The Similarity Differential Rate Hypothesis

Roy C. Major
Arizona State University

Eunyi Kim*
Gilbert, Arizona

After Contrastive Analysis was called into question because it could not predict areas of difficulty, it was modified into the so-called moderate version, claiming that similar phenomena are more difficult to acquire than dissimilar phenomena (see Oller & Ziahosseiny, 1970). Eckman (1977) added a very significant modification to further salvage Contrastive Analysis by incorporating markedness. In addition, the prolific work of Flege on similarity, including his Speech Learning Model (1995), has provided us with extensive data on degree of difficulty in second language acquisition. Although the bulk of his and his collaborators’ work supports his claim that similar sounds are harder to learn than dissimilar sounds, there are important exceptions (e.g., Bohn & Flege, 1992). Furthermore, other researchers have presented conflicting claims and results. The importance of this chapter is that it ties together similarity and markedness in a claim concerning rate of acquisition, rather than “degree of difficulty”—a vaguely defined term. The Similarity Differential Rate

---

*Eunyi Kim is a graduate of Arizona State University. Experiment 1 was based on an unpublished paper by Kim (1994) that was part of the requirements for the MTESL degree. The Similarity Differential Rate Hypothesis itself is exclusively the idea of Major.

Correspondence concerning this article should be addressed to Roy C. Major, Director, Program in Linguistics and TESL, Department of English, Arizona State University, Tempe, Arizona 85287-0302. Internet: Roy.Major@asu.edu
Hypothesis (SDRH) quite simply claims that similar phenomena are acquired at faster rates than dissimilar phenomena and that markedness is a mediating factor that slows rate. In making its claims, it calls into question some widespread time-honored concepts in L1 and L2 studies that are poorly defined or mean different things to different researchers. Specifically, what criterion level or percent accuracy is needed for a phenomenon to be “acquired”? What does it mean for \( x \) to be acquired before \( y \)? If at Stage I, \( x = 40\% \) accuracy and \( y = 20\% \) accuracy, we could say \( x \) is acquired before \( y \) and is less difficult; but if at Stage II, \( x = 50\% \) and \( y = 80\% \), we could say \( y \) is acquired before \( x \) and is more difficult. In light of this, the SDRH claims that we have been asking the wrong questions, and that rate is the significant factor. In the example above, what is consistent is rate: The rate of acquisition of \( y \) is faster than \( x \).

In L1 acquisition research, although order of acquisition has been a perennial concern (e.g., Jakobson, 1941, 1968, and current research on the minimal word involving morae), according to David Ingram (personal communication, 1998), rate has not been a major topic in L1 phonology. (Although Ingram himself [1991] has talked about gradual and abrupt rates, he says there is no explanation why these variable patterns exist.) However, because rates have predictive implications (with the caveat that any learner, L1 or L2, may fossilize at any point) and because L1 and L2 phenomena are reflected in a wide range of contexts—for example, historical change, languages in contact, and dialect variation—it follows that the investigation of rate is an important area of research for both L1 and L2 acquisition.

Many researchers have observed that adults learning a second language (L2) have difficulty achieving native speaker (NS) levels of accuracy. The Critical Period Hypothesis claims that after a certain “critical period” L2 learners cannot achieve NS-like competence. The exact age at which the critical period ends is still unresolved; so is the question of whether there is a critical period at all. Some, like Scovel (1988), have claimed the Critical Period ends at puberty and applies only to phonology; others claim it
applies to other additional areas of language as well; for instance, Long (1990) claimed that the Critical Period for phonology may end as early as 6 or 7 but for other areas of competence may extend to, even past, puberty. Ioup, Boustagui, El Tigi, and Moselle (1994) studied Arabic production in the spontaneous speech of three native speakers (NSs) and two nonnative speakers (NNSs) of Egyptian Arabic, who were NSs of American English. Eight of 13 NS listeners judged the NNSs as native. Although a remarkable feat, to be judged as native by a majority of listeners (62%) cannot truly be called passing for native: In contrast, 100% of the judges rated 100% of the NSs as native. Munro, Flege, and MacKay (1996) examined English vowel production by Italian immigrants in Canada. They found significant effects of the age of arrival, and in fact the latest-arriving group failed to produce any of the 11 English vowels in a consistent NS-like manner, although these immigrants had been long-term residents of Canada (average 32 years). In sum, phonologists have yet to demonstrate that anyone who has learned an L2 after puberty can pass for native 100% of the time in all areas for which NSs pass for native 100% of the time.

On some tests even NSs are not rated native 100% of the time. Therefore, “passing for native” might be qualified to mean an L2 speaker who does as well as NSs on a battery of perceptual and productive experiments. (For further discussion on criteria for NS-like accent, see Munro et al., in press). Although probably no NNS can truly pass for native, this does not mean that some aspects of phonology are not acquired with NS-like competence from the standpoint of physical acoustic measurements and perception tests that judge NS-like production. Flege (1987) found that experienced English NSs speaking L2 French produced French /ü/ authentically (no significant differences from French NSs) but not /u/. Bohn and Flege (1992), studying German NSs speaking L2 English, showed that experienced German NSs produced the dissimilar English /æ/ more authentically than the inexperienced speakers, but showed fewer differences with the similar vowels /i ɛ/. Major (1992) found that one American
English NS speaking Brazilian Portuguese produced voiceless stops in Portuguese with VOTs (voice onset times) authentically, but 4 other participants did not. Why some phenomena but not others can be produced authentically is a complex and puzzling question, not only to theoreticians but also to practitioners, who ponder feasibility regarding what to teach.

Contrastive Analysis alone does not offer any explanation of order or rate of acquisition because there is no theoretical basis to predict which areas should be more problematic than others: Any area of difference between L1 and L2 should cause difficulty (Lado, 1957). However, Oller and Ziahosseiny (1970) proposed a “moderate version” in an attempt to explain hierarchy of difficulty: “wherever patterns are minimally distinct in form or meaning in one or more systems, confusion may result” (p. 186). They found that ESL learners whose L1 used a non-Roman script, such as Japanese, found English spelling easier than did learners whose L1 used a Roman script, such as French. Applying this principle to phonology: An L2 sound that has a corresponding similar sound in the L1 (e.g., French dental /t/ vs. English alveolar /t/) will be harder to learn than a sound for which the corresponding sound in the L1 is dissimilar (e.g., French or German uvular /R/ vs. English /r/). Gross differences are more often noticed, due to their perceptual saliency, whereas minimal differences are more likely to be overlooked and to result in confusion or nonlearning.

Many researchers (including Bohn & Flege, 1992; Flege, 1987; Flege & Hillenbrand, 1984; Major, 1987a; Wode, 1978, 1983a, 1983b) have investigated the relationship between phonological similarity/dissimilarity (variously called new or different phenomena) and difficulty or order of acquisition. Wode (1983a) incorporated the notion of similarity/dissimilarity to claim that L1 transfer can take place only when “crucial similarity measures” (p. 180) occur between the L2 target items and the corresponding L1 items. That is, only the L2 elements that meet “specifiable similarity requirements” (p. 185) are substituted by L1 elements. Other L2 elements, which do not meet the similarity requirements, are acquired via the same developmental sequences.
believed to characterize L1 acquisition. A substitution of the latter type in German speakers of English is Wode’s example of [w] substitution for English /r/, rather than the L1 substitution German /R/. His other examples, of L1 substitutions in German speakers of English, are widely known: [ɛ] for /æ/, [l] for velarized [l], [s] for /θ/ and [z] for /ð/. If Wode’s claim about similarity requirements is true, Oller and Ziahosseiny’s (1970) claim—that different or dissimilar phenomena are easier to learn than similar ones—follows logically: Straight negative transfer (which should occur for similar phenomena) is nonlearning, because nothing of the L2 has been incorporated into the interlanguage (IL); however, developmental processes and substitutions (which should occur for dissimilar phenomena) is learning because the IL is passing through stages that progressively approach adult NS competence, and furthermore these stages are the same as (or similar to) those occurring in L1 acquisition. (See Major, 1995, for a discussion of the relationships and changes between Underlying Representations and Surface Representations.)

Much of Flege’s work has investigated similarities/dissimilarities between L1 and L2 and how they affect rate and order of acquisition, in both production and perception. Using the term “equivalence classification,” Flege (1987) claimed that “equivalent” or “similar” sounds are difficult to acquire because a speaker classifies or perceives them to be equivalent to those in his or her L1, whereas “new” (dissimilar or different) sounds are easier because the speaker readily notices the differences. Flege supported his hypothesis with a number of his own studies. For example, in his 1987 study of English NSs, he found that experienced L2 learners of French produced a dissimilar sound, French /ü/, authentically, whereas their French /u/, which has a close counterpart in English /u/, remained English-accented. In their study of German speakers’ production of dissimilar and similar sounds, Bohn and Flege (1992) found similar results. Their study provided further evidence that success or failure to learn L2 sounds is explicable in terms of sound correspondences between the L1 and L2 and in terms of L2 experience. The study showed
that extended L2 experience did not help German speakers produce the similar English sounds /i iɛ/ authentically (as determined by formant frequency measurements) because they did not establish “phonetic categories” for similar sounds, due to blocking of category formation caused by “equivalence classification” (Bohn & Flege, 1992, p. 132). On the other hand, these experienced German speakers produced the dissimilar sound /æ/ authentically because apparently they had concentrated on and had learned important phonetic characteristics of the dissimilar vowel category. Major (1987b) also found that a similar sound was harder to acquire than a dissimilar sound for Brazilian Portuguese NSs attempting English. The study revealed that as global foreign accent decreased, accuracy of the dissimilar sound /æ/ improved, whereas accuracy of the similar sound /ɛ/ actually decreased (Portuguese has /ɛ/ but no /æ/), which Major attributed to overgeneralization of /æ/ and /ɛ/.

Although the role of similarity and dissimilarity seems well-documented and convincing in these studies, what constitutes similar or dissimilar is not always clear. Though probably most linguists would agree that French and English ts are more similar than their rs, in many other areas of language similarities and dissimilarities are not so obvious; for example, is German /x/ more similar to English /k/ than German /s/ is to English /θ/?

Investigators have also invoked markedness to explain different levels of proficiency and order of acquisition. Eckman’s (1977) pioneering Markedness Differential Hypothesis (MDH) incorporates degree of difficulty of L2 phenomena based on markedness considerations: Less marked phenomena are acquired before more marked ones. Stating this in terms of hierarchical relationships means that the presence of more marked phenomena (i.e., they have been acquired) implies the presence of less marked phenomena, but not vice versa. Carlisle (1988) modified this hypothesis in his Intralingual Markedness Hypothesis (IMH), which incorporated markedness relationships within the target language (TL), not just between the L1 and TL as Eckman had done. Both the MDH and IMH can be supported by studies of
voicing contrasts in obstruents (summarized by Yavas, 1994). Korean has only voiceless obstruents; because English voiced obstruents in final position are more marked than in initial and medial position, Koreans should acquire English voicing contrasts in final position only after the other two positions have been acquired. This prediction was borne out in Major and Faudree's (1996) study of Korean speakers of English, who showed nearly 100% accuracy for voiced obstruents in initial and medial position but only about 50% accuracy in final position. Other work on L2 acquisition further supports the notion that markedness universals in the languages of the world also hold true for L2 acquisition (Carlisle, 1988, 1991, 1994; Eckman, 1981a, 1981b, 1985; Eckman & Iverson, 1993, 1994).

Eckman (1984) found word-final devoicing in the English of native Farsi speakers, which one would not predict on the basis of positive transfer because Farsi has voicing contrasts in this position (in addition to initial and medial positions). To account for these and other data, Eckman (1991) proposed the Structural Conformity Hypothesis (SCH), which is not based on the areas of difference, but simply and elegantly claims that interlanguages obey primary language universals. Thus, an IL can have word-initial and word-medial voice contrasts without having final voice contrasts, even if the L1 does have contrasts in all 3 positions (e.g., the Farsi study cited above). Eckman further supported his hypothesis with a study of the acquisition of consonant clusters by Korean and Japanese ESL learners (in Japanese and Korean there are no consonant clusters), finding that the less marked clusters (2 consonants) were acquired before the more marked clusters (3 consonants).

Thus, both similarities and differences between L1 and L2 and markedness are important in L2 acquisition. However, at times these 2 factors can confound our understanding of the complex nature of L2 phonological acquisition, because in certain instances they seem to be at odds. If similarity/difference considerations offered a complete explanation of the order of acquisition, dissimilar phenomena should be acquired more easily than similar
ones, regardless of markedness considerations. However, markedness universals predict that more marked phenomena are harder to acquire than less marked ones (Eckman, 1977, 1991), regardless of what similarities or differences exist between L1 and L2. Thus, these 2 factors—similarity and markedness—can compete. Consider an English speaker learning Arabic /x/ and /ʕ/: English /k/ is more similar to /x/ than to /ʕ/ (because both /k/ and /x/ are voiceless and velar, whereas /ʕ/ is voiced and pharyngeal) and accordingly /x/ should be harder to learn; however, from the standpoint of markedness, /ʕ/ should be harder to learn because it is much more marked than /x/. This case presents a dilemma, because the two claims predict opposite results. Perhaps there is a pecking order of similarity versus markedness; that is, a high degree of markedness will overrule the dissimilarity factor, or perhaps the dissimilarity factor will overrule a low degree of markedness. In the case of English learners of Arabic, it seems that the extreme markedness of /ʕ/ overrules the similarity of English /k/ and Arabic /x/: Although we are unaware of any published studies of these 2 sounds in English learners of Arabic, informal comments by Arabic teachers all agree that /ʕ/ is much more difficult to acquire than /x/. In a similar fashion, Xhosa clicks are very dissimilar to any English or Afrikaans phoneme, yet many of these clicks are very difficult to acquire (Lewis, 1994). Further studies illuminating these 2 competing factors are needed.

Markedness and similarity/dissimilarity can work together to the benefit or detriment of the learner. An L2 sound that is both more marked and similar to an L1 sound should be harder to acquire than an L2 sound that is less marked and dissimilar to an L1 sound (unlike the Arabic example above). We investigate this possibility by comparing the acquisition of a more marked L2 sound that is very similar to an L1 allophone, with a less marked L2 sound that does not have such a close similarity to an L1 sound.

Korean, unlike English, has only voiceless obstruents, but they contrast in three ways: Voiceless unaspirated (lenis), voiceless aspirated, and voiceless glottalized (fortis). The complete consonant and vowel system in Korean is as follows: /p pʰ p’ t tʰ t’
For the present study, which investigates acquisition of English /j/ and /z/, there is one relevant Korean phonological rule: The lenis obstruents—with the exception of /s/*—become voiced in intervocalic position. Thus, /c/* becomes [j] but /s/* does not become *[z] (however, /s/*->[[s] when followed by the high, front vowel /i/); the fortis and aspirated obstruents also do not participate in this voicing rule. According to Huh (1983), Korean once had a voiced fricative sound, similar to English /z/, but it became extinct in the 1600s. The English phoneme /j/* has a similar counterpart in Korean, because in Korean lenis /c/* becomes [j] between vowels (in initial and final positions there are only voiceless obstruents), whereas the English phoneme /z/* does not have a similar counterpart in Korean, because the lenis /s/* does not change to [z] under any condition. We can therefore reasonably conclude that the English phoneme /z/* is a dissimilar (new) sound for Koreans to acquire, whereas the English phoneme /j/* is a similar sound. Furthermore, /j/* is more marked than /z/*, because affricates are more marked than fricatives.

**Hypotheses**

Our general hypothesis is that similar sounds are more difficult to acquire than dissimilar sounds; that is, L2 experience will not affect L1 speakers' acquisition of a similar L2 sound as much as their acquisition of the dissimilar sound. Specifically, the amount of English L2 experience should not noticeably affect Korean NSs' production of English /j/*, which has a counterpart in Korean, whereas English experience will enable Korean NSs to produce an English-like /z/*, which has no counterpart in Korean. Furthermore, the two factors—dissimilarity and a low degree of markedness—should complement each other and foster acquisition of the dissimilar sound /z/*, which is less marked than the similar sound /j/*.

That dissimilar phenomena are easier to learn than similar ones is a very general statement, the specifics of which one can
interpret in at least three different ways. We formulate these possibilities into the following hypotheses:

Hypothesis 1. Competence with dissimilar phenomena is higher than with similar phenomena. This is the most unrestricted form of the similarity/dissimilarity hypothesis: Regardless of phenomena, exposure, experience, or stage, learners demonstrate more proficiency producing dissimilar sounds than producing similar sounds. Counterevidence would be learners who produce similar phenomena better than dissimilar ones, or learners who show no differences between the two types of phenomena.

Hypothesis 2. For advanced learners, competence with dissimilar phenomena is higher than with similar phenomena; for beginning learners, competence with similar phenomena is higher than with dissimilar phenomena. A second interpretation of the similarity/dissimilarity hypothesis: Due to transfer, beginning learners perform better with similar sounds because the nearest equivalent in the L1 is perceptually close to the corresponding sound in the L2 (e.g., a French NS learner of English using French dental stops in lieu of English alveolar stops). However, for dissimilar sounds these beginning learners do poorly, because they substitute their L1 equivalents, which are very different acoustically and perceptually from the L2 sounds (e.g., French NSs using French [R] for English [r]). On the other hand, advanced learners have noticed these large differences and have concentrated on these dissimilar sounds, while ignoring the less obvious differences of similar sounds. This form of the hypothesis draws support from much of Flege’s work (e.g., 1987); Flege found that English NSs producing French differed in their production of /ü/ depending on their experience. The experienced speakers produced /ü/ with native-like accuracy but the inexperienced speakers did not. However, neither group produced authentic /u/. Counterevidence for this hypothesis would be advanced learners whose competence with similar phenomena is better than with dissimilar ones, beginning learners whose competence with dissimilar phenomena is better than with similar ones, or beginning and advanced learners who show no differences in either similar or dissimilar phenomena.
Hypothesis 3. The Similarity Differential Rate Hypothesis (SDRH): An L2 phenomenon that is dissimilar to an L1 phenomenon is acquired faster than an L2 phenomenon that is similar to this same L1 phenomenon. This third interpretation of the similarity/dissimilarity hypothesis is similar to the second except that it makes no claims about ultimate competence; it simply hypothesizes that learners acquire dissimilar phenomena faster than the corresponding similar phenomena. In other words, if L1x is a phenomenon in the L1, and if in the L2 there is a corresponding dissimilar phenomenon L2x_d and a similar phenomenon L2x_s, then L2x_d is acquired faster than L2x_s. This hypothesis allows for variability in outcomes of different phenomena. Different phenomena have different rates, probably due to a variety of factors, such as intrinsic difficulty. For example, if x is a similar sound and y and z are both dissimilar sounds but y is more marked than z, then both y and z should be acquired more quickly than x; however, y should be acquired more slowly than z and, in fact, competence in y may never surpass that of x. When applied to cross-sectional data, this means the difference in competence between beginning and advanced learners for dissimilar phenomena should be greater than for similar ones, and if line graphs were made the upward slopes should be greater for dissimilar phenomena than for similar ones. Counterevidence for this claim would be learners whose rate of acquisition for similar phenomena is faster than the rate for dissimilar ones, or learners whose rate of acquisition for similar phenomena is equal to the rate for dissimilar ones; for cross-sectional data counterevidence would be slopes for dissimilar phenomena that are less steep than for similar ones.

The resemblance of the name—the Similarity Differential Rate Hypothesis—to Eckman’s (1977) Markedness Differential Hypothesis is intentional. Whereas Eckman’s MDH takes into account markedness differentials, our SDRH takes into account similarity differentials. However, besides the obvious difference between similarity and markedness, there are important differences between Eckman’s MDH and our SDRH. In the SDRH, rate of acquisition is the key, not areas of difficulty or ultimate outcome;
in the MDH the term “areas of difficulty” (Eckman, 1977, p. 321) is the key (marked phenomena are more difficult than unmarked ones). However, areas of difficulty, a problematic and vague term, may be interpreted in terms of stages or rate. If \( x \) is more difficult to acquire than \( y \), does that mean \( x \) is acquired more slowly than \( y \), or that at any given stage the level of competence in \( x \) is below that in \( y \), or both? However, what if the rate of acquisition of \( x \) is faster than \( y \), but at any given stage the competence level of \( x \) is below \( y \)? Which, then, is the more difficult phenomenon? Furthermore, although the MDH does not take into account similarity, the SDRH takes into account both markedness and similarity: The SDRH claims that a dissimilar phenomenon is acquired more quickly than a similar one but markedness is a mediating factor; that is, given a similarity differential, a greater degree of markedness will decrease rate of acquisition. (This could be called a reinterpretation of the MDH in terms of rate.) Eckman himself has pointed out problems with his MDH. Although we believe it is still fundamentally sound if confined to rate (marked phenomena are acquired more slowly than unmarked ones), his Structural Conformity Hypothesis (SCH, 1991) remedies these purported problems, claiming that interlanguages obey primary language universals. The SDRH is also consistent with the SCH: Although we claim that a dissimilar phenomenon is acquired more quickly than a similar one, universals still necessarily apply, among them markedness principles.¹

Method

**Speakers**

Three groups of 10 adult speakers (5 men, 5 women per group) participated in this study: a group of inexperienced Korean NS speakers of English, a group of experienced Korean NS speakers of English, and a group of American English NSs. The third group (G-NS) consisted of monolingual English NSs, ages from 24
to 44, with a mean of 35. Three of the 10 speakers in G-NS were from California, 2 from Michigan, 2 from Kansas, and the remaining 3 from New Jersey, Ohio, and New York, respectively.

The two groups of Koreans differed in their amount of English language exposure in the U.S. The speakers in the inexperienced group (G-I) had lived in the U.S. for a mean of 0.5 years, whereas the mean for the experienced group (G-E) was 11.6 years. Both groups had studied English in school for 6 years before they came to the U.S. (in Korea, education in English as a foreign language (EFL) is obligatory in junior-high and high school). The age range for G-I was 22 to 27, with a mean of 26; for G-E the age range was 32 to 42, with a mean of 36. There was a 10-year difference in the mean age between G-I and G-E because we chose only college-educated adults. The speakers in G-E had come to the U.S. after puberty (either before college or after college), had attended a 4-year college, and had lived here for a minimum of 10 years. On the other hand, those in G-I were college graduates who were recent arrivals to the U.S.

Speech Materials

We asked members of all three groups individually to read a list of 20 words, randomly ordered, with /j/ and /z/ in initial position, consisting of 5 minimal pairs repeated twice. Because vowel quality can influence the preceding consonant (e.g., universally, palatalization of dental and alveolar obstruents before high front vowels is common; lower vowels favor voicing of obstruents: Yavas, 1994), we chose minimal pairs to include high, mid, and low vowels, both in the front and back dimension: /i e æ u o/ (we excluded other vowels because there are no minimal pairs, e.g., /je/ jay but no */ze/, */ja/ */za/). The list of words appears in Table 1.

We recorded the speakers as they produced each word in the carrier phrase “I will say ___.” (They did not read the number of each sentence.) After we had recorded all the speakers, we dubbed the tapes and randomized the order of production by each speaker from the 3 different groups. We deleted the number code of each
group in order to avoid any possibility of rater bias. We dubbed the randomized oral productions of the 30 speakers with new numbers (i.e., Speaker No. 1, Speaker No. 2, . . . Speaker No. 30.)

**Experiment 1**

A group of 5 listeners, NSs of American English, rated the recorded and edited tape. These listeners were not from G-NS. There were 3 men and 2 women, ages 22–43, from 5 different states. We told these NS listeners that they would be listening to a number of NSs and NNSs of English saying, “I will say ___,” and that after they heard each sentence they would decide which of 2 words they heard by circling the word they heard. For example, if the speaker said, “I will say ___” and they heard *jest*, they should circle *jest*; if they heard *zest*, they should circle *zest*. It took the listeners about 40 minutes to judge the 30 speakers. We compared the /j/ and /z/ sounds produced by the 3 groups respectively through single-factor Analysis of Variance (ANOVA, n=3000: 20 words × 30 speakers × 5 listeners). To test interrater reliability of the 5 raters, we analyzed the data using Kendall Tau-B coefficients.

**Experiment 2**

We played the same tape to a different group of 5 American English NS listeners (3 women and 2 men, ages 20–40, from 5 different states). We told the raters that they would be listening NSs and NNSs of English, saying, “I will say ___.” We gave them
the word list and told them that after each sentence they should decide how NS-like the key word sounded, on a scale from 1 “very heavy foreign accent” to 5 “no foreign accent.” It took about 40 minutes for the listeners to complete this task. Similarly to Experiment 1, we analyzed the data by a single-factor ANOVA \((n=3000: 20\text{ words } \times 30\text{ speakers } \times 5\text{ listeners})\) and tested inter-rater reliability using Kendall Tau-B coefficients. Thus, the difference between Experiments 1 and 2 is that 1 is a categorical either/or task, whereas 2 is a scalar rating task.

**Results**

**Experiment 1**

Table 2 and Figure 1 indicate the results. The G-NS speakers scored 100% by all 5 judges, meaning these NS listeners heard exactly the sounds the G-NS speakers intended to make. Both groups of Korean NSs performed better for the similar English sound /j/ than for the dissimilar English sound /z/. The G-E speakers performed much better for the dissimilar sound /z/ than did G-I (69.2% vs. 10.4%); however, G-E actually performed worse than the G-I for the similar sound /j/ (88.6% vs. 96.8%). These differences between the 2 groups in their production of both /j/ and /z/ were significant at \(p<0.01\) on single-factor ANOVA analysis. The differences in standard deviations for /j/ and /z/ production between the 2 groups show there was more variability in both /j/ and /z/ in G-E compared to G-I (4.81 vs. 2.12, 14.07 vs. 10.24). In addition, the standard deviations of both groups for /z/ were much greater than for /j/: (z: G-I: \(SD=10.24\), G-E: \(SD=14.07\); /j/: G-I: \(SD=2.12\), G-E: \(SD=4.81\)). There was more variability in competence with the dissimilar sound for both groups than with the similar sound. There was good interrater reliability: Kendall Tau-B analysis showed the differences in scores by the 5 raters were not significant \((p<0.001)\).
Table 2

Experiment 1: Categorical Judgments of /ʃ/ and /z/ by Native Speakers of English

<table>
<thead>
<tr>
<th></th>
<th>G-I</th>
<th></th>
<th>G-E</th>
<th></th>
<th>G-NS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>SD</td>
<td>%</td>
<td>SD</td>
<td>%</td>
</tr>
<tr>
<td>/ʃ/</td>
<td>1,500</td>
<td>96.8</td>
<td>2.12</td>
<td>88.6</td>
<td>4.81</td>
<td>100.0</td>
</tr>
<tr>
<td>/z/</td>
<td>1,500</td>
<td>10.4</td>
<td>10.24</td>
<td>69.2</td>
<td>14.07</td>
<td>100.0</td>
</tr>
</tbody>
</table>

G-I=Inexperienced Korean speakers of English; G-E=Experienced Korean speakers of English; G-NS=Native speakers of English.

Figure 1. Experiment 1: Categorical judgments of /ʃ/ and /z/ by native speakers of English.
Experiment 2

The results are shown in Table 3 and Figure 2. The productions of G-NS were given nearly perfect ratings by the NS listeners (4.96 and 4.93 for /j/ and /z/, respectively). As in Experiment 1, the Korean groups produced the similar English sound /j/ better than the dissimilar English sound /z/. Group G-E speakers performed better than G-I for both sounds; however, the differences were greater for the dissimilar sound /z/ (2.71 vs. 1.25) than for the similar sound /j/ (3.00 vs. 2.84). Both differences proved significant.

Table 3

Experiment 2: Ratings of /j/ and /z/ by Native Speakers of English, on a Scale of 1–5 (1=very heavy foreign accent and 5=no foreign accent)

<table>
<thead>
<tr>
<th></th>
<th>G-I</th>
<th>G-E</th>
<th>G-NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>/j/</td>
<td>1,500</td>
<td>2.84</td>
<td>1.19</td>
</tr>
<tr>
<td>/z/</td>
<td>1,500</td>
<td>1.25</td>
<td>0.61</td>
</tr>
</tbody>
</table>

G-I=Inexperienced Korean speakers of English; G-E=Experienced Korean speakers of English; G-NS=Native speakers of English.
at $p<0.05$ on single-factor ANOVA analysis. Interrater reliability was also good: The differences in scores by the 5 raters were not significant ($p<0.001$) on Kendall Tau-B analysis. The differences in standard deviations for /j/ and /z/ production between G-E and G-I groups show, unlike Experiment 1, little variability in /j/ between the two groups (1.22 vs. 1.19); however, like Experiment 1, there was more variability in /z/ ratings in G-E compared to G-I (1.41 vs. 0.61), suggesting that, although G-E speakers had higher overall competence, there was more variability in their competence.

**Summary of Experiments 1 and 2**

Native speaker judgments of competence in phonology of NNSs involve both categorical judgments (“Did I hear this word or that word?”) as well as how NS-like the speaker sounds. In light of this, we combined the results of Experiments 1 and 2, yielding an overall competence rating or a rating of how NS-like the two groups of NNSs were, by comparing their scores with the scores of NSs (i.e., the scores of the NNSs divided by the scores of the NSs). Because the scoring of the two experiments was different (percentage for Experiment 1 and scalar for Experiment 2), we had to convert the scores of Experiment 2 to percentages by taking the scores of the NNSs and dividing them by the scores of the NSs (e.g., /z/ for G-I is $1.25/4.93=25.4\%$). These results were then averaged with those from Experiment 1, giving an overall level of competence expressed in percentage, the baseline being NSs, who by definition are 100% NS competent. These calculations appear in Table 4 and Figure 3.

Because Experiment 2 was converted to a percentage score and the two experiments were based on different tasks, it would not be appropriate to statistically analyze the two experiments combined. However, the difference in the graphic representation (Figure 3) of the two sounds /j/ and /z/ is dramatic and corroborates the trends of the two individual experiments: There was very little difference in /j/ production between G-I and G-E (77.1 vs. 74.6);
### Table 4

**Experiments 1 and 2 Combined: Categorical and Rating Judgments of Native Speakers of English Judging Production of /j/ and /z/ by Korean Speakers of English**

<table>
<thead>
<tr>
<th></th>
<th>Inexperienced Korean Speakers</th>
<th>Experienced Korean Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experiment 1</td>
<td>2</td>
</tr>
<tr>
<td>/j/</td>
<td>96.8</td>
<td>57.3</td>
</tr>
<tr>
<td>/z/</td>
<td>10.4</td>
<td>25.4</td>
</tr>
</tbody>
</table>

**Figure 3.** Experiments 1 and 2 combined: Categorical and rating judgments of native speakers of English judging production of /j/ and /z/ by Korean speakers of English.
However, there was a clear difference in /z/ production between G-I and G-E (17.9 vs. 62.1).

Discussion

Hypotheses

The data do not support Hypotheses 1 or 2; however, they do support Hypothesis 3: the SDRH.

Hypothesis 1. Competence with dissimilar phenomena is higher than with similar phenomena. This hypothesis is disconfirmed; both the beginning and advanced learners produced the similar sound /j/ more accurately than the dissimilar sound /z/ (Figures 1, 2, & 3).

Hypothesis 2. For advanced learners, competence with dissimilar phenomena is higher than with similar phenomena; for beginning learners, competence with similar phenomena is higher than with dissimilar phenomena. This hypothesis is also disconfirmed. Although, as predicted, beginning learners produced /j/ more accurately than /z/, so too did the advanced learners, contrary to the hypothesis (Figures 1, 2, & 3).

Hypothesis 3. The Similarity Differential Rate Hypothesis (SDRH): An L2 phenomenon that is dissimilar to an L1 phenomenon is acquired faster than an L2 phenomenon that is similar to this same L1 phenomenon. The data support the SDRH: The differences between productions of the similar sound /j/ and the dissimilar sound /z/ were much greater for the advanced learners than the beginning learners (Figures 1, 2, & 3); in fact, in the categorical judgment task (Figure 1), beginning learners slightly outscored the advanced learners in the similar sound, giving a slight negative slope if a line graph had been made (Figure 3).

General Discussion

The data support Hypothesis 3, the SDRH, but do not support the other two hypotheses. Although other studies support
Hypotheses 1 and 2 (e.g., Oller & Ziahosseiny, 1970, supports Hypothesis 1; Bohn & Flege, 1992, partially supports Hypothesis 2; and Flege, 1987, supports Hypothesis 2), we will argue that Hypotheses 1 and 2 logically can be true only if learners’ competence levels and types of phenomena are so highly specified and restricted that the hypotheses’ generality is lost. Therefore, we argue that Hypothesis 3, the SDRH, can best account for all the data.

Hypothesis 1 may hold true for advanced learners; for example, Flege (1987) found English NSs performed better with French /ü/ (the dissimilar sound) than with /u/. However, the hypothesis does not stipulate proficiency levels; it can be shown to be false: It does not hold true for beginners because it naïvely implies that even true beginners will more proficiently produce dissimilar sounds than similar sounds. This would imply instantaneous acquisition—an “immaculate conception” in L2 acquisition. However, a true rock-bottom beginner should necessarily transfer all or nearly all L1 phonology into L2, having almost no L2 exposure and thus having learned very little of the L2. Accordingly, the L1 substitution for the dissimilar phenomenon should be acoustically and perceptually further off the mark than for the similar one. The dissimilar phenomenon has no L1 equivalent (perceptually, acoustically, or articulatorily) whereas the similar phenomenon has an L1 near-equivalent; therefore, partial positive transfer occurs for the similar phenomenon. Thus, true beginners tend to perform better for similar sounds (albeit not native-like) than for dissimilar sounds. Later, as learners start to acquire more of the L2, their proficiency for dissimilar sounds can increase and even surpass their proficiency for similar sounds.

Hypothesis 2 is also false because it does not allow for different rates of acquisition for different phenomena. Bohn and Flege (1992) claimed that L2 experience does not affect similar sounds, whereas extensive L2 experience enables learners to produce NS-like dissimilar (“new”) sounds. Perhaps their claim is too restrictive: It does not allow for improvement in similar phenomena nor does it allow for some dissimilar phenomena never
being mastered, due to markedness considerations (intrinsic difficulty). Furthermore, it is unclear how much L2 experience learners would have to have for researchers to legitimately test the hypothesis. Hypothesis 2 resembles Bohn and Flege’s, except that it does not constrain similar phenomena to a zero rate of acquisition. As discussed below, Hypothesis 2 is really a subset of the more general SDRH.

Hypothesis 3, the SDRH, covers a range of phenomena that Hypotheses 1 and 2 necessarily omit or cannot account for. The SDRH claims that the most general principle involving dissimilarity and similarity is rate: Dissimilar phenomena are acquired faster than similar ones. Evidence for the SDRH comes from documented general principles of L2 acquisition, from the present study, and from other studies. Specifically the SDRH can account for, or follows logically from, (a) different rates and different stages, (b) overgeneralization and interaction of phenomena, (c) markedness principles, and (d) evidence from other studies.

Different Rates, Different Stages

It has long been known that different phenomena have different rates of acquisition. The SDRH encompasses this observation by claiming that the rates of acquisition for dissimilar phenomena are faster than for similar phenomena but that the exact rates and stages for different phenomena may differ. Figure 4, a hypothetical graph, shows several possible rate differences. These straight lines represent an idealized learner. Real learners often show curvilinear and jagged learning graphs: They plateau, fossilize, backslide, leap to new heights, and forget. However, even though a frictionless surface and a perfect vacuum are not found on Earth, they are still important theoretical constructs; so too is the idealized learner.

At t₁ in Figure 4, because of a high level of L1 transfer due to a very low level of acquisition, both dissimilar phenomena y and z are at a lower level of accuracy than both similar phenomena w and x. Dissimilar phenomena’s L1 substitutions in the L2 are
acoustically, perceptually, and articulatorily further from the L2 than substitutions for similar phenomena. It also follows that the L1 substitutions for \( w \) must be the closest perceptually and acoustically to the L2 and the L1 substitutions for \( z \) must be the furthest from the L2. In addition, in terms of ranking from most similar to least similar, it follows that \( w > x > y > z \). At \( t_2 \), \( y \) has surpassed \( x \) and is equal to \( w \), but \( x \) and \( z \) are still below \( w \), even though \( w \) accuracy has decreased. At a much later stage, \( t_n \), some hypothetical time in the future, \( y \) and \( z \) will necessarily surpass \( w \) and \( x \) because the rates of acquisition of dissimilar phenomena are faster than for similar ones.

We do not claim that all learners will ultimately reach stage \( t_n \): It is a theoretical construct, a point that many learners may never achieve. In addition, we do not claim that all learners will in fact show learning curves at all: Some may never acquire anything—they may fossilize from the start. For example, some
L2 learners of Highland Chontal, a language of Mexico (Turner, 1972), may never show progress with the voiceless alveolar laterally released affricate ejective ([t’]). However, we do claim that if the acquisition process starts from ground zero, then the changes observed will be those specified by the SDRH. Furthermore, because we claim for similar phenomena that the slope may be positive, negative, or zero, at some unspecifiable time in the future, \(t_{n+1}\) (not shown) that \(w\) may well level off or show an upturn. Otherwise, if the graph continued in a linear fashion, proficiency would eventually reach zero, an improbable event because at earlier stages the learner demonstrates a rather high proficiency with \(w\). In addition, not all studies should be expected to show all these trends because of problems of internal validity, which can involve methodological design. The shortcomings of cross-sectional studies over longitudinal studies have long been known, as have the problems of lumping speakers together rather than looking at individual learning patterns. Generally cross-sectional and group studies can obscure and blur true stages, which in some instances can be seen only in longitudinal studies of individual learners.

Hypothesis 1 could be presumed in Figure 4 only if the learners under investigation happen to be very advanced; that is, they are at stage \(t_n\) (but not at \(t_1\) or \(t_2\)). Hypothesis 2 would hold true if one happens to pick two groups of learners, one from stage \(t_1\) and the other from \(t_n\) (but not \(t_2\)). Thus, Hypotheses 1 and 2 are only partially true; that is, they are true if the researcher fortuitously picks learners at the right level. However, Hypothesis 3, the SDRH, can account for learners at all levels and stages.

The data from the present study roughly fit the patterns of Figure 4. In Figures 1, 2, and 3, /z/, the dissimilar sound, shows a sharp increase (from G-I to G-E). The similar sound, /j/, shows a very slight decrease in Figures 1 and 3 (from G-I to G-E), but a very slight increase in Figure 2. If this rate of acquisition continues, an extrapolation to some group, for example, a very experienced group (G-VE) would mean that /z/ would surpass /j/.
Interactions of phenomena and overgeneralizations are legion in L1 and L2 acquisition, speech disorders, historical change, and dialect variation. Complete overgeneralization is a merger or unity between two or more phenomena. For example, consider the acquisition of English liquids. A speaker whose L1 has /l/ but no /r/ (e.g., Korean) produces /l/ for both /l/ and /r/ in English. As acquisition proceeds, interaction of /l/ and /r/ occurs, although attempts at /r/ become more r-like, attempts at /l/ also become more r-like but less l-like. This we may term partial overgeneralization. If interaction continues in the same pattern to its logical conclusion (happily this does not occur for many learners), both /l/ and /r/ may be completely overgeneralized to /r/.

Consider Figure 4 in light of these points, and as it pertains to English learners of French acquiring /ü/ and /u/. Let \( z = /ü/ \) and \( w = /u/ \). Typically, beginning learners will overgeneralize both /ü/ and /u/ to /u/ and substitute their English /u/ for both French sounds. This is reflected at \( t_1 \), which shows a higher proficiency for /u/ (\( w \)) than for /ü/ (\( z \)). At \( t_2 \), /u/ is still below /ü/, despite the fact that /u/ is decreasing and /ü/ increasing rapidly. As these trends continue, at \( t_n \), /ü/ proficiency has surpassed /u/. One explanation for the decrease in /u/ proficiency is that both /ü/ and /u/ have been overgeneralized to /ü/ whereas at \( t_1 \) they were both overgeneralized to /u/.

Figure 1 suggests that overgeneralization may have occurred, although the difference in proficiency for /j/ between G-I and G-E is perhaps too slight to warrant this claim.

**Markedness**

A number of researchers (Carlisle, 1988, 1991, 1994; Eckman, 1977, 1984, 1991; Eckman & Iverson, 1993, 1994) have shown that more marked phenomena are acquired later than unmarked ones. Markedness itself does not affect the basic claims of the SDRH; it is simply a mediating factor: When similarity/dissimilarity
specifications are held constant, relatively more marked phenomena will exhibit slower rates than relatively less marked ones. That is, if in Figure 4, y and z have the same degree of similarity to a phenomenon in L1, and z is more marked than y, then z will be acquired more slowly than y.

However, markedness and similarity can interact and vary in ways that can produce mutually facilitating or opposing forces. For example, comparing the acquisition of a dissimilar but marked phenomenon to a similar but unmarked one should enable researchers to see which factor has the more powerful effect, because similarity and markedness should have opposing effects: for example, an English speaker acquiring Arabic /x/ and /ʕ/, discussed in the introduction. Considering English /k/ (or even /h/) as the basis of comparison, similarity/dissimilarity considerations would predict that the dissimilar element /ʕ/ should be acquired before the similar element /x/. However, markedness considerations would predict the opposite, because /ʕ/ is more marked than /x/. Perhaps the second prediction holds true—many teachers of Arabic observe that often English NSs never acquire /ʕ/, although they acquire /x/. However, no formal study has investigated these teachers’ observations to verify them.

Evidence From Other Studies

When evaluating evidence for a theory, whenever possible it is important to examine studies that were not specifically designed to test the theory; rather, previous studies can be reanalyzed, framed, and examined in light of the new theory. This process assures that no built-in biases influence an experiment specifically designed to test the theory. Obviously, the following studies were not designed to test the SDRH, because we put forth this hypothesis for the first time here. Because much research in L2 phonology involves no comparisons of stages, levels, or chronology, this research is difficult to use or evaluate the SDRH. However, the data from the following studies do involve levels or longitudinal stages and thus we can reinterpret them to evaluate our hypothesis.
Flege and Hillenbrand (1984) found two groups of American speakers of French produced /ü/ authentically (as determined by acoustical measurements) but /u/ unauthentically. Hence, the rate of acquisition for /ü/ was faster than for /u/, and in fact was NS-like for both groups. In labeling tasks, both groups produced /ü/ equally well, but for /u/ the experienced group outperformed the less experienced group. The results of the acoustical measurements and labeling tasks taken together suggest that acquisition of /ü/, the dissimilar sound, was faster than of /u/, the similar sound.

Major’s (1986) longitudinal study of 4 American English NSs acquiring the two kinds of Spanish rs, the flap /ɾ/ and trill /ɾ/, showed some patterns predicted by the SDRH. Many of the learning graphs were jagged up and down, and the first 2 learners (Speakers 1 and 2) showed virtually zero proficiency with the trill. This could mean that their acquisition of the trill had not yet begun (and that therefore these data neither support nor refute the SDRH). However, the other 2 learners (Speakers 3 and 4), did support our claim. In intervocalic position, Speaker 3 showed a slight increase in the trill (the dissimilar sound) but an overall decrease in the flap (the similar sound); in many instances she actually substituted the trill for the flap (overgeneralization). In intervocalic position, Speaker 4 showed a dramatic increase in proficiency in the trill but a lesser increase in the flap.

Flege (1987), in a study (somewhat similar to the 1984 study cited above) involving acoustic measurements of English NSs learning French /u/ and /ü/, found that neither the inexperienced nor the experienced speakers produced French /u/ authentically: The similar sound was difficult to learn. However, only the experienced speakers produced French /ü/ authentically. Flege’s data and interpretation therefore imply that the rate of acquisition for the dissimilar sound is higher than for the similar sound.

Major (1987b) investigated Brazilian Portuguese speakers learning English /ɛ/ and /æ/. The regression analyses of the two experiments, involving global foreign accent and judgments of /ɛ/ and /æ/ production, revealed that as global foreign accent decreased, proficiency in the dissimilar sound /æ/ increased rapidly,
but that proficiency in the similar sound /ɛ/ actually decreased dramatically, probably due to overgeneralization.

Bohn and Flege's (1992) study of German NS learners of English producing /i ɛ æ/ involved both acoustic measurements and labeling. The authors hypothesized that L2 English experience would not affect German NSs' productions of the “similar” vowels /i ɛ/ but that L2 experience would enable them to produce native-like English /æ/, the “new” (dissimilar) vowel. Their results from acoustic measurements largely support their hypothesis; however, the results of the labeling task for /æ/ and /ɛ/ did not support their model. Neither /ɛ/ nor /æ/ production approached NS-like norms in terms of labeling identification; furthermore, the inexperienced speakers performed better for /ɛ/ than the experienced speakers, but the experienced speakers performed better for /æ/ than the inexperienced speakers. However, if we recast their findings in terms of rate, the SDRH can account for them: Proficiency in the dissimilar sound increased but proficiency in the similar sound decreased. The SDRH makes no claim about positive or negative rate for similar phenomena; it may be positive, negative, or zero. It simply claims that the rate for similar phenomena will be slower than for dissimilar ones. A zero or negative rate is thus less than any positive rate exhibited by dissimilar phenomena. Therefore, both of Bohn and Flege’s experiments are consistent with the SDRH.

Oh’s (1996) longitudinal study of Korean NS learners of English investigated the effects of intense formal instruction for specific L2 sounds for a period of 1 month and 4 training sessions. Because Korean has /l/ but no /r/, English /r/ is the dissimilar sound and /l/ is the similar sound. Although the learners’ proficiency in the similar sound was superior to that in the dissimilar sound even at the end of the last session, the rate of improvement was greater for the dissimilar sound: /r/ improved steadily throughout the 4 sessions—22%, then 46%, then 50%, then 77% accuracy; however, /l/ showed very little change—95%, then 89%, then 95%, then 95% accuracy. Thus, Oh’s results support the SDRH.2
Summary and Implications

Previous studies investigating the role of markedness, similarities, and differences in L2 acquisition have produced various claims and results. We have attempted to reconcile these conflicting hypotheses and disparate results by confining these notions to rate, rather than to ultimate achievement. We hypothesized that dissimilar phenomena are acquired faster than similar ones and, ceteris paribus, that markedness is merely a mediating factor affecting rate. Our data support the hypothesis; recasting and reanalyzing data from other studies gave additional supporting evidence.

Because similarity and markedness can have opposite effects on rate of acquisition, it is important to control for one factor in order to determine the effects of the other. If a high degree of markedness slows rate and dissimilarity speeds up rate, then a study comparing a marked but dissimilar phenomenon to an unmarked but similar phenomenon would isolate neither the effects of similarity nor the effects of markedness, because the two factors would act at odds. The ideal test for the SDRH would compare two phenomena, one similar and one dissimilar to an L1 phenomenon but both having the same degree of markedness. In this way, the degree of similarity would be the only variable; markedness would be constant. Likewise, an adequate test of the effects of markedness should hold the degree of similarity constant, allowing the degree of markedness to vary.3

Our claims remain to be tested by either longitudinal studies dealing with stages or cross-sectional studies of learners at different levels. Although we have limited our discussion to phonology, it should be possible to test our hypothesis in other areas of language learning, providing that researchers can adequately work out and rigorously define similarity/dissimilarity specifications and markedness relationships.
Markedness has been defined both in terms of frequencies and implicational hierarchies. In our present study of a fricative and affricate, both definitions apply, because the presence of an affricate implies the presence of a fricative and, accordingly, in the languages of the world, fricatives are necessarily more frequent than affricates. Defining markedness solely in terms of frequencies is the less restricted definition, because for other phenomena, frequencies in the languages of the world may not imply presence or absence in a particular language: Researchers generally claim that voiced obstruents are more marked than voiceless obstruents, because all languages have voiceless obstruents but not all have voiced obstruents. Although applying to classes of sounds, this principle does not necessarily apply to individual sounds: /b/ but not /p/ occurs in Arabic, although Arabic does have other voiceless obstruents.

The other results of Oh’s (1996) study, which included just two other sounds, /j/ and /c/, conformed to markedness considerations; that is, the less marked sound or position was acquired faster than the more marked. Specifically, final /c/ was acquired faster than final /j/ (45% vs. 23% improvement) and initial /j/ faster than final /j/ (55% vs. 23% improvement). Because Korean has /c/ and allophonic [j] (intervocally), this aspect of his study is not relevant to similarity differentials, because both Korean /c/ and [j] are similar to English /c/ and /j/.

Although to keep one factor as close to constant as possible is important, it may be impossible to keep it absolutely constant while varying the other. For example, do two L2 sounds ever have exactly the same degree of similarity to an L1 sound, or exactly the same degree of markedness, in terms of frequency definitions? (See Note 1.)

References


