295	.001880	.246359	.007550	.194022	.196364	.001559	.053058	.001374	.024055	.000857	.013460
ㅔ	.137667	.022304	.015435	.003264	.007501	.015669	.003714	.006942	.000307	.105391	.004003	.426866	.187829	.000054	.053841	.000072	.008150	.000072	.000920
ㅕ	.182574	.000194	.055890	-	-	.074154	.090085	.073199	.000919	.000106	-	.362142	-	-	.000088	.003819	-	.071484	.085346
ㅖ	.592072	-	-	-	-	.067219	-	-	-	-	-	.268037	-	-	-	-	-	.047609	.025063
ㅗ	.091109	.005532	.050595	.121823	.005196	.175778	.060872	.133483	.002629	.078869	.001452	.087706	.093915	.009026	.016137	.002055	.029164	.012698	.021962
ㅘ	.443425	.001953	-	-	-	-	-	-	-	.000146	.000049	.059748	.034170	.000049	.000244	.000976	-	-	.459240
ㅙ	.104195	-	-	.365359	.010825	-	-	-	-	.186739	.108254	.058187	-	-	-	.116373	-	-	.050068
ㅚ	.012893	.001769	.005503	.650393	.000079	.003577	-	.003970	-	.006958	-	.079481	.010888	.002123	.036989	-	.006525	-	.178852
ㅛ	.386396	-	.000076	-	-	.073769	.014375	-	.000076	.000530	-	.343194	-	-	-	-	-	.141636	.039949
ㅜ	.143967	.003651	.022856	.042513	.003178	.010537	.192020	.135456	.004309	.081504	.000587	.145413	.138884	.000215	.027495	.000086	.005204	.025168	.016958
ㅝ	.145183	.000598	-	-	-	-	-	-	-	-	-	.822262	-	-	-	-	-	-	.031957
ㅞ	.901515	-	-	-	-	-	-	-	-	-	-	.037879	-	-	-	-	-	-	.060606
ㅟ	.067082	.000748	.001995	.026185	.078180	-	-	-	-	.142519	-	.491022	.026060	.001122	.121197	.020948	.006359	-	.016584
ㅠ	.055318	-	.003795	-	-	.314546	.000095	.000095	-	.000190	-	.594079	.000190	-	-	-	-	-	.031692
ㅡ	.197779	.018821	.067749	.113385	.011868	.052537	.000212	.000191	.002983	.054278	.051029	.322406	.011635	.000180	.005584	.049861	.023726	.001178	.014596
ㅢ	-	-	.014831	-	.016330	-	-	-	-	-	.000899	.789513	-	-	-	-	.000749	-	.177678
ㅣ	.075417	.004569	.010207	.007123	.000416	.088225	.030603	.017392	.000510	.116167	.006186	.466245	.125191	.004625	.031328	.000264	.000076	.008653	.006883

(b) 음절의 길이 L= 2

표 7. 음절의 길이에 따른 조건부 확률p(x|y)
Table 7. Conditional probability p(x|y) as a function of world length.

Y \ X	ㄱ	ㄲ	ㄴ	ㄷ	ㄸ	ㄹ	ㅁ	ㅂ	ㅃ	ㅅ	ㅆ	ㅇ	ㅈ	ㅉ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
ㅏ	.059497	.004945	.064719	.514158	.012978	.017202	.051917	.025684	.003821	.029313	.003031	.044923	.038601	.000775	.008252	.000153	.002985	.004419	.112626
ㅐ	.033354	.023793	.139145	.272051	.009926	.047185	.035726	.051491	.012334	.183557	.001168	.035252	.031931	.010656	.020728	.001679	.014269	.004635	.071124
ㅑ	.003181	-	.006893	-	-	.093849	-	.001060	.000795	.004242	-	.732768	.000265	.000265	-	-	-	-	.156681
ㅒ																			
ㅓ	.072607	.013666	.015888	.033459	.038054	.122038	.049785	.084085	.004208	.144843	.000488	.147502	.213983	.002440	.027939	.003854	.007877	.006076	.011209
ㅔ	.023976	.017982	.040000	.190092	.014067	.066300	.021896	.019450	.003180	.266055	.005627	.023853	.195719	.005505	.056147	-	.002936	.016758	.030459
ㅕ	.132827	.002249	.021945	-	-	.152097	.103222	.132340	.006565	.000851	-	.319088	.000061	.000061	.004134	.000304	-	.057082	.067173
ㅖ	.686309	-	-	-	-	.043420	-	-	-	-	-	.227293	-	-	-	-	-	.029685	.013292
ㅗ	.124787	.017170	.053251	.108458	.005925	.087924	.155353	.074059	.003383	.090955	.001604	.132473	.039034	.006101	.016153	.011382	.042886	.008018	.021081
ㅘ	.445741	-	-	-	-	-	-	.002139	-	-	.000389	.094710	.022170	-	.000972	-	-	-	.433878
ㅙ	.400000	.024138	-	.055172	-	-	-	-	-	.327586	.003448	.048276	-	-	-	.062069	.024138	-	.055172
ㅚ	.105704	.033838	.004189	.105382	-	.077022	.009668	-	-	.077345	.000967	.157912	.011602	-	.046729	-	.012246	-	.357396
ㅛ	.313543	-	.001131	-	-	.109132	.010178	-	.001131	.000565	-	.374329	.002262	-	-	-	-	.170766	.016964
ㅜ	.104675	.026107	.034779	.037917	.004101	.030074	.173853	.160721	.019026	.096943	.000627	.137056	.070298	.004146	.046208	.000314	.008605	.036483	.008067
ㅝ	.165628	.001099	-	.001099	-	-	-	-	-	-	-	.827043	-	-	-	-	-	-	.005130
ㅞ	.038462	.480769	-	.057692	-	-	-	-	-	-	-	.057692	-	-	.019231	.057692	.269231	-	.019231
ㅟ	.106586	.010912	.001364	.074045	.011691	-	-	-	-	.023772	-	.623149	.004287	-	.099182	.008769	.002338	-	.033905
ㅠ	.075644	-	.034871	.001073	-	.175429	-	-	-	.003219	.000536	.667918	.002146	-	-	-	-	.015021	.024142
ㅡ	.247250	.022159	.036509	.155188	.009463	.318262	.000078	.001266	.016746	.036159	.012988	.061647	.015636	.000058	.004323	.002122	.016960	.010164	.033024
ㅢ	-	-	.001684	.000561	.046042	-	-	-	-	-	.029759	.807412	-	-	-	-	.003369	-	.111173
ㅣ	.112582	.010145	.085628	.006479	.000422	.172302	.025029	.038823	.000932	.074876	.005275	.163300	.153269	.004774	.050840	.012721	.000923	.011657	.070023

(c) 음절의 길이 L= 3

Y \ X	ㄱ	ㄲ	ㄴ	ㄷ	ㄸ	ㄹ	ㅁ	ㅂ	ㅃ	ㅅ	ㅆ	ㅇ	ㅈ	ㅉ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
ㅏ	.042119	.002099	.023868	.455599	.002942	.015619	.013769	.019461	.001266	.024768	.001270	.041308	.018906	.000615	.009099	.000968	.007786	.003272	.315266
ㅐ	.064887	.028066	.124247	.142999	.001118	.031233	.039242	.058367	.008196	.253089	.000683	.017013	.051785	.001366	.016020	.001676	.022353	.013971	.123688
ㅑ	.026134	-	.028107	-	-	.168639	-	-	.000493	.016272	-	.671105	.000986	-	-	-	-	-	.088264
ㅒ	-	-	-	-	-	-	-	-	-	-	-	1.000000	-	-	-	-	-	-	-
ㅓ	.046156	.006458	.014949	.011283	.084506	.226452	.011017	.036127	.003571	.098771	.000893	.265732	.142705	.002982	.022546	.004692	.002621	.002203	.016335
ㅔ	.025790	.008344	.052592	.059418	.036410	.087231	.020986	.020228	.026549	.203793	-	.076359	.257901	.001770	.052592	.006068	.017699	.001264	.045006
ㅕ	.201881	.000812	.011704	-	-	.182126	.163182	.067248	.003518	.000609	-	.193086	.013531	-	.026182	.003315	-	.075096	.057709
ㅖ	.653763	-	-	-	-	.089606	-	-	-	-	-	.158423	-	-	-	-	-	.093190	.005018
ㅗ	.105903	.012853	.029515	.142639	.007188	.084603	.039320	.122466	.003909	.086226	.004538	.143136	.103584	.005333	.021664	.001226	.033192	.020803	.031900
ㅘ	.264102	.000236	-	-	-	-	-	-	-	.000944	-	.150578	.016993	.000472	.001180	-	-	-	.565494
ㅙ	.098039	-	-	.068627	.002451	-	-	-	-	.183824	-	.007353	-	-	-	.639706	-	-	-
ㅚ	.036658	.002291	.002291	.806739	-	.004852	-	-	-	.015768	.000135	.016712	.004717	.002156	.011995	.000135	.020216	-	.075337
ㅛ	.092551	-	-	-	-	.015185	.015185	-	.005435	.000480	-	.727781	.000160	.000160	.000160	-	-	.134910	.007992
ㅜ	.092000	.007280	.010882	.034022	.010325	.021468	.132298	.199079	.008617	.108416	.002043	.093782	.148083	.002451	.080486	.000223	.009025	.017308	.022211
ㅝ	.090226	-	-	-	-	-	-	-	-	-	-	.899033	-	-	.003222	-	-	-	.007519
ㅞ	.035714	-	-	-	-	-	-	-	-	-	-	.821429	-	-	.035714	-	-	-	.107143
ㅟ	.157243	.003770	.021002	.050619	.047388	-	-	-	-	.002693	-	.369952	.018848	.000539	.181475	.018848	.001077	-	.126548
ㅠ	.057779	-	.000696	-	-	.151758	.001392	.001044	-	.001740	.003481	.754264	-	-	-	-	-	-	.027845
ㅡ	.361691	.041141	.027291	.186817	.035934	.090278	.023446	.004322	.002348	.059312	.029469	.046143	.016402	.001940	.008167	.005921	.030762	.006704	.021914
ㅢ	-	-	.003739	-	.010684	-	-	-	-	-	.002671	.854701	-	-	-	-	-	-	.128205
ㅣ	.069319	.003448	.081541	.004643	.000205	.136402	.036000	.035419	.001178	.107640	.004967	.245152	.168852	.013860	.042623	.013024	.000717	.024512	.010498

(d) 음절의 길이 L= 4

앞음절 초성이 X_p일 때 바로 다음 음절의 초성이 X_n일 조건부 확률$p(x_n|x_p)$를 표10(a)에 보인다. 조건부 확률p(ㄹ|ㄱ)와 p(ㅈ|ㄱ)는 각각 초성의 발생확률p(ㄹ)와 p(ㅈ)보다 크고 조건부 확률p(ㄴ|ㄱ)와 p(ㅇ|ㄱ)는 각각 초성의 발생확률p(ㄴ)와 p(ㅇ) 보다 작다. 조건부 확률p(ㄷ|ㄷ)와 p(ㄹ|ㄷ)는 각각 초성의 발생확률p(ㄷ)와 p(ㄹ)보다 크고 조건부 확률p(ㄱ|ㄷ)와 p(ㄴ|ㄷ)는 각각 초성의 발생확률p(ㄱ)와 p(ㄴ)보다 작다. 조건부 확률p(ㄷ|ㄹ)와 p(ㅎ|ㄹ) 는 각각 초성의 발생확률p(ㄷ)와 p(ㅎ)보다 크고 조건부 확률p(ㄱ|ㄹ)와 p(ㄹ|ㄹ)는 각각 초성의 발생확률p(ㄱ)와 o(ㄹ)보다 작다. 조건부 확률p(ㄹ|ㅅ) 는 초성의 발생확률p(ㄹ)보다 크고 조건부 확률p(ㅇ|ㅅ)는 초성의 발생확률 p(ㅇ)보다 작다. 조건부 확률 p(ㄷ|ㅎ)는 초성의 발생확률p(ㄷ)보다 매우 크다.

앞음절 중성이 Y_p일 때 바로 다음 음절의 중성이 Y_n일 조건부 확률$p(y_n|y_p)$를 표10(b)에 보인다. 조건부 확률p(ㅏ|ㅏ)는 중성의 발생확률p(ㅏ)보다 크고 조건부 확률p(ㅜ|ㅏ)는 중성의 발생확률p(ㅜ) 보다 작다. 조건부 확률p(ㅓ|ㅓ)는 중성의 발생확률p(ㅓ)

표 8. 단어내에서의 음절의 위치에 따른 초성, 중성, 종성의 발생확률

Table 8. Probabilities of a choseong, a jungseong, and a jongseong as a function of syllable's location in a word.

X	ㄱ	ㄲ	ㄴ	ㄷ	ㄸ	ㄹ	ㅁ	ㅂ	ㅃ	ㅅ	ㅆ	ㅇ	ㅈ	ㅉ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
첫번째	.137906	.008502	.059221	.074035	.015038	.026973	.053834	.052089	.003603	.070754	.006327	.329946	.066834	.002643	.017909	.002935	.006675	.010660	.054558
두번째	.085899	.006530	.032302	.305974	.006571	.135948	.035568	.029050	.001825	.084579	.003337	.097077	.071873	.003876	.023800	.002875	.012119	.009678	.051118
마지막	.108140	.003775	.047985	.313712	.010177	.076566	.031530	.024632	.001522	.059350	.003990	.214147	.048248	.002850	.013302	.001470	.006434	.007099	.025502

(a) 초성의 발생확률 p(x)

Y	ㅏ	ㅐ	ㅑ	ㅒ	ㅓ	ㅔ	ㅕ	ㅖ	ㅗ	ㅘ	ㅙ	ㅚ	ㅛ	ㅜ	ㅝ	ㅞ	ㅟ	ㅠ	ㅡ	ㅢ	ㅣ
첫번째	.214839	.038321	.004121	.000057	.088097	.054099	.027441	.003423	.104516	.018592	.000909	.011352	.004220	.073391	.004254	.000258	.007532	.003542	.159268	.032803	.148523
두번째	.413781	.033991	.008969	-	.100531	.018832	.029272	.005245	.067745	.011294	.000459	.004793	.010776	.064389	.004879	.000016	.003492	.006487	.058410	.002976	.153662
마지막	.378636	.025956	.005367	.000022	.070757	.047294	.018974	.003392	.072455	.017555	.000605	.003[illegible]5	.004649	.048118	.004763	.000182	.004579	.003760	.135980	.031662	.121675

(b) 중성의 발생확률 p(y)

Z	공백소	ㄱ	ㄲ	ㄱㅅ	ㄴ	ㄴㅈ	ㄴㅎ	ㄷ	ㄹ	ㄹㄱ	ㄹㅁ	ㄹㅂ	ㄹㅅ	ㄹㅌ	ㄹㅍ	ㄹㅎ	ㅁ	ㅂ	ㅂㅅ	ㅅ	ㅆ	ㅇ	ㅈ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
첫번째	.548823	.035049	.001713	.000078	.115682	.000791	.005089	.005727	.124467	.001646	.000756	.000720	.000006	.000019	.000017	.000983	.023511	.011759	.005507	.026607	.013602	.056686	.002666	.002387	.000009	.007739	.003603	.004356
두번째	.673827	.053791	.000110	.000011	.070485	.000040	.000282	.000727	.044802	.001077	.000009	.000085	.000002	-	.000012	.000028	.043011	.020794	.002288	.011850	.000012	.073049	.000626	.000543	.000163	.000587	.000431	.003439
마지막	.637001	.037279	.000829	.000076	.103037	-	-	.000816	.099626	.001161	.000039	.000015	.000008	-	-	-	.033378	.011881	.000305	.025869	-	.042609	.000375	.002213	.000105	.001517	.001858	-

(c) 종성의 발생확률 p(z)

보다 크고 조건부 확률p(ㅗ|ㅓ)는 중성의 발생확률 p(ㅗ)보다 작다. 이것은 양성모음(ㅏ,ㅗ)은 양성모음끼리, 음성모음(ㅓ,ㅜ)은 음성모음끼리 어울리는 모음조화 현상이 있기 때문이다. 조건부 확률p(ㅣ|ㅜ)는 중성의 발생확률p(ㅣ)보다 크고 조건부 확률p(ㅡ|ㅜ)는 중성의 발생확률p(ㅡ)보다 작다. 조건부 확률p(ㅓ|ㅡ)와 p(ㅗ|ㅣ)는 각각 중성의 발생확률 p(ㅓ)와 p(ㅗ)보다 크고 조건부 확률p(ㅡ|ㅡ)는 중성의 발생확률p(ㅡ)보다 작다. 조건부 확률p(ㅏ|ㅣ)는 중성의 발생확률p(ㅏ)보다 크고 p(ㅣ|ㅣ)는 중성의 발생확률 p(ㅣ)보다 작다.

앞음절 종성이 Z_p일 때 바로 다음 음절의 종성이 Z_n일 조건부 확률$p(z_n|z_p)$를 표10(c)에 보인다. 조건부 확률p(공백소|ㅆ)와 p(공백소|ㅎ)는 각각 0.99보

표 9. 단어내에서의 음절의 위치에 따른 조건부 확률 p(x|y)

Table 9. Conditional probability p(x|y) as a function of syllable's location in a word.

X \ Y	ㄱ	ㄲ	ㄴ	ㄷ	ㄸ	ㄹ	ㅁ	ㅂ	ㅃ	ㅅ	ㅆ	ㅇ	ㅈ	ㅉ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
ㅏ	.192802	.013372	.078027	.058341	.015332	.001112	.104763	.065197	.004479	.090893	.005913	.096978	.061888	.001836	.020203	.001114	.007492	.010437	.169820
ㅐ	.049682	.014757	.126517	.195358	.130145	.000141	.031720	.059320	.009838	.172642	.000164	.023327	.042110	.013043	.032577	.001644	.013630	.004978	.078409
ㅑ	.006987	-	.004476	-	-	.000437	-	-	.015939	.002729	-	.844869	.000437	.000328	-	-	-	-	.123799
ㅒ	-	-	-	-	-	-	-	-	-	-	-	.936508	.063492	-	-	-	-	-	-
ㅓ	.251233	.004892	.030685	.032963	.014472	.000041	.033162	.062663	.002946	.118289	.006159	.190337	.173853	.000327	.051367	.002089	.010948	.003605	.009968
ㅔ	.006636	.009048	.019600	.023567	.004815	.001048	.002561	.004574	.000474	.064514	.000008	.793078	.057811	.000466	.006353	.000033	.000299	.001264	.003850
ㅕ	.167268	.000607	.006443	-	-	.001312	.087741	.101921	.004984	.000082	-	.454310	.000033	.000016	.007541	.003344	-	.083150	.081249
ㅖ	.421661	-	-	-	-	-	-	-	-	-	-	.518796	-	-	-	-	-	.048896	.010647
ㅗ	.123436	.013072	.049143	.198022	.034693	.071246	.091605	.107170	.002750	.079019	.001859	.089406	.072107	.008105	.012650	.003168	.017458	.008561	.016529
ㅘ	.521922	.001863	-	-	-	-	-	-	-	.000121	.000194	.334132	.012219	.000290	.000363	.000605	-	-	.128291
ㅙ	.114356	.034653	-	.144554	.003960	-	-	-	-	.017822	.040594	.542574	-	-	-	.055446	-	-	.046040
ㅚ	.027701	.008560	.003804	.667314	.000040	-	-	.004003	-	.024174	.000079	.091979	.021003	.002140	.042046	-	.006975	.000159	.100024
ㅛ	.355826	-	-	-	-	.000853	.021320	-	.004371	.001066	-	.349323	.000853	-	-	-	-	.219273	.047117
ㅜ	.117945	.008073	.042486	.054537	.003359	.000257	.147920	.161197	.019144	.123651	.001103	.104428	.146424	.000650	.024378	.000147	.003224	.021510	.019566
ㅝ	.100042	.005393	-	.000106	-	-	.059433	-	-	-	-	.806895	-	-	-	-	-	-	.028130
ㅞ	.235602	.043630	-	.005236	-	-	-	-	-	-	-	.684119	-	-	-	.006981	-	-	.024433
ㅟ	.084936	.000239	.005555	.202246	.043483	-	-	.000119	-	.074005	-	.481364	.016485	-	.064150	.000179	.003882	-	.023354
ㅠ	.065667	-	.005462	.000254	-	.002540	.000127	.000127	-	.000762	-	.860663	.000635	-	-	-	-	.003556	.060206
ㅡ	.153418	.009971	.156110	.099501	.005420	.115350	-	.000489	.000054	.007098	.013027	.394464	.008047	.002390	.002627	.011256	.008595	.000158	.012025
ㅢ	-	-	.000055	.000014	.002757	-	-	-	-	-	.000809	.977920	-	-	-	-	.000151	-	.018295
ㅣ	.047485	.002578	.002856	.000382	.003383	.002299	.026055	.027524	.000382	.076515	.014070	.668907	.089649	.002614	.018498	.001048	.000124	.013152	.005452

(a) 첫번째 음절

표 9. 단어내에서의 음절의 위치에 따른 조건부 확률 p(x|y)

Table 9. Conditional probability p(x|y) as a function of syllable's location in a word.

X \ Y	ㄱ	ㄲ	ㄴ	ㄷ	ㄸ	ㄹ	ㅁ	ㅂ	ㅃ	ㅅ	ㅆ	ㅇ	ㅈ	ㅉ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
ㅏ	.039288	.002941	.020434	.668927	.001417	.046243	.018155	.019254	.000870	.041274	.000737	.021887	.034942	.001263	.008545	.000475	.006810	.004363	.062175
ㅐ	.072746	.007686	.067896	.157570	.002608	.125841	.049275	.065151	.001167	.142998	.000755	.029556	.059866	.010157	.034245	.000297	.033056	.008373	.130759
ㅑ	.019246	-	.050455	-	-	.202427	-	-	.000260	.000607	-	.601994	-	-	-	-	-	-	.125011
ㅒ																			
ㅓ	.080618	.004973	.006853	.010086	.043871	.186853	.021757	.032640	.001439	.240430	.006876	.129570	.154275	.003860	.028688	.002529	.029453	.001029	.014201
ㅔ	.300921	.027664	.042735	.011272	.008175	.045708	.006152	.002849	.000413	.097031	.011024	.015277	.294397	-	.115942	.000165	.018911	.000289	.001074
ㅕ	.153828	.000611	.091935	-	-	.216145	.157679	.079637	.000956	.000213	-	.186926	.005525	-	.001222	.001063	-	.060989	.043272
ㅖ	.727057	-	-	-	-	.122461	-	-	-	-	-	.073388	-	-	-	-	-	.048184	.028910
ㅗ	.064953	.007288	.008057	.144289	.003099	.358076	.038887	.045407	.001240	.080173	.000448	.054910	.070853	.012752	.017378	.003225	.047484	.013969	.027513
ㅘ	.333563	.001377	-	-	-	-	-	.000757	-	.000069	-	.035663	.028709	-	-	.000895	-	-	.598967
ㅙ	.023729	.001695	-	.022034	-	-	-	-	-	.459322	-	.006780	-	-	-	.464407	.011864	-	.010169
ㅚ	.040720	-	.014114	.032284	.000162	.059377	.004867	-	-	.015899	.000162	.142927	.025308	-	.019306	.000162	.030500	-	.614212
ㅛ	.238185	-	.000216	-	-	.088102	.010174	-	-	.000216	-	.519446	-	-	-	-	-	.128364	.015297
ㅜ	.151169	.014008	.014057	.043764	.004577	.034912	.184559	.114421	.003212	.099109	.000712	.140771	.082263	.002850	.050792	.000205	.010422	.025783	.022413
ㅝ	.120797	.000319	-	.001275	-	-	-	-	-	-	-	.873785	-	-	.000478	-	-	-	.003347
ㅞ	.400000	-	-	-	-	-	-	-	-	-	-	.050000	-	-	.100000	-	.350000	-	.100000
ㅟ	.053440	.012469	.002004	.007793	.009352	-	-	-	-	.020708	-	.629259	.003563	.002227	.169895	.041862	-	-	.047428
ㅠ	.049023	-	.005873	-	-	.474290	.000479	-	-	.001079	.000120	.466858	-	-	-	-	-	-	.002277
ㅡ	.146450	.021193	.087130	.129556	.013326	.296440	.000346	.000333	.014897	.072433	.022751	.137677	.005711	.000226	.007162	.012594	.012221	.008613	.010943
ㅢ	-	-	.025608	-	.001045	-	-	-	-	-	.001307	.892083	-	-	-	-	-	-	.079958
ㅣ	.094890	.008577	.066183	.014609	.000081	.234108	.031156	.017772	.000531	.131567	.004539	.148296	.151479	.010100	.054934	.006781	.000663	.006168	.017564

(b) 두번째 음절

X \ Y	ㄱ	ㄲ	ㄴ	ㄷ	ㄸ	ㄹ	ㅁ	ㅂ	ㅃ	ㅅ	ㅆ	ㅇ	ㅈ	ㅉ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
ㅏ	.070072	.000507	.027001	.722215	.003151	.021964	.029604	.019547	.000333	.028441	.001271	.016160	.026479	.000699	.006512	.000576	.003127	.004514	.017827
ㅐ	.077474	.005061	.084129	.119677	.150528	.071009	.030106	.073470	.000520	.115171	.000052	.043468	.039967	.020885	.047022	.000780	.019498	.005148	.096036
ㅑ	.018607	-	.044506	-	-	.168804	-	.000335	.012153	.000251	-	.639091	.000084	.000503	-	-	-	-	.115665
ㅒ	-	-	-	-	-	-	-	-	-	-	-	.833333	.166667	-	-	-	-	-	-
ㅓ	.302096	.002639	.007693	.019723	.013021	.044258	.005118	.048321	.000299	.233391	.010593	.054698	.175902	.001729	.035014	.001380	.034022	.001793	.008310
ㅔ	.073035	.010387	.032066	.031923	.002454	.012337	.000685	.000961	.000114	.037060	.002112	.687882	.074738	-	.029070	.000228	.004613	.000114	.000219
ㅕ	.111649	.000261	.081608	-	-	.134623	.164805	.114968	.009626	.000640	-	.247789	-	-	.000119	.000403	-	.072101	.061408
ㅖ	.642260	-	-	-	-	.090596	-	-	-	-	-	.209577	-	-	-	-	-	.035814	.021754
ㅗ	.106435	.015641	.004309	.246554	.044152	.285056	.054335	.022470	.000149	.073937	.000155	.027351	.043115	.014225	.010822	.006513	.023153	.007047	.014579
ㅘ	.484880	.001999	-	-	-	-	-	-	-	.000051	.000179	.316206	.011968	.000256	-	.000615	-	-	.183845
ㅙ	.008929	.047619	-	-	-	-	-	-	-	.086310	.000744	.781994	-	-	-	.037202	-	-	.037202
ㅚ	.013695	.006636	.013836	.014118	.000141	.012848	-	-	-	.052097	-	.187068	.064662	-	.005506	-	.012142	.000565	.616688
ㅛ	.296855	-	.000290	-	-	.117658	.006870	-	-	.000484	-	.324528	.000097	.000097	.000097	-	-	.236091	.016933
ㅜ	.156959	.008031	.032891	.060173	.002235	.017708	.182258	.119795	.018101	.156847	.000944	.064259	.091213	.000748	.022037	.000196	.004759	.028862	.031985
ㅝ	.105497	.004439	-	.000094	-	-	.053079	-	-	-	-	.835002	-	-	-	-	-	-	.001889
ㅞ	.064198	-	-	-	-	-	-	-	-	-	-	.903704	-	-	-	.002469	.017284	-	.012346
ㅟ	.063464	.000688	.004716	.261715	.000196	-	-	-	-	.007270	-	.558699	.009235	.000884	.060811	.022203	-	-	.010119
ㅠ	.049414	-	.006939	-	-	.415171	-	-	-	.001077	-	.523929	.000239	-	-	-	-	-	.003230
ㅡ	.106954	.005250	.191655	.083153	.005174	.154452	.000079	.000129	.001019	.015996	.003262	.409469	.004764	.003014	.003781	.003030	.001102	.000457	.007259
ㅢ	-	-	.001435	.000014	.000043	-	-	-	-	-	-	.994203	-	-	-	-	-	-	.004305
ㅣ	.079160	.003501	.029875	.007661	.003760	.113705	.024787	.018624	.000725	.098405	.017707	.432338	.103385	.003561	.025527	.000714	.000166	.004710	.035236

(c) 마지막 음절

표 10. 음절간 초성, 중성, 종성 단위의 조건부 확률

Table10. Conditional probabilities for a choseong, a jungseong, and a jongseong between syllables.

x_n \ x_p	ㄱ	ㄲ	ㄴ	ㄷ	ㄸ	ㄹ	ㅁ	ㅂ	ㅃ	ㅅ	ㅆ	ㅇ	ㅈ	ㅉ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
ㄱ	.094906	.008342	.018298	.202857	.000802	.175594	.038096	.023673	.003249	.070712	.002933	.087887	.132145	.001068	.028771	.001474	.010042	.005896	.093257
ㄲ	.011995	.039212	.040513	.320150	.011513	.073655	.025820	.027458	.000385	.058673	.002168	.088058	.238114	.013103	.016571	.000289	.002938	.001831	.027554
ㄴ	.062635	.010163	.023202	.405530	.001450	.141452	.057553	.019717	.004082	.028929	.002486	.097961	.032573	.004143	.011162	.002449	.012539	.007165	.074808
ㄷ	.044353	.003897	.015024	.363867	.001275	.118370	.026065	.028828	.000767	.076827	.000759	.132338	.047318	.001706	.021143	.001471	.007489	.010274	.098228
ㄸ	.032555	.008930	.042761	.199112	.034682	.204498	.115101	.011435	.000473	.004300	.000614	.099036	.008033	.000142	.002882	.000898	.000662	.001323	.232565
ㄹ	.044609	.003694	.044450	.502194	.001966	.007751	.023663	.018457	.002910	.052236	.002921	.033880	.028107	.000966	.010093	.002762	.002830	.006467	.210045
ㅁ	.049928	.003938	.053348	.245023	.004315	.109495	.023244	.009554	.000876	.065644	.013069	.143810	.096926	.001366	.034513	.007877	.005512	.005107	.126455
ㅂ	.070018	.008782	.025464	.300288	.005436	.117450	.021462	.036213	.003168	.050674	.008941	.085276	.071808	.005183	.025229	.003571	.049737	.014602	.096700
ㅃ	.043106	.010647	.019995	.284082	.026487	.303558	.002337	.001039	.023890	.015580	.000649	.111789	.092833	.011296	.017788	.000519	.001298	.000779	.032329
ㅅ	.133451	.006852	.023666	.121244	.004588	.146906	.046020	.032310	.001870	.106924	.001418	.100584	.067907	.004594	.021251	.015265	.012601	.021481	.131069
ㅆ	.025253	.000957	.006505	.610101	.001244	.043811	.007940	.010331	.001339	.007174	.023914	.180696	.005070	-	.000096	-	.007653	.000383	.067534
ㅇ	.075759	.003429	.051634	.379743	.015798	.141050	.030313	.020300	.001375	.092375	.000848	.056324	.045199	.006376	.011148	.002722	.004653	.003839	.057115
ㅈ	.121347	.003590	.028379	.257322	.000707	.062590	.038405	.053556	.000834	.089014	.001362	.110232	.075519	.000286	.035133	.003571	.004838	.010746	.102567
ㅉ	.124505	.015439	.004949	.360451	.001188	.075812	.069873	.012272	.002573	.040182	.003959	.073436	.032462	.021180	.000396	.001188	.010095	.016429	.133610
ㅊ	.088146	.000961	.025076	.268841	.000559	.093554	.038441	.035446	.000268	.075273	.001073	.100013	.083408	.000939	.043291	.001564	.004157	.017276	.121715
ㅋ	.025502	.009218	.021610	.785436	.000717	.021610	.004199	.002356	.000102	.006862	.003687	.022941	.019357	-	.002458	.022122	.000205	.005121	.046497
ㅌ	.064360	.018019	.069253	.143457	.000864	.092971	.019861	.039664	.000806	.107017	.001785	.093662	.105866	.000288	.010880	.000748	.025157	.013989	.191353
ㅍ	.076172	.002022	.015780	.177575	.005714	.095029	.040086	.044745	.000527	.085579	.000791	.161575	.116918	.001363	.022768	.000176	.016747	.016615	.120039
ㅎ	.033096	.003975	.015131	.772882	.000577	.022348	.010098	.014277	.000299	.031100	.002322	.020751	.023064	.000378	.006179	.000169	.004023	.004240	.035071

(a) 조건부 확률 $p(x_n|x_p)$

표 10. 음절간 초성, 중성, 종성 단위의 조건부 확률

Table10. Conditional probabilities for a choseong, a jungseong, and a jongseong between syllables.

x_n / y_p	ㅏ	ㅐ	ㅑ	ㅒ	ㅓ	ㅔ	ㅕ	ㅖ	ㅗ	ㅘ	ㅙ	ㅚ	ㅛ	ㅜ	ㅝ	ㅞ	ㅟ	ㅠ	ㅡ	ㅢ	ㅣ
ㅏ	.623735	.019913	.003662	-	.029037	.006959	.017646	.001775	.035337	.004204	.000110	.007040	.006061	.040601	.001913	.000007	.003567	.002985	.053847	.000488	.141115
ㅐ	.391273	.028928	.009787	-	.061259	.013359	.042745	.000628	.097527	.037446	.000079	.013516	.016132	.099987	.005574	-	.003127	.008832	.022530	.001688	.145584
ㅑ	.436395	.035265	.005540	-	.038888	.008630	.018645	.001918	.099936	.005114	-	.006179	.010121	.081078	.000959	-	.001811	.004155	.026529	.000852	.217984
ㅒ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.012658	-	.987342
ㅓ	.401677	.035575	.004050	-	.130487	.019298	.028931	.001045	.038942	.006440	.000021	.006566	.004496	.070777	.003608	.00011	.002117	.005788	.047094	.004769	.188311
ㅔ	.105613	.008560	.002006	-	.416076	.173131	.020084	.036891	.039677	.001783	.000089	.001917	.001382	.050198	.003277	-	.000914	.003700	.022112	.000446	.112144
ㅕ	.312812	.046771	.024344	-	.131267	.030287	.050067	.006220	.071195	.040210	.000568	.011287	.003573	.075450	.008396	.000065	.001478	.007178	.037595	.005229	.136009
ㅖ	.262910	.025059	.016752	-	.137478	.003600	.011076	.001938	.117403	.007476	.000277	.058563	.005261	.062301	.003877	-	.000277	.001800	.123910	.004707	.155337
ㅗ	.449163	.041770	.015223	.000006	.057889	.009515	.024749	.003308	.067007	.008851	.000170	.006071	.009418	.059587	.004136	.000023	.003223	.007767	.071013	.002338	.158763
ㅘ	.269780	.032742	.007955	-	.136703	.023812	.068514	.083721	.107046	.010878	.000162	.010174	.010174	.074467	.005899	.000054	.001948	.007522	.022026	.001624	.124797
ㅙ	.399381	.007740	.003870	-	.037152	.000774	.003096	-	.027864	.015480	.008514	.022446	.002322	.053406	-	-	-	.000774	.016254	-	.400929
ㅚ	.812098	.008333	.003700	-	.017503	.001158	.005341	.004408	.046847	.002735	-	.007336	.005663	.042342	.000901	.000225	.000676	.000644	.010618	.009041	.020431
ㅛ	.383702	.028100	.011240	-	.087320	.006603	.033087	.000632	.091254	.051633	-	.030699	.004355	.073902	.004285	-	.000070	.047278	.013137	.005971	.126730
ㅜ	.312714	.023399	.006615	-	.114208	.030820	.037416	.001652	.059123	.017230	.000872	.008136	.016364	.077015	.004412	-	.006130	.005697	.060198	.010595	.207403
ㅝ	.323147	.056261	.001254	-	.026955	.012381	.033380	.005642	.097320	.007836	-	.054067	.006739	.094963	.004388	-	.007836	.005015	.021627	.000627	.240558
ㅞ	.437186	.075377	-	-	.165829	-	.035176	-	.170854	-	-	-	.020101	.020101	-	.040201	-	-	-	.035176	-
ㅟ	.548685	.037537	.003919	-	.066713	.005574	.027608	.000174	.046508	.003571	-	.002177	.014806	.042850	.056652	-	.003745	.000261	.027696	.001481	.111043
ㅠ	.198821	.037768	.010464	-	.094419	.019846	.122925	.006736	.071085	.022492	.011787	.006375	.017681	.062786	.029468	-	.000962	.006014	.014193	.011306	.254871
ㅡ	.374903	.028410	.003326	-	.173148	.009826	.015425	.001823	.174891	.001743	.000036	.001133	.001852	.014111	.002055	-	.000871	.002535	.036341	.000123	.157447
ㅢ	.459819	.026882	.005801	-	.030985	.001132	.045133	.003962	.073147	.005235	-	.040040	.009479	.065507	.071307	-	.001415	.002830	.006084	.012450	.138936
ㅣ	.650706	.025153	.006750	-	.065465	.009198	.013198	.003081	.034736	.004299	.000715	.004750	.013074	.047412	.004318	.000003	.001243	.003850	.041931	.000636	.069481

(b) 조건부 확률 $p(y_n|y_p)$

x_n / y_p	공백소	ㄱ	ㄲ	ㄱㅅ	ㄴ	ㄴㅈ	ㄴㅎ	ㄷ	ㄹ	ㄹㄱ	ㄹㅁ	ㄹㅂ	ㄹㅅ	ㄹㅌ	ㄹㅎ	ㅁ	ㅂ	ㅂㅅ	ㅅ	ㅆ	ㅇ	ㅈ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
공백소	.779719	.033813	.000017	.000001	.044293	.000260	.000085	.000789	.029691	.001046	.000003	.000034	.000003	.000003	.000013	.000019	.037536	.015223	.000267	.012286	.000104	.042177	.000390	.000164	.000174	.000325	.000244	.004319
ㄱ	.606478	.082879	-	-	.085244	-	-	.001040	.064102	.000065	-	.000026	-	-	-	.000013	.024585	.032590	.000325	.003988	.000013	.097952	.000286	.000273	.000143	.000598	.000390	.000013
ㄲ	.931280	.000474	-	-	.009479	-	-	-	.000948	-	-	-	-	-	-	-	.054185	.000474	-	.005687	-	.000474	-	-	-	-	-	-
ㄱㅅ	.450000	-	-	.050000	.400000	-	-	-	-	-	-	-	-	-	-	.050000	-	-	.050000	-	-	-	-	-	-	-	-	-
ㄴ	.538414	.082086	.000570	-	.102898	-	.000885	.000362	.078975	.000477	.000069	.000015	-	-	-	-	.032315	.019465	.000870	.004373	.000046	.135383	.001047	.000554	-	.000462	.000693	.000038
ㄴㅈ	.991266	-	-	-	.008248	-	-	-	-	-	-	-	-	-	-	-	.000485	-	-	-	-	-	-	-	-	-	-	-
ㄴㅎ	.996377	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.003623	-	-	-	-	-	-	-	-	-	-	-
ㄷ	.971201	.000082	-	-	.001969	-	-	.000164	.013046	-	.000164	-	-	-	-	-	.010912	.000820	-	-	-	.001641	-	-	-	-	-	-
ㄹ	.695854	.033368	.000089	-	.072878	.000024	.000032	.001477	.058681	.000292	-	.000016	-	-	-	.000089	.037165	.019082	.001022	.001138	-	.067175	.002986	.002158	.000049	.001193	.000625	.001606
ㄹㄱ	.911913	.008925	-	-	.040357	-	-	.002716	.008149	.000388	-	-	-	-	-	-	.005433	-	-	-	-	.018238	-	.000776	-	-	.002716	.000388
ㄹㅁ	.939055	-	-	-	.041045	-	-	-	-	-	-	-	-	-	-	-	.014925	.004975	-	-	-	-	-	-	-	-	-	-
ㄹㅂ	.948490	.020722	-	-	.001184	-	-	-	-	-	-	-	-	-	-	-	.000592	-	-	-	-	-	-	-	-	-	.029011	-
ㄹㅅ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ㄹㅌ	1.000000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ㄹㅎ	1.000000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ㄹㅎ	.987854	-	-	-	.000450	-	-	-	-	-	-	-	-	-	-	-	.011709	-	-	-	-	.009897	-	-	-	-	-	-
ㅁ	.548474	.139375	-	.000314	.059201	-	.001520	.000410	.064569	.003282	-	-	-	-	-	.000024	.054862	.042235	.002751	.003771	.000024	.094703	.001641	.001303	-	.000531	.000372	.000097
ㅂ	.762686	.051357	-	-	.046635	.000065	-	.000032	.052715	.000032	-	-	-	-	-	-	.017205	.013066	.000065	.002539	-	.048996	.000065	.000065	-	.001649	.000129	-
ㅂㅅ	.999756	-	-	-	.000182	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.000061	-
ㅅ	.730634	.022429	.001958	-	.027854	-	-	.000783	.002634	.000112	-	.004027	-	-	-	.000056	.038089	.006936	.001308	.011354	.000559	.026176	.000783	.014297	.000168	.005369	.004195	.000280
ㅆ	.999802	-	-	-	.000198	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ㅇ	.494966	.105717	.000052	-	.127898	-	-	.000132	.066256	.000361	-	.000007	-	-	-	-	.045481	.029305	.000177	.002973	.000015	.124983	.000191	.000499	.000029	.000383	.000861	.000015
ㅈ	.923512	.006458	-	-	.034929	-	-	.000162	.001455	-	-	-	-	-	-	-	.013422	.002264	.000323	-	.000323	.009379	.001294	.000808	-	.000372	-	.002587
ㅊ	.644917	.007587	-	-	.034143	-	-	.000305	.185888	.000759	-	-	-	-	-	-	.003794	.013657	-	-	-	.036419	-	.026832	-	.018209	.022762	-
ㅋ	.555556	-	-	-	.444444	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ㅌ	.941348	.001624	-	-	.006224	.000068	.000068	-	.027763	.001150	-	.000135	-	-	-	-	.002368	.009809	.000609	.009135	-	.000385	-	.000203	-	.000180	.000676	.000406
ㅍ	.933348	.008170	-	-	.014621	-	-	.000215	.014836	-	-	.000430	-	-	-	-	.003440	.007095	-	-	-	.016986	.000430	-	-	.000430	-	-
ㅎ	.994674	-	-	-	.002292	-	-	-	-	-	-	-	-	-	-	-	.002697	-	-	.000337	-	-	-	-	-	-	-	-

(c) 조건부 확률 $p(z_n|z_p)$

x_n / z_p	ㄱ	ㄲ	ㄴ	ㄷ	ㄸ	ㄹ	ㅁ	ㅂ	ㅃ	ㅅ	ㅆ	ㅇ	ㅈ	ㅉ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
공백소	.060405	.006686	.039945	.407827	.007172	.146813	.027484	.020228	.001979	.061177	.000909	.070538	.051990	.002264	.015491	.003563	.009460	.005509	.060560
ㄱ	.103994	.002534	.014554	.162846	.005744	.027938	.070222	.029952	.001312	.118977	.003366	.085075	.096068	.003444	.018400	.000715	.013878	.012475	.228507
ㄲ	.015640	.000948	.003318	.696682	-	-	-	.001422	-	.017062	-	.264929	-	-	-	-	-	-	-
ㄱㅅ	.050000	-	-	.350000	-	-	.050000	-	-	-	-	.150000	.350000	-	-	-	-	-	.050000
ㄴ	.111168	.003434	.026364	.103583	.005659	.042610	.046467	.037328	.000731	.123264	.001001	.067141	.166651	.006445	.036165	.006568	.014260	.014291	.186871
ㄴㅈ	-	-	-	.953906	-	-	-	-	-	-	-	.009704	-	-	-	-	-	-	.036390
ㄴㅎ	.000518	-	-	.824103	-	-	-	-	-	-	-	.175380	-	-	-	-	-	-	-
ㄷ	.018215	.000328	.000985	.887430	.000738	-	.000410	.002379	.000246	.004102	.000328	.045372	.003938	-	.025107	-	.000246	.000738	.009436
ㄹ	.092390	.006555	.006928	.269512	.005135	.125402	.041749	.027933	.001777	.045449	.025053	.150073	.044832	.007650	.018952	.006060	.006442	.022018	.096090
ㄹㄱ	.023283	-	.001940	.744664	.005821	-	.005821	.016686	-	.005433	-	.077610	.006985	-	.000388	-	.004657	.000388	.106325
ㄹㅁ	.330846	-	-	.575871	-	-	-	.004975	-	.000622	-	.075871	.011816	-	-	-	-	-	-
ㄹㅂ	.001184	-	-	.696270	-	-	-	-	-	.000592	-	.253996	.019538	.001184	-	-	-	-	.027235
ㄹㅅ																			
ㄹㅂ	-	-	-	1.000000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ㄹㅍ	-	-	-	.924528	-	-	-	-	-	-	-	-	.075472	-	-	-	-	-	-
ㄹㅎ	-	-	-	.798021	-	.016644	-	-	-	-	-	.161044	.024291	-	-	-	-	-	-
ㅁ	.099891	.019476	.023314	.153855	.002220	.028961	.031978	.050778	.006806	.142343	.010909	.075492	.136986	.014094	.038325	.008375	.014070	.022614	.119512
ㅂ	.055820	.000453	.007341	.481388	.001876	.024999	.012677	.018725	-	.082274	.006177	.059830	.079137	.000776	.020277	.000032	.002102	.004657	.141457
ㅂㅅ	-	-	.000406	.955639	-	-	-	-	-	-	.000244	.042330	.001300	-	-	-	-	-	.000081
ㅅ	.084345	.002685	.041837	.294088	.002797	-	.050003	.069299	.001566	.025840	.000895	.088204	.044186	.003580	.002125	.000447	.003132	.001063	.283908
ㅆ	-	-	.000066	.999901	-	-	-	-	-	-	-	-	.000033	-	-	-	-	-	-
ㅇ	.160507	.001722	.025493	.045348	.000854	.044678	.042986	.057425	.000640	.126558	.000361	.124947	.092440	.001575	.038894	.003238	.011944	.014020	.206370
ㅈ	.008571	.004366	.004366	.715071	.000647	.001617	.006630	.010026	-	.029107	.000485	.079722	.013098	-	.075841	-	.003234	-	.047219
ㅊ	.227618	-	.125190	.416540	.002276	-	.016692	.080425	-	.031866	.014416	.069803	.004552	-	.003035	-	-	.006829	.000759
ㅋ	.333333	-	-	.444444	-	-	.111111	.111111	-	-	-	-	-	-	-	-	-	-	-
ㅌ	.017115	.003991	.028819	.704641	.000338	.000068	.007983	.004397	.000609	.003991	.000609	.210729	.012042	.000203	.002638	.000271	-	.000338	.001218
ㅍ	.017201	.000645	.043217	.687594	.008385	-	.011825	.004945	.000430	.030746	.000215	.157386	.025586	.000215	.007525	-	-	.002150	.001935
ㅎ	.002899	-	-	.968786	-	-	-	-	-	.002832	-	.021371	.001146	-	.002832	-	-	-	.000130

(d) 조건부 확률 $p(x_n|z_p)$

다 크다. 조건부 확률p(ㄴ|ㄴ)는 종성의 발생확률 p(ㄴ)보다 크고 조건부 확률p(공백소|ㄴ)는 종성의 발생확률 p(공백소)보다 작다. 조건부 확률 p(공백소|공백소)는 종성의 발생확률 p(공백소)보다 크고 조건부 확률p(ㄴ|공백소)와 p(ㄹ|공백소)는 각각 종성의 발생확률 p(ㄴ)와 p(ㄹ)보다 작다. 조건부 확률 p(공백소|ㄹ)는 종성의 발생확률 p(공백소) 보다 크고 조건부 확률 p(ㄹ|ㄹ)는 종성의 발생확률p(ㄹ) 보다 작다.

앞음절 종성이 Z_p일 때 바로 다음 음절의 초성이 X_n일 조건부 확률 $p(x_n|z_p)$를 표10(d)에 보인다. 조건부 확률 p(ㄷ|공백소)와 p(ㄹ|공백소)각각 초성의 발생확률 p(ㄷ)와 p(ㄹ)보다 크고 조건부 확률 p(ㄱ|공백소)와 p(ㅇ|공백소)는 각각 초성의 발생확률 p(ㄱ)와 p(ㅇ)보다 작다. 조건부 확률 p(ㅈ|ㄴ)와 p(ㅎ|ㄴ)는 각각 초성의 발생확률 p(ㅈ)와 p(ㅎ)보다 크고 조건부 확률 p(ㄷ|ㄴ)와 p(ㅇ|ㄴ)는 각각 초성의 발생확률 p(ㄷ)와 p(ㅇ)보다 작다. 조건부 확률 p(ㄷ|ㅎ)는 0.97보다 크다.

Ⅳ. 결 론

한국어 다음절 단어에 있어서 빈도수와 종류에 따른 음절의 길이 분포를 계산하였다. 한국어 단어는 빈도수를 고려했을 때 1음절 단어의 발생확률이 가장 높고 단어의 종류로는 2음절 단어가 가장 많다. 전체 음절의 초성, 중성, 종성 단위의 조건부 확률을 계산하였다. 다음절 단어의 길이와 단어내에서의 음절의 위치에 따른 초성, 중성, 종성 단위의 발생확률과 조건부 확률을 계산하였다. 한국어 단어내에서 인접음절의 초성－초성, 중성－중성, 종성－종성, 종성－초성 사이의 음절간 조건부 확률을 계산하였다. 이 결과로부터 한국어 단어에는 음절간의 상관관계가 있음을 알 수 있다.

1956년 문교부에서 조사한 "우리말 말수 사용의 잦기 조사"를 한국어 단어의 빈도분포로 사용하였는데 한국어 단어의 발생빈도가 30년동안 근본적으로는 변하지 않았다는 점을 고려하면 이 논문에서 계산된 한국어 다음절 단어의 초성, 중성, 종성 단위의 음절간 조건부 확률은 한국어 음성인식 및 합성, 자연언어처리, 암호법, 언어학, 음성학, 한글부호 표준화 연구등에 이용될 것으로 기대된다.

參 考 文 獻

[1] C. Shannon, "Prediction and entropy of printed English," *Bell Syst., Tech. J.*, vol. 30, pp. 50-64, Jan. 1951.

[2] 이재홍, 오상현, "한글 음절의 초성, 중성, 종성 단위의 발생확률, 엔트로피, 평균상호정보량", 전자공학회논문지, 제26권, 제 9 호, pp. 1－9, 1989년 9월.

[3] 이주근, 최홍문, "학국어 음절의 entropy에 관한 연구", 전자공학회지, 제11권, 제3호, pp. 15-21, 1974년 6월.

[4] 안수길, 안지환, "공백소를 포함한 한글 자소발생 확률과 엔트로피", 전자공학회지, 제17권, 제2호, pp. 23－28, 1980년 4 월.

[5] 남궁건, 한글 낱말의 발생빈도 분포의 엔트로피에 관한 연구. 석사학위논문, 서울대학교, 1979.

[6] 과학기술처, 컴퓨터 관련 표준화 규격. 과학기술처, 1982.

[7] 한국표준연구소, 한글 정보처리 표준화연구, 과학기술처, 1986.

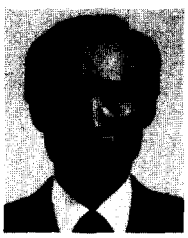
[8] 한국표준연구소, 한자부호 표준시안 작성을 위한 연구. 과학기술처, 1987.

[9] 문교부, 우리말 말수 사용의 잦기 조사. 문교부 1956.

[10] 정우상, "국민학교 교과서 어휘 연구", 국어연구소 연구보고서 제1집, pp. 651－872, 1987.

[11] R. Manprino, "Printed Portuguese entropy-statistical calculation," *IEEE Trans. Inform, Theory*, vol. IT-16, p.122, Jan. 1970.

著 者 紹 介



李 在 弘(正會員)

1953年 12月 7日生. 1976年 2月 서울대학교 전자공학과 졸업. 공학사학위취득. 1978年 2月 서울대학교 대학원 전자공학과 졸업, 공학석사학위 취득. 1986年 8月 미국 미시간 대학교 전기공학과 및 컴퓨터 과학과 졸업. 공학박사학위 취득. 1978年～1981年 해군사관학교 전자공학과 교관, 전임강사. 1987年 2月～현재 서울대학교 전자공학과 조교수, 부교수. 1991年 1月～현재 AT & T 벨 연구소 방문교수. 주관심분야는 채널코우딩, 확산대역 시스템, 디지탈 통신이론, 위성통신, 이동통신, 음성합성 등임.

李 在 鶴(學生會員)

1967年 5月 1日生. 1989年 2月 서울대학교 전자공학과 졸업. 공학사학위 취득. 1991年 2年 서울대학교 대학원 전자공학과 졸업. 공학석사학위 취득. 1991年 3月～현재 서울대학교 대학원 전자공학과 대학원 박사과정 재학중. 주관심분야는 디지탈 통신이론, 채널코우딩, 이동통신, 위성통신 등임.