Onset-to-Onset Probability and Gradient Acceptability in Korean*

Hahn Koo and Young-il Oh
(University of Illinois at Urbana-Champaign)


Bi-phone probability, or the relative frequency with which two segmentally adjacent phonemes co-occur in a language, has effectively explained the influence of speakers' sensitivity to the phonotactic probability of sound sequences on lexical processing. In this paper, we argue that speakers are also sensitive to onset-to-onset probability in their native language, and therefore bi-phone probability alone is not a sufficient estimate of sequential probability since onset-to-onset probability reflects how often two onsets that are segmentally non-adjacent co-occur in the language. To support our argument, we present an experiment which shows that adult Korean speakers are sensitive to onset-to-onset probability in their language and that their sensitivity is manifested in their gradient acceptability judgment of non-words.

Keywords: phonotactic probability, gradient acceptability, Korean

1. Introduction

In classical generative phonology, phonotactic constraints define the set of possible sound patterns in a given language by restricting the position of phonological segments and sequences of segments. Speakers' sensitivity to the phonotactic constraints in their language is manifested in various linguistic tasks they perform. Speakers will judge a sound pattern to be grammatical if it is phonotactically legal and ungrammatical if it is phonotactically illegal. Their sensitivity is also manifested in their performance in on-line speech processing tasks, such as speech perception and production. For example, speakers perceive non-words with phonotactically legal onset clusters more accurately than non-words beginning with illegal onset clusters (Brown and Hildum 1956). Speakers rarely produce speech errors that violate the phonotactic constraints in their language (Fromkin 1971, Stemberger 1982).

More recent studies show that speakers are not only sensitive to the phonotactic legality of sound patterns but also to the frequency with which the sound

* We thank Jennifer Cole and the three anonymous reviewers for their corrections and helpful comments. Of course, all remaining faults are ours.
patterns occur in the language. Speakers judge non-words consisting of more frequent sound patterns to be more acceptable than non-words consisting of less frequent sound patterns (Coleman and Pierrehumbert 1997, Treiman et al. 2000). Speakers perceive non-words with high-frequency syllables more quickly than non-words with low frequency syllables (Vitevitch et al. 1997). When speakers produce speech errors, they tend to replace less common phonemes or phoneme sequences by more common phonemes or sequences (Motley and Baars 1975). Speakers produce words faster in picture naming tasks if the constituent phonemes occur in more probable positions and sequences (Vitevitch et al. 2004).

The frequency with which sub-lexical sound patterns are observed in a language is called phonotactic probability. As phonotactic constraints restrict what are possible sound positions and sound sequences in a given language, phonotactic probability encodes the frequency with which sounds occur in certain positions and sequences in the language (Jusczyk et al. 1994, Vitevitch and Luce 2004). There are many ways to represent sound positions and sound sequences, and there are many positions and sequences to consider given a particular representational framework. In other words, we are faced with the problem of how to define phonotactic probability, and a psycholinguistically effective definition should only include frequency of sound patterns to which speakers are sensitive. One way to do this is to first define phonotactic probability in a particular way and examine whether speakers process sound patterns differently as a function of the phonotactic probability of the sound patterns in the language. If their behavior does differ as a function of the phonotactic probability, that particular definition of phonotactic probability is shown to be psycholinguistically effective and adopted in successive studies on the effect of phonotactic probability on speech processing.

However, the definition of phonotactic probability commonly adopted in studies on language processing is based on several assumptions open to challenge with the discovery of recent experimental results. For example, it is commonly assumed that phonotactic probability is computed over segmental representation. However, Goldrick (2004) suggests that feature representation of sound patterns may also be necessary. Subjects in his study learned two phonotactic constraints: a categorical segmental constraint restricting the position of a specific segment, and a gradient featural constraint restricting the position of segments that shared a phonological feature. When a segment was restricted to a position by both the segmental constraint and the featural constraint, its tendency to stick to the restricted position in speech errors was strong. But when the two constraints were contradictory, its tendency to stick to the position restricted by the segmental constraint became weaker.

The assumption of particular interest to the present study is the one regarding the type of sound sequences over which phonotactic probability is com-
The common practice is to estimate the probability of sound sequences in terms of what is called *bi-phone probability*, the conditional probability of observing a phoneme given the preceding phoneme. Often in conjunction with *positional probability*, or the probability with which a phoneme occurs in a word position, the use of bi-phone probability has been successful in explaining the frequency effect on various linguistic tasks that speakers perform (Jusczyk et al. 1994, Vitevitch et al. 1997, Vitevitch and Luce 1998, Bailey and Hahn 2001, Leigh and Charles-Luce 2002, Vitevitch et al. 2004, Storkel et al. 2006). But is bi-phone probability alone enough to capture the effect of frequency of sound sequences on lexical processing? By definition, the exclusive use of bi-phone probability limits the locus of effective sequential statistical regularity to two adjacent segments. However, phonological dependencies between non-adjacent segments such as consonant harmony and vowel harmony do exist in some languages, and speakers of such languages are indeed sensitive to the non-adjacent phonological dependencies.¹

For example, speakers of Semitic languages are sensitive to the co-occurrence restrictions of their language on consonants that are root-adjacent but not string-adjacent (Berent and Shimron 1997, Frisch and Zawaydeh 2001). Root morphemes in Semitic languages are typically a sequence of three consonants whose co-occurrence is restricted in various ways. For example, repetition of first two consonants is prohibited while repetition of final two consonants is acceptable, henceforth Obligatory Contour Principle (OCP) following McCarthy (1986). Repetition of homorganic consonants is also prohibited, henceforth OCP-Place following McCarthy (1988). Berent and Shimron (1997) shows that when asked to rate the acceptability of artificially created verbs, Hebrew speakers rate the verbs that respect OCP to be more acceptable than those that violate OCP. Similarly, Frisch and Zawaydeh (2001) shows that Jordanian Arabic speakers rate artificially created verbs of the form $C_1aC_2aC_3a$ that respect OCP-Place to be more acceptable than those that violate OCP-Place.

In sum, the speakers of Semitic languages are sensitive to the co-occurrence pattern of segmentally non-adjacent consonants, a phonological dependency which is beyond the scope of bi-phone probability. So for a better statistical account of the effect of speakers' sensitivity to the sequential frequency in their language on lexical processing, one would have to deploy some measure of

¹ By "non-adjacent", we mean the phonemes are segmentally non-adjacent, rather than being non-adjacent within a tier of their own. For example, the root consonants (e.g., /k/, /t/, /b/) in Arabic words (e.g., /kutib/) are adjacent to each other when we list them in a tier separate from a tier consisting solely of intervening vowels (e.g., /u/, /i/). However, when we represent a word as a sequence of the constituent phonemes, the consonants are non-adjacent in that they are separated from each other by the intervening vowels. We follow the latter use of the term adjacency/ non-adjacency in this paper.
how often the phonologically dependent non-adjacent segments co-occur in the language. But what about languages which are not known to have phonological dependencies between non-adjacent segments? Is there a priori reason to believe that the co-occurrence frequency of non-adjacent segments will affect lexical processing even in such languages? Recent phonotactic learning studies using artificially created languages seem to suggest that speakers' sensitivity to the distribution of co-occurrence of non-adjacent consonants may be a more general one and may even exist in languages without templatic morphology, consonant harmony, or consonant disharmony.

For example, in Newport and Aslin (2004) and Bonatti et al. (2005), adult speakers of English (Newport and Aslin 2004) and French (Bonatti et al. 2005) listened to a continuous stream of tri-syllabic sequences which exhibited the following statistical regularity: the transitional probability between onset consonants within a word was 1.0, while the transitional probability between onset consonants spanning word boundary was 0.5. After listening to the stream for approximately 20 minutes, the subjects could correctly identify tri-syllabic sequences that formed a word from tri-syllabic sequences that were concatenation of syllables from two words. In H Koo and Cole (2006), adult English speakers first listened to and repeated a set of tri-syllabic non-words that exhibited liquid-harmony; the vast majority of non-words had either an /l/-/l/ or an /r/-/r/ onset sequence. After encountering approximately 70 such items, they perceived novel words following liquid-harmony more quickly than novel words following liquid-disharmony. In addition, the subjects judged liquid-harmony words to be more grammatical than liquid-disharmony words.

Neither English nor French has phonological dependencies between non-adjacent onset consonants, so it is unlikely that the experimental results reflect the subjects' knowledge of their own language. Therefore, the subjects must have learned the non-adjacent phonological dependencies embedded in the words of the artificially created languages. Within the statistical language learning framework (e.g., Saffran et al. 1996), they must have tracked the frequency with which two onsets co-occur within a word, henceforth onset-to-onset probability, and that their acquired sensitivity to the onset-to-onset probability affected how they process new sound patterns. If this is indeed how speakers learn the phonotactic probability of their language, we would expect speakers to be sensitive to co-occurrence frequency of non-adjacent phonemes such as onset-to-onset probability of their language, regardless of whether the language is known for non-adjacent phonological dependencies or not. As a consequence, one could argue that we must reconsider the commonly adopted definition of phonotactic probability and extend the types of sequential phonotactic probability to include onset-to-onset probability as well as bi-phone probability.

However, speakers' ability to learn the phonotactic probability of a sound
pattern in an artificial language may not directly translate to their ability to learn the same phonotactic probability and maintain their sensitivity in their native language. Data in natural language is much noisier and therefore the pattern may not be as salient as it is in artificial language. In addition, there may be too many other salient patterns in their native language for the speaker to pay attention to a particular pattern. Therefore, to further support an argument based on the results of the artificial language experiments, one must provide parallel experimental results using natural language.

As a response, the question we address in this paper is the following: the phonotactic learning studies show that speakers can acquire sensitivity to onset-to-onset probability embedded in artificial languages, but are they also sensitive to the same probability in their native language so that it affects their lexical processing behavior? We argue that speakers are indeed sensitive to onset-to-onset probability in their native language, and therefore we must also consider co-occurrence frequencies of non-adjacent phonemes, such as onset-to-onset probability, in computing the phonotactic probability of a sound sequence, as suggested by the phonotactic learning studies. To support our argument, we present an experiment which shows that adult Korean speakers are sensitive to onset-to-onset probability in their native language and that their sensitivity to the probability is manifested in their gradient acceptability judgment of non-words.

The paper is organized as follows. In section 2, we first review a few measures of lexical statistics which could potentially affect speakers' acceptability judgment. In the experiment whose methods are described in section 3, adult Korean subjects were asked to rate the acceptability of two groups of non-words. Non-words in one group had higher onset-to-onset probability than non-words in the other group, but they did not differ in terms of other lexical statistics reviewed in section 2. The result summarized in section 4 shows that they rated the non-words with high onset-to-onset probability to be more acceptable than the non-words with low onset-to-onset probability. We summarize and conclude the paper in section 5.

2. Lexical Statistics and Gradient Acceptability

Various measures of lexical statistics have been proposed to account for the effect of frequency of sound patterns on speakers' linguistic behavior. Some are specific types of phonotactic probability such as positional probabilities and bi-phone probability that specialize in encoding the frequency information regarding particular types of sub-lexical sound patterns. Another measure is neighborhood density which measures the number of words that are phonologically similar to a given word. The idea is that what appears to be speakers'
sensitivity to the frequency of a sound pattern in their language is a by-product of simultaneously accessing all words in the language that share the sound pattern. For example, speakers judge non-words with more frequent constituent sound patterns to be more acceptable because there are more words in the language that share the sound patterns and sound similar to the non-words.

In this section, we briefly review three measures of lexical statistics that could potentially affect speakers' acceptability judgment and therefore must be controlled for in the experiment discussed below: neighborhood density, positional probability, and bi-phone probability. Studies that we cite in this section below show that these measures correlate relatively well with acceptability judgment ratings collected from experiments. In addition, they have been successful in explaining the effect of lexical statistics on other speech processing domains such as perception and production. As different studies define these measures somewhat differently, we try to focus on the underlying basic idea and give our own definition of these measures that we followed in our experiment.

2.1. Neighborhood Density

In general, neighbors of a word refer to the set of words in the speaker's mental lexicon that sound similar to the given word. The neighborhood density of a word measures the number of neighbors of the given word. Studies differ with respect to how they measure similarity and how they count the number of neighbors. The most common practice (e.g., Luce 1986) is to follow the notion of minimal edit distance in Kruskal (1983) and measure the similarity between two words in terms of the number of phoneme edit-operations (substitution, insertion, or deletion) required to derive one word from the other. Two words are neighbors if one can be derived from the other with a single edit-operation. For example, the neighbors of pat in English would be words such as pan (via substitution), spat (via insertion), and at (via deletion). Neighborhood density can be measured by simply counting the number of such neighbors, but to account for the token-frequency effect, the number of neighbors is often weighted by log token-frequency of the individual neighbors. This simple definition of neighborhood density has been useful in explaining the frequency effect found in many studies on lexical processing (e.g., Charles-Luce and Luce 1990, Vitevitch and Luce 1998).

Bailey and Hahn (2001) proposes a significantly more refined neighborhood density measure to explain speakers' acceptability judgment of non-words. Their Generalized Neighborhood Model differs from the commonly used measure of neighborhood density primarily in the following aspects. Firstly, substitution costs less if the substituted phonemes are phonologically more similar. Specifically, the substitution cost between two phonemes is one minus
their phonological similarity as measured in Frisch et al. (2004). Therefore, *bat* and *pat* are considered closer to each other than *bat* and *cat*. Secondly, all words in the mental lexicon are considered to be neighbors of different perceived similarity, rather than ignoring words whose minimal edit-distance is beyond a threshold. To measure the perceived similarity of two words on a continuous scale, they adopted the exponential version of the Generalized Context Model (Nosofsky 1986). To account for the token-frequency effect, especially in a potentially non-monotonic way as observed in inflectional morphology (e.g., Bybee 1995), they added a quadratic frequency weighting term to the original similarity equation of the exponential Generalized Context Model. As a result, the neighborhood density of a word \( i \), \( S_i \), is computed in this model according to (1), where \( f_j \) is the log token-frequency of a neighbor \( j \), and \( d_{ij} \) is the edit-distance between \( i \) and \( j \). All parameters written in upper-case are free parameters whose best-fitting values are determined by regression.

\[
S_i = \sum_j (A \cdot f_j^2 + B \cdot f_j + C) \cdot e^{-D \cdot d_{ij}}
\]

Bailey and Hahn (2001) and Albright (2006) show that the generalized neighborhood model explains the speakers' acceptability judgment better than the simple measure of neighborhood density. Despite its superior performance, one disadvantage of the model is that its free parameters require the user to identify the best-fitting values from the existing acceptability rating data before the model is put to use. It may be an excellent explanatory model, but it may not be efficient for controlling stimuli for experimental design when there is no previous data.

2.2. Phonotactic Probability

Phonotactic probability seeks to capture frequency information regarding two types of sound patterns: which sounds appear in which positions, and which sounds appear next to which. We will refer to them in this section as positional probability and sequential probability, respectively.

2.2.1. Positional Probability

Different studies define positional probability differently depending on how they define the term “position”. In Coleman and Pierrehumbert (1997), and similarly in Frisch et al. (2000), a position is specified along three dimensions: onset vs. rhyme, word-initial syllable vs. word-final syllable, and stressed vs. unstressed. For example, the position of /p/ in /karpat/ is specified as onset in an unstressed word-final syllable. The positional probability of a sound in a given position is the ratio of the frequency of that sound occurring in the given position to the frequency of any sound occurring in the given position, where
the frequency is counted in terms of type-frequency. The positional probability of a word is equal to the product of the individual positional probabilities over all word positions.

On the other hand, Vitevitch and Luce (2004) number each segmental slot in a word from left to right and specify the position in terms of its index. For example, the position of /p/ in /karpat/ is specified as the fourth position of the word. The positional probability of a sound in a given position is computed in the same way as in Coleman and Pierrehumbert (1997) except that the frequency is counted in terms of log token-frequency instead of type-frequency. The positional probability of a word is the sum of the individual positional probabilities over all word positions.

2.2.2. Sequential Probability

As mentioned in the introductory section, sequential probability is most commonly captured in terms of bi-phone probability (Jusczyk et al. 1994, Vitevitch et al. 1997, Vitevitch and Luce 1998, Bailey and Hahn 2001, Leigh and Charles-Luce 2002, Vitevitch et al. 2004, Storkel et al. 2006). To compute bi-phone probabilities for a given word, we first add the null-symbol at the word boundaries and extract all substrings consisting of two phonemes, or bi-phones. For example, given the word /keet/ , we extract the four bi-phones {#k, kee, eet, t#}. The bi-phone probability of 'kee' is the conditional probability of 'ee' given 'k'. That is, it is the ratio of frequency of 'kee' to the sum of frequencies over all bi-phones beginning with 'k', where the frequency is the log token-frequency. The positional bi-phone probability in Vitevitch and Luce (2004) is slightly different in that bi-phones in different word positions are counted separately. For example, the bi-phone 'kee' in /keet/ spans the first and second word positions, and therefore it is different from the bi-phone 'kee' in /skeet~r/ as it spans the second and third word positions. The bi-phone probability of a word is usually the mean of the individual bi-phone probabilities (e.g., Vitevitch et al. 1997, Vitevitch and Luce 1998, Bailey and Hahn 2001), but Vitevitch and Luce (2004) uses the sum of the individual bi-phone probabilities.

2.3. Measures of Lexical Statistics Controlled for in Our Experiment

To summarize, measures of lexical statistics that may affect speakers' acceptability judgment include neighborhood density, positional probability, and bi-phone probability. However, we also saw that there are different ways to define and compute them. Rather than trying all different versions of the same statistical measure, we adopted the basic idea underlying the measures and used the following definitions for the present study. Firstly, the neighborhood density of a word was defined as the log token-frequency of words that could be derived by a single edit-operation. Despite its superior performance (Bailey
and Hahn 2001, Albright 2006), we did not use the Generalized Neighborhood Model to compute neighborhood density because the free parameters of the model must be fit using the results from other acceptability experiments, access to which we do not have at this point.

Secondly, positional probability was defined as the probability of observing the phoneme in a given syllable position, where a syllable position was onset, nucleus, or coda. For example, the positional probability of observing a phoneme X as an onset was equal to the rate of token frequency of syllables whose onset is X to the sum of token frequencies over all syllables, and likewise for nucleus and coda. The positional probability of a word was the mean of positional probabilities over all syllable positions in the word.

Finally, the bi-phone probability of a two phoneme sequence XY was defined as the conditional probability of Y given X. To compute the bi-phone probability of XY, we divided the token frequency of XY by the sum of token frequencies over all bi-phones beginning with X. The bi-phone probability of a word was the mean of bi-phone probabilities over all bi-phones in the word.

2.4. Onset-to-Onset Probability

Recall that the main argument in the present study is that speakers are sensitive to onset-to-onset probability in their native language in addition to the lexical statistics above, and therefore the length of the sound sequence characterized by phonotactic probability should be extended to include at least two onsets in adjacent syllables. Following Newport and Aslin (2004) and Bonatti et al. (2005), we defined onset-to-onset probability as the conditional probability of observing an onset consonant given the onset consonant in the preceding syllable. That is, the onset-to-onset probability of an onset consonant X given the onset consonant Y in the preceding syllable was computed as in (2), where \( \text{onset}_i = X \) denotes that X occupies the onset of the \( i \)th syllable.

\[
(2) \quad P(\text{onset}_i = X \mid \text{onset}_{i-1} = Y) = \frac{\text{token frequency of bisyllables with its first onset } Y \text{ and second onset } X}{\text{token frequency of bisyllables whose first onset is } Y}
\]

In the experiment discussed below, subjects were asked to rate the acceptability of two groups of non-words that differed in their onset-to-onset probability but not in the other three lexical statistics: neighborhood density, positional probability, and bi-phone probability. Our hypothesis is that the subjects will rate non-words with higher onset-to-onset probability to be more acceptable than non-words with lower onset-to-onset probability, as the phonotactic learning studies conducted in artificial languages would suggest (Newport and Aslin 2004, Bonatti et al. 2005, H Koo and Cole 2006). The following section de-
scribes the methods of the experiment.

3. Methods

3.1. Subjects

Twenty subjects from the University of Illinois at Urbana-Champaign community volunteered for the experiment. All subjects were adult native speakers of Korean and reported no history of a hearing or speech disorder at the time of participation.

3.2. Materials

The key manipulation involved selecting two groups of 20 bi-syllabic non-words which differed in their onset-to-onset probability but not in the other lexical statistics mentioned in section 2. To this end, we consulted the result of Research on Usage Frequency in Contemporary Korean conducted by the National Institute of the Korean Language in 2003, henceforth NIKL database. The NIKL database lists 58437 Korean words with their part-of-speech tag and token frequency, collected from 176 documents. We transcribed the pronunciation in X-SAMPA (Wells 2000) and syllabified all words in the database. The resulting frequency-annotated Korean pronunciation dictionary was used to compute the lexical statistics including the onset-to-onset probability.

We generated 40 bi-syllabic non-words of the form $C_1V_1C_2V_2C_3$. The first consonant ($C_1$) was /s/ for all stimuli while the second consonant ($C_2$) was chosen from {/k/, /r/, /p/, /tʃ/}, resulting in four different onset-types. Onset types of /s/-/k/ and /s/-/r/ had higher onset-to-onset probabilities than onset types of /s/-/p/ and /s/-/tʃ/. Each onset-type had ten members depending on how the remaining positions ($V_1$, $V_2$, $C_3$) were filled. To eliminate the potential influence of vowels and rhymes on subjects' judgment, the remaining positions were filled symmetrically for the four onset-types. For example, there were four versions of /sa.Cem/: [sa.kem], [sa.rem], [sa.pem], and [sa.tʃem]. The full list of stimuli is given in Appendix A.

The 40 non-words thus generated were divided into two groups: words with high onset-to-onset probability (H-words), and words with low onset-to-onset probability (L-words). While the two groups differed in their onset-to-onset probability, they did not differ in the other three lexical statistics defined above: positional probability, bi-phone probability, and neighborhood density. Mean lexical statistics for the four onset-types are summarized in Table 1.
Table 1. Summary of mean lexical statistics for the four onset-types

<table>
<thead>
<tr>
<th>Onset-type</th>
<th>Onset-to-onset probability</th>
<th>Positional probability</th>
<th>Bi-phone probability</th>
<th>Neighborhood density</th>
</tr>
</thead>
<tbody>
<tr>
<td>/s/-/k/</td>
<td>0.1478</td>
<td>0.1858</td>
<td>0.0051</td>
<td>1.0256</td>
</tr>
<tr>
<td>/s/-/r/</td>
<td>0.1294</td>
<td>0.1758</td>
<td>0.0045</td>
<td>1.3260</td>
</tr>
<tr>
<td>/s/-/p/</td>
<td>0.0317</td>
<td>0.1744</td>
<td>0.0040</td>
<td>0.9550</td>
</tr>
<tr>
<td>/s/-/tʃʰ/</td>
<td>0.0285</td>
<td>0.1713</td>
<td>0.0037</td>
<td>1.0958</td>
</tr>
</tbody>
</table>

As summarized in Table 1, onset-to-onset probability was high for the two onset-types /s/-/k/ and /s/-/r/, while it was low for the two onset-types /s/-/p/ and /s/-/tʃʰ/. However, there was no significant difference between the four onset-types in positional probability ($F(3,36) = 0.714, p = .550$), bi-phone probability ($F(3,36) = 0.419, p = .740$), or neighborhood density ($F(3,36) = 0.269, p = .847$). Therefore, the twenty words of the two onset-types /s/-/k/ and /s/-/r/ belonged to the H-words, while the other twenty words of the two onset-types /s/-/p/ and /s/-/tʃʰ/ belonged to the L-words. Onset-to-onset probability was significantly higher for H-words than for L-words ($F(1,38) = 2555.775, p < .0001$). However, there was no significant difference between H-words and L-words in positional probability ($F(1,38) = 1.188, p = .283$), bi-phone probability ($F(1,38) = 1.097, p = .302$), or neighborhood density ($F(1,38) = 0.244, p = .624$).

The words were produced in a sound-attenuated booth by a male Korean speaker of Seoul dialect, while the session was recorded at 44.1 KHz sampling rate and with 16 bit resolution. Individual words were extracted from the session recording and stored as separate .WAV files.

3.3. Procedures

Subjects were tested individually. Each subject was seated in front of a computer placed in a sound-attenuated booth. At the beginning of the session, the investigator told the subjects in Korean that they would hear a set of non-words one by one and that their task was to rate on a five-point scale how likely it would be that each stimulus could be a new word in Korean.

A typical trial began with a five-point scale displayed on the computer monitor in front of the subjects: with 1 labeled “very unlikely” and 5 labeled “very likely”. One of the stimulus items was presented to subjects through headphones at a comfortable listening level. Subjects were then asked to rate the acceptability of the presented stimulus by pressing the corresponding number key on the keyboard. Subjects were asked to click the mouse to begin the next trial. Prior to experimental trials, subjects had four practice trials on [satʃʰel], [soret], [sipup], and [sukiʃ], respectively. The purpose of these trials was to
familiarize the subjects with the task, so they did not count in the final data analysis. Stimuli presentation and response collection was controlled using the E-prime software throughout the session.

Stimuli were randomly ordered with the following four restrictions: (1) the stimuli formed a sequence of ten groups of four, with the four members of each group from four different onset types, (2) stimuli of the same onset type did not appear next to each other, (3) stimuli with the same vowel in the first syllable did not appear next to each other, and (4) stimuli with the same rhyme in the second syllable did not appear next to each other. An example ordered list of stimuli is given in Appendix B.

4. Results and Discussion

H-words consisted of non-words whose mean onset-to-onset probability was significantly higher than non-words that constituted L-words. In addition, H-words consisted of non-words from the two onset-types /s/-/k/ and /s/-/r/, while L-words consisted of non-words from the two onset-types /s/-/p/ and /s/-/j/>. By grouping two onset-types into a single group, we assumed that we could ignore the difference in onset-to-onset probability between the two onset-types within each group. Accordingly, in conjunction with the commonly held assumption that speakers find sound patterns with higher phonotactic probability to be more acceptable, our hypothesis was two-fold: subjects would rate H-words to be more acceptable than L-words, and there would be no difference in acceptability rating between the two onset-types within each group.

To test our hypothesis, we first compared mean acceptability ratings between H-words and L-words, and then compared mean acceptability ratings between the two onset-types within each group. Mean acceptability ratings for H-words, L-words, and their respective constituent onset-types are summarized in Figure 1. Acceptability ratings for each non-word averaged over subjects are listed in Appendix A.

The results were consistent with our hypothesis. One-tailed paired t-test on mean ratings for H-words and L-words revealed that subjects considered H-words to be more acceptable than L-words ($t(19) = 2.794, p = .0058$). However, subjects did not rate one onset-type to be more acceptable than the other onset-type within each group; mean acceptability rating was different in neither /s/-/k/ vs. /s/-/r/ ($t(19) = 1.023, p = .319$) nor /s/-/p/ vs. /s/-/j/ ($t(19) = 1.042, p = .310$).

To assess the individual effect of lexical statistics on acceptability ratings, a post-hoc multiple regression analysis was conducted using the mean acceptability rating averaged over subjects per item as the dependent variable.
Neighborhood density, positional probability, bi-phone probability, and onset-to-onset probability of each item were entered as independent variables. The result of the analysis is summarized in Table 2.

Table 2. Multiple regression analysis with acceptability rating as the dependent variable. Independent variables were Neighborhood Density (ND), Positional Probability (POS), Bi-phone Probability (BI), and Onset-to-Onset probability (OTO)

<table>
<thead>
<tr>
<th>Variables</th>
<th>B</th>
<th>SE</th>
<th>Beta</th>
<th>t</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td>4.211</td>
<td>0.438</td>
<td>0.070</td>
<td>9.606</td>
<td>0.000</td>
</tr>
<tr>
<td>ND</td>
<td>0.046</td>
<td>0.051</td>
<td>0.070</td>
<td>0.909</td>
<td>0.370</td>
</tr>
<tr>
<td>POS</td>
<td>-14.841</td>
<td>2.876</td>
<td>-0.550</td>
<td>-5.161</td>
<td>0.000</td>
</tr>
<tr>
<td>BI</td>
<td>253.875</td>
<td>23.043</td>
<td>1.153</td>
<td>11.018</td>
<td>0.000</td>
</tr>
<tr>
<td>OTO</td>
<td>1.269</td>
<td>0.873</td>
<td>0.113</td>
<td>1.454</td>
<td>0.155</td>
</tr>
</tbody>
</table>

Bi-phone probability and positional probability were significant predictors of acceptability ratings, as previous studies on phonotactic probability mentioned in section 2.2 would suggest. However, the regression coefficient for positional probability was negative, implying that speakers rated a non-word to be more acceptable as a Korean word if its positional probability was lower. Despite the effect of neighborhood density on acceptability ratings suggested in Bailey and Hahn (2001), neighborhood density was not a significant predictor of acceptability rating in this model. Contrary to what we would expect from the result of the t-test above, onset-to-onset probability was not a significant predictor, perhaps due to the following two possibilities.

One possibility may be that lexical processing in general is influenced more by statistical regularity between adjacent linguistic units such as bi-phone probability than statistical regularity between non-adjacent linguistic units. As far as we know, this remains a hypothesis to be tested. However, there are studies which suggest that adult speakers are biased towards conditional probabil-
ity between adjacent linguistic units rather than conditional probability between non-adjacent linguistic units when it comes to artificial grammar learning. For example, Gomez (2002) studied whether adult speakers could learn dependency between non-adjacent elements embedded in three element strings (e.g., pel-wadim-jic) in four different conditions which differed in terms of the conditional probability between the adjacent elements (e.g., pel-wadim and wadim-jic). The conditional probability of the third element (e.g., jic) given the first element (e.g., pel) was fixed to 1.0 in all four conditions, while the conditional probabilities between the adjacent elements were 0.5, 0.167, 0.083, and 0.042, respectively. Evidence of learning was observed only when the adjacent conditional probability was the lowest of all (0.042), suggesting that the speakers may have focused on adjacent conditional probability by default and then shifted their focus onto non-adjacent conditional probability when the predictability of adjacent conditional probability became “sufficiently unreliable”.

Another possibility may be that the variability of onset-to-onset probability was too small compared with the other lexical statistics in our experiments. The subjects may have attempted to rate all 40 non-words as differently as possible and classify them into as many groups as possible. To this end, they may have relied more on the features of a stimulus with greater variability when rating its acceptability. Among the 40 non-words, onset-to-onset probability had only four different values, while the other lexical statistics varied to a greater extent; there were 40 different bi-phone probabilities, 38 different positional probabilities, and 15 different neighborhood densities. As a result, the subjects may have relied more on bi-phone probability and/or positional probability than onset-to-onset probability.

In brief, the results of the t-test show that Korean speakers are sensitive to the onset-to-onset probability in their native language and that their sensitivity was manifested in their gradient acceptability judgment of non-words. However, care must be taken since the result of the post-hoc multiple regression analysis suggests that while the effect of onset-to-onset probability on acceptability ratings may be present as shown by the t-test, it is relatively weak compared with the effect of bi-phone probability. In relation to our research question, and the notion of gradience in psycholinguistics and phonology, the results have the following implications.

Firstly, recall that our research question was whether speakers are indeed sensitive to the onset-to-onset probability in their native language as suggested by the phonotactic learning studies using artificially created languages. Newport and Aslin (2004) and Bonatti et al. (2005) show that adults can learn the onset-to-onset probability embedded in a continuous stream of syllables and utilize that probability for word segmentation. H Koo and Cole (2006) shows that adults can learn co-occurrence restrictions between onsets in two adjacent
Onset-to-Onset Probability and Gradient Acceptability in Korean

syllables and that their acquired sensitivity affects their perception and grammaticality judgment of novel sound patterns. If the artificial languages in these studies were good models of natural languages, we would expect the speakers to have learned the onset-to-onset probability in their native language so that their acquired sensitivity to the probability affects their linguistic behavior.

However, the artificial languages in these studies are unrealistically simple compared with natural languages. Their stimuli had phonologically simpler structure; they were mere sequences of CV syllables generated with small phoneme inventories. There was no meaning attached to the stimuli, so the subjects were encouraged to devote their attention to the formal aspect of the stimuli. The size of the vocabulary was tiny, so there were only a small number of sound patterns whose statistical distribution subjects had to track. In addition, the investigators carefully controlled the distribution of sound patterns, so any regularities related to a sound pattern had only a few exceptions, if any. Therefore, while the phonotactic learning studies suggest the possibility that speakers are sensitive to onset-to-onset probability in their language, their suggestion remained a hypothesis to be tested. Our results show that the hypothesis raised from these studies is valid; Korean speakers were indeed sensitive to onset-to-onset probability in their language.

Secondly, taken together with the results from the phonotactic learning studies and the studies on Semitic speakers' sensitivity to OCP effects (Berent and Shimron 1997, Frisch and Zawaydeh 2001), our results show that co-occurrence frequency of non-adjacent phonemes should also be included in estimating the phonotactic probability of novel words. Studies that explain the frequency effect found in adult speakers' behavior in terms of phonotactic probability commonly use bi-phone probability to encode the frequency of sound sequences (e.g., Vitevitch et al. 1997). Moreover, phonotactic probability is defined as consisting of positional probability and bi-phone probability in studies that attempt to tease apart the effect of phonotactic probability from the effect of neighborhood density on perception (Vitevitch and Luce 1998), production (Vitevitch et al. 2005) and adult word learning (Storkel et al. 2006). To differentiate the two effects, these studies examine the change in speakers' behavior towards non-words as the non-words orthogonally vary in their neighborhood density and phonotactic probability.

The problem, however, is that since phonotactic probability and neighborhood density may well be positively related (Vitevitch and Luce 1998, Landauer and Streeter 1973, Frauenfelder et al. 1993), the validity of how their values are orthogonally manipulated is determined by how well their respective definitions are empirically supported. For example, two non-words that do not differ in their neighborhood density may or may not differ in their phonotactic probability depending on whether you include their onset-to-onset probability or not. Ignoring onset-to-onset probability, while there is evidence that it
affects speakers' speech processing behavior, could lead to an orthogonal manipulation that is problematic and any results based on such orthogonal manipulation are open to criticism.

Our results show that capturing the statistical distribution governing sound sequences with bi-phone probability is not enough. H-words and L-words did not differ in their bi-phone probability. In fact, none of the lexical statistics commonly assumed to affect speech processing can explain why our subjects rated H-words to be more acceptable than L-words. This suggests that we may have to consider the statistical distribution of a wider range of sound patterns than currently assumed. In particular, considering the co-occurrence frequency of onsets in two adjacent syllables in addition to positional probability and bi-phone probability may lead to a better estimate of the phonotactic probability of multi-syllabic words.

Finally, our results show that speakers can make gradient acceptability judgments and suggest that resorting to the lexical statistics of the speakers' language is a productive approach to explain the observed gradience. All non-words in our experiment had two onsets whose co-occurrence is phonotactically legal in Korean. If we limited speakers' sensitivity to phonotactic constraints to mean their ability to categorically distinguish phonotactically legal sound patterns from illegal ones, we would not expect our subjects to rate one group of non-words to be more acceptable than another. However, our subjects rated H-words to be more acceptable than L-words. In other words, our results reflect gradience in speakers' acceptability judgment. Furthermore, the gradience in their acceptability judgment reflects difference in how frequently two onsets co-occur in Korean. H-words consisted of onsets which co-occur with high frequency in Korean, while L-words consisted of onsets which co-occur with low frequency in Korean. As our subjects rated H-words to be more acceptable than L-words, the results suggest that the gradience observed in our subjects' acceptability rating is related to the difference in statistical distribution of onset co-occurrence pattern in Korean. In brief, our results add to the growing body of evidence of gradient acceptability and its relation to the statistical distribution of sound patterns in the language (Ohala and Ohala 1986, Coleman and Pierrehumbert 1997, Vitevitch et al. 1997, Frisch et al. 2000, Treiman et al. 2000, Bailey and Hahn 2001, Albright 2006).

Despite such implications, it is also true that care must be taken since our results are rather preliminary and raise many questions to be addressed in future research. For example, one question relates to the exact nature of onset-to-onset probability, defined here as the conditional probability of \( C_2 \) given \( C_1 \) in \( C_1V.C_2V.C_3 \) words. However, it is not clear from our experiment alone whether it estimates co-occurrence frequency of two onsets in adjacent syllables or co-occurrence frequency of two non-adjacent consonants separated by a vowel. Specifically, given \( C_1V.C_2.C_3V.C_4 \), will lexical processing be affected by the con-
ditional probability of \( C_2 \) given \( C_1 \), or \( C_3 \) given \( C_1 \)? Similarly, given \( C_1C_2V.C_3V \), will lexical processing be affected by the conditional probability of \( C_3 \) given \( C_1C_2 \), or \( C_3 \) given \( C_2 \)? Experiments that control for differences in consonant clusters will have to be conducted to resolve these issues.

Another question relates to the applicability of our results to other languages. The results are encouraging since Korean is not known for phonological dependencies between non-adjacent consonants, unlike Semitic languages such as Arabic or Hebrew. However, syllable structure in Korean is simple compared with languages such as English, and there is reason to believe that difference in syllable complexity affects how lexical statistics influence lexical processing tasks. For example, syllable structure is simpler in Cantonese than in English, and Kirby and Yu (2007) shows that correlation between bi-phone probability and speakers’ wordlikeness judgment is weaker for Cantonese than for English, while correlation between neighborhood density and wordlikeness judgment is stronger for Cantonese than for English. Therefore, our results must be replicated with speakers of languages with more complex syllable structure to further support our claims.

5. Conclusion

Phonotactic probability has been widely used to account for the frequency effect found in speakers’ behavior. In capturing the frequency effect of sound sequences on speech processing, psycholinguistic studies commonly limit the sound sequences whose distribution is probabilistically characterized to two adjacent phonemes. This approach has been effective in explaining the frequency effect and maintaining the statistical reliability of phonotactic probability used to characterize the frequency of sound patterns. Nevertheless, this approach is limited as there is prior reason to believe that speakers’ lexical processing behavior is also affected by phonological dependencies between non-adjacent phonemes. Experimental evidence from previous studies shows that Semitic language speakers are sensitive to co-occurrence restrictions on consonants that are root-adjacent but segmentally non-adjacent. Furthermore, recent phonotactic learning studies suggest the possibility that sensitivity to co-occurrence patterns on non-adjacent consonants may exist for speakers of languages without templatic morphology. Specifically, they show that speakers can quickly acquire sensitivity to how frequently onsets in two adjacent syllables co-occur in an artificially created language to which they are exposed. As the onsets in two adjacent syllables are segmentally non-adjacent, these studies suggest that co-occurrence frequency of non-adjacent phonemes such as onset-to-onset probability should be included to better estimate the overall phonotactic probability of novel sound sequences.
However, the simplicity of the artificial languages in which onset-to-onset probability was embedded leads one to question whether speakers can also learn onset-to-onset probability in their native language so that the acquired sensitivity affects how they process novel sound patterns. In this paper, we argued that speakers are indeed sensitive to onset-to-onset probability in their native language and therefore to better estimate the phonotactic probability of novel sound sequences, co-occurrence frequency of non-adjacent phonemes such as onset-to-onset probability should be considered as well as positional probability and bi-phone probability. To support our argument, we presented an experiment where adult Korean speakers rated acceptability of two groups of bi-syllabic non-words: non-words with higher onset-to-onset probability in Korean (H-words) and non-words with lower onset-to-onset probability (L-words). The results showed that subjects rated H-words to be more acceptable than L-words, implying that Korean speakers are sensitive to the onset-to-onset probability in their language and that their sensitivity to the probability is manifested in their gradient acceptability judgment of non-words.

References


Appendix A: List of non-words and mean acceptability ratings

<table>
<thead>
<tr>
<th>H-words</th>
<th>L-words</th>
</tr>
</thead>
<tbody>
<tr>
<td>saren</td>
<td>2.85</td>
</tr>
<tr>
<td>sarep</td>
<td>3.05</td>
</tr>
<tr>
<td>sarit</td>
<td>4.10</td>
</tr>
<tr>
<td>sirum</td>
<td>3.25</td>
</tr>
<tr>
<td>sirut</td>
<td>3.35</td>
</tr>
<tr>
<td>sorel</td>
<td>2.70</td>
</tr>
<tr>
<td>sorep</td>
<td>2.35</td>
</tr>
<tr>
<td>surel</td>
<td>2.40</td>
</tr>
<tr>
<td>suren</td>
<td>2.45</td>
</tr>
<tr>
<td>surit</td>
<td>4.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>H-words</th>
<th>L-words</th>
</tr>
</thead>
<tbody>
<tr>
<td>satj\textsuperscript{\texten}</td>
<td>2.20</td>
</tr>
<tr>
<td>satj\textsuperscript{\textep}</td>
<td>2.10</td>
</tr>
<tr>
<td>satj\textsuperscript{\textit}</td>
<td>3.50</td>
</tr>
<tr>
<td>sitj\textsuperscript{\textum}</td>
<td>3.25</td>
</tr>
<tr>
<td>sitj\textsuperscript{\textut}</td>
<td>2.80</td>
</tr>
<tr>
<td>satj\textsuperscript{\textel}</td>
<td>2.60</td>
</tr>
<tr>
<td>sortj\textsuperscript{\texten}</td>
<td>2.75</td>
</tr>
<tr>
<td>satj\textsuperscript{\textel}</td>
<td>2.65</td>
</tr>
<tr>
<td>satj\textsuperscript{\texten}</td>
<td>1.95</td>
</tr>
<tr>
<td>satj\textsuperscript{\textit}</td>
<td>3.85</td>
</tr>
</tbody>
</table>

Appendix B: An example ordered list of stimuli

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>supen</td>
<td>9</td>
<td>soper</td>
<td>17</td>
<td>sirut</td>
<td>25</td>
<td>surit</td>
<td>33</td>
</tr>
<tr>
<td>2</td>
<td>sakit</td>
<td>10</td>
<td>sukit</td>
<td>18</td>
<td>suken</td>
<td>26</td>
<td>sapen</td>
<td>34</td>
</tr>
<tr>
<td>3</td>
<td>sorel</td>
<td>11</td>
<td>satj\textsuperscript{\texten}</td>
<td>19</td>
<td>sotj\textsuperscript{\textel}</td>
<td>27</td>
<td>sitj\textsuperscript{\textut}</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>satj\textsuperscript{\textep}</td>
<td>12</td>
<td>sirum</td>
<td>20</td>
<td>supit</td>
<td>28</td>
<td>sakep</td>
<td>36</td>
</tr>
<tr>
<td>5</td>
<td>siput</td>
<td>13</td>
<td>saperp</td>
<td>21</td>
<td>sokel</td>
<td>29</td>
<td>sotj\textsuperscript{\textel}</td>
<td>37</td>
</tr>
<tr>
<td>6</td>
<td>surel</td>
<td>14</td>
<td>soper</td>
<td>22</td>
<td>sipum</td>
<td>30</td>
<td>sapit</td>
<td>38</td>
</tr>
<tr>
<td>7</td>
<td>sitj\textsuperscript{\textum}</td>
<td>15</td>
<td>satj\textsuperscript{\textit}</td>
<td>23</td>
<td>suren</td>
<td>31</td>
<td>soperp</td>
<td>39</td>
</tr>
<tr>
<td>8</td>
<td>saken</td>
<td>16</td>
<td>sukel</td>
<td>24</td>
<td>sotj\textsuperscript{\texten}</td>
<td>32</td>
<td>sarit</td>
<td>40</td>
</tr>
</tbody>
</table>

Hahn Koo  
Department of Linguistics  
4080 Foreign Languages Building, MC-168  
University of Illinois at Urbana-Champaign  
707 S. Mathews Avenue  
Urbana, IL, 61801, U.S.A.  
Email: hahnkoo@uiuc.edu

Received: September 19, 2007  
Revised version received: December 18, 2007  
Accepted: December 21, 2007