SSLA, **29**, 407–448. Printed in the United States of America.
DOI: 10.1017/S027226310707026X

A DYNAMIC LOOK AT L2 PHONOLOGICAL LEARNING:

Seeking Processing Explanations for Implicational Phenomena

Pavel Trofimovich, Elizabeth Gatbonton, and
Norman Segalowitz
Concordia University

This study investigates whether second language (L2) phonological learning can be characterized as a gradual and systematically patterned replacement of nonnative segments by native segments in learners' speech, conforming to a two-stage implicational scale. We adopt a dynamic approach to language variation based on Gatbonton's (1975, 1978) gradual diffusion framework. Participants were 40 Quebec Francophones of different English proficiency levels who produced 80 tokens of English /ð/ in eight phonetic contexts. In Analysis 1, production accuracy data are subjected to implicational scaling, with phonetic contexts ordered solely by a linguistic criterion—sonority hierarchy. In Analysis 2, the production accuracy data are similarly analyzed but with phonetic context ordering determined by psycholinguistic (processing) criteria—cross-language perceptual similarity and corpus-based estimates of lexical frequency. Results support and extend Gatbonton's framework, which indicates that L2 phonological learning progresses gradually, conforming to an implicational scale, and that perceived cross-language similarity and lexical frequency determine its course.

This research was made possible through grants to Pavel Trofimovich, Norman Segalowitz, and Elizabeth Gatbonton from the Social Sciences and the Humanities Research Council of Canada (SSHRC) and support from the Centre for the Study of Learning and Performance at Concordia University. The authors gratefully acknowledge the assistance of Melanie Barrière and Randall Halter in all aspects of data collection and analysis. Many thanks are extended to Dawn Cleary, Winnie Grady, Eva Karchava, Nootan Kumar, Magnolia Negrete Cetina, and Alin Zdrite for their help in various stages of this study. The authors wish to thank Tracey Derwing and Murray Munro for sharing their speech elicitation materials. Sarita Kennedy, Randall Halter, and five anonymous *SSLA* reviewers provided helpful suggestions on earlier drafts of this manuscript.

Address correspondence to: Pavel Trofimovich, Department of Education/TESL Centre, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; e-mail: pavel.trofimovich@concordia.ca.

How do second language (L2) learners learn to perceive and produce the sounds (phonology) of their L2 like native speakers (NSs) of that language? This question might initially seem straightforward, but it has no simple answer. The following example, an excerpt from an audio recording of a Francophone learner of English describing a picture story, illustrates the complexity involved in learning the sound system of a L2.

> *Th*e /də/ story is uh beginning in Manhattan. It's uh wi*th* /wɪdə/ uh a lady and a man who were walking and *th*ey /dej/ had a similar suitca- suitcase. *Th*ey /dej/ hurts *th*ey /dej/ hurt uh each o*th*er /ʌdəɹ/ and after apologize, *th*ey /ðej/ took *th*e /də/ wrong suit- suitcase and *th*ey /dej/ continue *th*eir /dejɹ/ way. At *th*e /də/ hotel, *th*e /ðə/ guy realize *th*at /dət/ he wa- he w- he took *th*e /də/ suitcase of *th*e /ðə/ lady and *th*e /də/ lady realize at her office *th*at /dət/ she took *th*e /də/ man's suitcase.

Even to a casual listener, this learner's speech reveals several generalizations about his production of English /ð/, a difficult consonant for Francophone learners of English (Jamieson & Morosan, 1986). First, the learner's L2 speech is variable. In some instances (*they took*, *the guy*), the learner produces the English /ð/ as a /ð/ and in others (*each other, took the suitcase*) as a /d/. Moreover, even when the same word is used (e.g., *they*), the learner produces English /ð/ as a /ð/ on one occasion and as a /d/ on another (*they took*, *they continue*). Second, the learner's native language (L1) plays an important role in L2 phonological learning. Because /ð/ is not a part of the consonant inventory of French (Fougeron & Smith, 1999), this learner (a NS of Quebec French) likely substitutes the closest L1 consonant, French /d/, for English /ð/ (we will present experimental evidence supporting this claim; see also Wenk, 1979, for evidence that speakers of European French often substitute /z/ for English /ð/). Finally, the nature of input appears crucial in L2 phonological learning. This learner might have been exposed to accented (non-native) English speech or have acquired English without much exposure to the spoken language (i.e., through reading and translation). Taken together, these generalizations define the goal of the present study: to develop a descriptive framework of adult L2 phonological learning, an elaboration of a gradual diffusion framework (referred to as the gradual diffusion model in Gatbonton, 1975, 1978), with a view to explaining the roles of learners' L1 and their L2 input and clarifying the sources of variability in L2 phonological learning.[1]

THE DYNAMIC PARADIGM

The descriptive framework used in the present study is based on the dynamic approach to describing language variation and change (see Preston, 1996, for review) first articulated in Bailey's (1973a) wave model. Most work in this area has focused on historical change, variation between language communities and between individuals, and, to a lesser extent, on intraindividual variation.

Underlying Bailey's model is the assumption that speech variability in a language user reflects a gradual spread of a speech pattern over time. In this model, a given speech pattern—for example, the raising of /æ/ to /ɛ/ in words like *January*, *had*, or *hatch* by speakers of American English from New England states (Labov, 1971)—first emerges in the speech of one speaker group. It then appears in the speech of another group, with changes in speech patterns proceeding in a wavelike fashion between socially and geographically proximitous groups (Bailey, 1973b).

The earliest applications known to us of the dynamic approach to language variation and change are in studies of synchronic language change in creole language communities (Bickerton, 1971, 1975; DeCamp, 1971). In these studies, language change in the community was seen as nonrandom. For example, in Bickerton's study of Guyanese English Creole, the coexistence of several forms with the same function (e.g., infinitive *tu* and *fu* markers) could be placed on a continuum ranging between creole features (basilectal) and standard English features (acrolectal), as determined by the linguistic (and often geographical or social) contexts in which such forms occurred. Two properties of such continua are noteworthy. First, the linguistic contexts in which creole forms occurred are arranged in a systematic way, rank-ordered from contexts more favorable to change (called *heavy*) to those that are less so (called *light*). Second, the distribution of creole forms along the continuum is implicational, such that the presence of a particular form in one context (e.g., the use of *tu* after so-called psychological verbs) implies the presence of that feature in a higher ranked, heavier context (i.e., after modal verbs). Seen dynamically, a linguistic change along this continuum (e.g., the replacement of *fu* by *tu* in all contexts of its use) starts in the heaviest context (after modal verbs) and propagates, in a wavelike fashion, to each consecutive lighter context. In this change process, individual speakers or speaker groups are assigned to particular stages along a continuum of those who are not yet participating in the change, those who are undergoing it, and those who have completed it.

THE DYNAMIC PARADIGM AND L2 DEVELOPMENT

If the dynamic paradigm can be used to characterize language change in creole communities (Bickerton, 1971, 1975) and L1 speaker groups (Bailey, 1973b; Chen & Wang, 1975), it can then be hypothesized that the same paradigm might apply to L2 development as well. Gatbonton (1975, 1978) adopted this hypothesis in designing a dynamic framework to describe variability in L2 learners' speech. In her original research, Gatbonton set out to determine whether learning L2 phonology can be conceptualized as the gradual replacement of nontarget (nonnative) segments by target (native) segments in learners' speech—this replacement resembling language change as conceptualized within the dynamic paradigm in creole studies (Bickerton). More specifically, Gatbonton (1975, 1978) proposed a two-stage framework to describe variability in L2

Table 1. A schematic representation of Gatbonton's gradual diffusion framework

Stage	Context 1	Context 2	Context 3
Acquisition phase			
Stage 1	0	0	0
Stage 2	01	0	0
Stage 3	01	01	0
Stage 4	01	01	01
Replacement phase			
Stage 5	1	01	01
Stage 6	1	1	01
Stage 7	1	1	1

Note. Target L2 forms gradually replace nontarget L2 forms across time (learning stages) in the direction from easy to difficult speech contexts (as indicated by the arrow). 0 indicates the presence of nontarget forms in the learner's L2 (e.g., /d/ for /ð/); 1 indicates the presence of target forms in the learner's L2 (e.g., /ð/); 01 indicates variation of target and nontarget forms in the learner's L2 (e.g., both /d/ and /ð/ for /ð/).

learners' speech (Table 1), a framework that departed from the dynamic paradigm of Bailey and Bickerton in one important way. In Bailey's and Bickerton's versions of the paradigm, the coexistence of alternate (variable) forms, caused by the entry of a new form into one environment, is resolved (i.e., the old form is replaced by the new form) before the new form begins its entry into another environment. In Gatbonton's (1975, 1978) version, however, the alternation of the old (nontarget) and the new (target) forms is allowed to occur across all environments before the new form begins to replace the old form.

This innovation resulted in a two-stage implicational framework. The first stage is the acquisition phase, in which target segments (e.g., English /ð/) first appear in L2 learners' speech and coexist with nontarget segments (e.g., /d/ used in place of /ð/) in all contexts of their use. The second stage is the replacement phase, in which target segments gradually supplant nontarget segments in all contexts of their use in the same order in which target segments originally appeared in those contexts. Although the idea of using implicational relations to describe speech phenomena is not a novel one (see, e.g., Trubetzkoy, 1939/1969, for early attempts), Gatbonton's (1975, 1978) innovation allowed the dynamic paradigm to address what is now known as Type 1 linguistic variation (Bayley & Regan, 2004; Mougeon, Rehner, & Nadashi, 2004). Type 1 variation refers to L2 learners' alternation between target and nontarget forms in the course of SLA (Adamson & Regan, 1991; Tarone, 1988; Young, 1991). By contrast, Type 2 variation refers to learners' alternation among socially appropriate target forms. Gatbonton's (1975, 1978) account of variation in L2 interlanguage is therefore a specific application of the dynamic paradigm to SLA (cf., e.g., Bailey, 1973a).

To test this framework, Gatbonton (1975, 1978) examined whether adult Francophone learners of English from Montreal would demonstrate the predicted pattern of (learning-driven) change in their production of English /ð/ from inaccurate forms (e.g., /ð/ produced as /d/) via a stage of variable performance (e.g., /ð/ produced as either /d/ or /ð/) to accurate forms (e.g., /ð/ consistently produced as /ð/). The accuracy data obtained for 21 of the 27 Francophone learners in her studies (analyzed from taped readings of two paragraphs as well as recorded spontaneous speech) patterned well along the predicted path, revealing a gradual learning-driven diffusion of target forms into the learners' L2. In particular, the direction of this learning diffusion was determined by the speech context immediately preceding the L2 segment (English /ð/), with contexts ordered from most vowellike (heavy) to most consonantlike (light) according to the sonority hierarchy (Clements, 1990; Jespersen, 1912). *Sonority hierarchy* refers to the classification of sounds according to the degree of their spontaneous voicing, with vowels and glides being the most sonorous and voiceless obstruents being the least sonorous. Thus, in the acquisition phase, the target forms first appeared in the learners' speech among nontarget forms (representing a stage characterized by variable performance) in heavy contexts—that is, contexts ostensibly most favorable to the target segment (i.e., /ð/ between two vowels, as in *either*). It was not until later in the learning process that the target forms appeared in lighter contexts, which were ostensibly less favorable to the target segment (e.g., /ð/ after a consonant, as in *ask the teacher*). Similarly, in the replacement phase, target forms first occurred consistently in one context and gradually replaced nontarget forms in all contexts, proceeding from most to least favorable.

Gatbonton's (1975, 1978) findings indicate that phonetic variability found in L2 speech is systematic (e.g., Cardoso, 2007; Dickerson, 1976), revealing the process by which L2 learners master new elements of the target language. Gatbonton's findings also suggest that the dynamic paradigm can successfully capture this learning process within the three dimensions of an implicational framework: time, context difficulty, and degree of phonetic variability. As such, the dynamic paradigm can exemplify what Meisel, Clahsen, and Pienemann (1981) argued to be the preferred approach to describing L2 development. According to Meisel et al., what best describes L2 development is not the degree to which L2 forms differ from the target L1 norm but, rather, the state of the learner's language at a given time, characterized by the emergence of new forms in the learner's L2 and, consequently, by the extent to which these forms are variable (see Dinnsen, 1984, and Leonard, Newhoff, & Mesalam, 1980, for a similar claim in L1 phonological development).

MOTIVATIONS FOR THE CURRENT STUDY

Recognized as a valuable approach to describe L2 development, the dynamic paradigm and the use of implicational scales as a methodological tool have found their due place in SLA research (see Rickford, 2002, for review). However, there exist at least two research gaps that need to be filled. The first research gap is that aside from Gatbonton's (1975, 1978) studies, there have been very few applications of the dynamic paradigm to the study of L2 phonological learning (see Amastae, 1978, and Nagy, Moisset, & Sankoff, 1996, for rare exceptions). Indeed, virtually all investigations of L2 development (or attrition) using the dynamic paradigm have focused on L2 morphology and syntax (Andersen, 1978; Bayley, 1999; Hyltenstam, 1977; Pienemann, 1998; Pienemann & Mackey, 1993; Politzer, 1976; Trudgill, 1986). The first objective of the present study, therefore, is to again apply the dynamic paradigm to describing L2 phonology and test the assumptions underlying Gatbonton's original framework with a larger dataset.

The second research gap relates to the challenge of offering psycholinguistic (processing) accounts of variability and its role in L2 development. Articulated by researchers investigating SLA from both psycholinguistic (Meisel et al., 1981; Pienemann, 1998) and sociolinguistic, variationist perspectives (Preston, 1996, 2000; Tarone, 2002), this challenge entails the need to offer psycholinguistic explanations for the variable and implicational nature of the phenomena described within dynamic accounts of L2 learning. One attempt to address this challenge, albeit exclusively in the realm of morphology and syntax, is found in Pienemann's processability theory (Pienemann; see also Young, 1991, for a prototype explanation of variation compatible with connectionist literature). Drawing on psycholinguistic principles of language processing and learning (e.g., Kempen & Hoenkamp, 1987; Levelt, 1989), Pienemann outlined a number of processing skills (e.g., procedures responsible for access to a word's syntactic information or generation of phrase structure) and posited a hierarchical, implicational relationship among them. He argued that the emergence of new forms in the learner's L2 and the degree to which such forms are variable are determined by the processing skills available to the learner in each learning stage. According to this view, L2 learning entails "the acquisition of the *skills* needed for the processing of the language" (Pienemann, p. 39), and the nature of each skill at each learning stage determines a specific patterning of L2 forms, of the type revealed in dynamic investigations of L2 morphosyntactic learning (Hyltenstam, 1977; Pienemann & Mackey, 1993).

Although Pienemann's (1998) processability hierarchy might be revealing of L2 morphosyntactic learning, the processing operations outlined in it do not apply to learning L2 phonology. To date there have been few attempts at a psycholinguistic account of the dynamic and variable nature of L2 phonological learning. A rare exception is Escudero and Boersma's (2004) optimality theoretic account of L2 phonological learning (discussed in more detail in the General Discussion section) compatible with psycholinguistic approaches to language development. Hence, the second objective of the present study is to add a psycholinguistic (processing) dimension to the dynamic framework of L2 phonological learning investigated here. This dimension is exemplified by two factors—cross-language similarity, a perceptual measure of L1-L2 distance (e.g., Guion, Flege, Akahane-Yamada, & Pruitt, 2000), and lexical frequency, a measure of input richness (e.g., Bradlow & Pisoni, 1999); both of these factors will be discussed in subsequent sections.

With the overall goal of developing a dynamic framework of L2 phonological learning, one that includes a psycholinguistic (processing) dimension, the present study investigates the accuracy with which adult Francophone speakers of L2 English produce English /ð/. To address the first objective—that of testing whether L2 phonological learning can be characterized as a systematic diffusion of target forms into learners' L2—speakers' accuracy data are subjected to implicational scaling and are tested against the predictions of Gatbonton's (1975, 1978) original framework. This is reported as Analysis 1. To address the second objective—that of adding a processing dimension to Gatbonton's framework—the obtained accuracy data are again subjected to implicational scaling to determine if the diffusion of target forms into learners' L2 is predicted by two psycholinguistic factors: the perceptual similarity between English /ð/ and segments in the learners' L1 (cross-language similarity) and the frequency with which English /ð/ occurs in spoken language (lexical frequency). This is reported as Analysis 2.

THE CURRENT STUDY

Participants

The participants in this study were 40 adult Francophones (27 female, 13 male) from Quebec (mean age: 35.6, range: 18.1–61.0). All were NSs of Quebec French and were born and raised in Montreal (n = 6) or in Granby, Quebec (n = 34), in homes where only French was used. All participants had received primary and secondary education in French in Quebec. With the exception of two (whose first exposure to some English occurred between birth and age 2 through interaction with an English-speaking parent), the participants started learning English as children at an average age of 9.3 as part of primary English as a second language (ESL) instruction in Quebec. Participants had studied English on average 60 min per week in elementary school, 90 min per week in high school, and 180 min per week in junior college. Five adult English NSs (two male, three female) from Quebec (mean age: 29), all born and raised in monolingual homes, also participated for comparison purposes.

Prior to testing, participants rated their proficiency in speaking, listening, reading, and writing in English and French on a 9-point scale (1 = extremely poor; 9 = extremely fluent). The analysis of the French self-ratings yielded consistently high mean proficiency scores (8.4–9.0), which suggests that the participants estimated their ability in French at a NS level. The analysis of the English self-ratings revealed that the participants represented different ability levels in English, from beginning to advanced. The mean speaking, listening, reading, and writing self-rated scores in English were intermediate (5.5–6.7), ranging from a low of 1 to a high of 9. The participants also estimated their daily use of French and English on a 0–100% scale. The analysis of these self-ratings indicated that the participants, on average, used French 80% (30–100%) and English 20% (0–70%) of the time per day.

Although revealing, the participants' self-ratings might not have provided an accurate measure of their L2 proficiency. Therefore, an accent rating task was administered to obtain another measure of the participants' English proficiency. In this task, five English NS judges (mean age: 38.2; all exposed to English from birth) rated a story read by each participant. The first paragraph of the story recorded by each participant was used for this test. These speech samples (mean duration: 18 s) were randomly presented one at a time binaurally via a Platronics (DSP-300) stereo headset, and the judges were asked to rate the degree of nonnative or nativelike accent in each speech sample on a 9-point scale (1 = heavy nonnative accent; 9 = nativelike accent). An accent score was computed for each participant by averaging the five judges' accent ratings (interrater reliability: $\alpha = .96$). These scores ranged from 1.8 to 9.0, with a mean of 5.3 ($SD = 2.2$). This result shows that participants represented different ability levels (particularly with respect to their ability to speak English), from beginning to advanced.

Materials and Procedure

The materials used in the present study included a 440-word text adapted from a short story used by Gatbonton (1978). This reading (Appendix A) included 80 target tokens of the English voiced interdental fricative /ð/ distributed across eight phonetic contexts, with 10 tokens per context. The phonetic contexts were (a) intervocalic (e.g., *ano<u>th</u>er*, *to <u>th</u>e*), (b) following voiceless stop consonants /p, t, k/ (e.g., *at <u>th</u>e*, *seek <u>th</u>e*), (c) following voiced stop consonants /b, d, g/ (e.g., *wanted <u>th</u>e*, *shed <u>th</u>e*), (d) following voiceless fricatives/affricates (e.g., *produce <u>th</u>e*, *touch <u>th</u>em*), (e) following voiced fricatives/affricates (e.g., *achieve <u>th</u>is*, *stage <u>th</u>e*), (f) following nasal consonants /m, n, ŋ/ (e.g., *directing <u>th</u>e*, *on <u>th</u>e*), (g) following liquids /l/ and /ɹ/ (e.g., *tell <u>th</u>e*, *clear <u>th</u>at*), and (h) sentence-initially, following a pause (e.g., *<u>Th</u>e. . .*, *<u>Th</u>en. . .*). The complete list of the target tokens is presented in Appendix B.

Testing, conducted individually in a quiet location using a personal computer, lasted approximately 60 min. The experimenter, a balanced French-English bilingual research assistant, gave testing instructions in English, providing clarification in French. The participants performed several tasks as part of a larger research project; two of these tasks—a cross-language perceptual identification task and a reading task—are used in the current study. For all participants, the cross-language perceptual identification task was performed first, followed by the reading task. Although spontaneous speech production tasks (e.g., describing pictures or telling a simple story; Moyer, 1999) are indeed more ecologically valid because they provide speech samples most representative of natural speech, the reading task was chosen to obtain speech samples that were equivalent in terms of the degree of formality involved—known to influence L2 learners' performance (e.g., Cardoso, 2007)—and therefore maximally comparable across all participants. The reading task also permitted elicitation of precisely the same number of target segment tokens from each participant.

ANALYSIS 1: TESTING THE GRADUAL DIFFUSION FRAMEWORK

The objective of this analysis is to determine the participants' English /ð/ production accuracy and to examine, using implicational scaling, if their production accuracy patterns in a systematic manner, similar to the gradual diffusion documented by Gatbonton (1975, 1978). As mentioned earlier, aside from Gatbonton's (1978) study, only two other investigations have applied the dynamic paradigm to describe L2 phonological learning. In a small-scale study of Mexican-American Spanish-English bilinguals' acquisition of five English vowels, Amastae (1978) reported that seven of nine bilinguals displayed vowel production patterns that fit the implicational order of vowels (from easiest to most difficult): /o/, /æ/, /ʊ/, /ɛ/, /ɪ/. Analyzing variable phenomena in the speech of 20 young Anglophone Montrealers, Nagy et al. (1996) likewise used the dynamic paradigm to examine the rate of /l/ deletion in French personal pronouns *il(s)* and *elle(s)*. Although many data cells contained missing data and several participants' responses did not conform to the predicted pattern, Nagy et al. demonstrated a systematic ordering of pronominal contexts favoring /l/ deletion. The purpose of the current analysis is, therefore, to apply the dynamic paradigm again to L2 phonology, testing the assumptions underlying Gatbonton's original framework with a larger dataset.

Method

Materials and Procedure. The English /ð/ production accuracy of the 40 Francophone participants was analyzed in their performance on the reading task (Appendix A). Participants read the text twice and were recorded directly onto a computer using a Platronics (DSP-300) head-mounted microphone. The second (usually more fluent) reading was used in all further analyses. The 40 Francophone speakers' recordings (one per participant) were subsequently presented to 10 native English listeners (mean age: 36.5; range: 24–50) for global pronunciation accuracy judgment. The listeners, recruited from a pool of ESL teachers or teachers-in-training at a local university, had, on average, 5 years (range: 0–14) of teaching experience. Although all listeners had some knowledge of French and, in some cases, another language (e.g., Italian, German), all had been exposed to English from birth and spoke English natively.

The five English NSs' recordings along with seven randomly selected Francophone speakers' recordings from the original set of 40 were presented to another 10 native English listeners (mean age: 40.8; range: 27–63). Recruited from the same teacher pool, these listeners had, on average, 2 years (range: 0–8) of teaching experience. This second set of listener judgments was collected to determine English NSs' English /ð/ production accuracy and also to obtain a measure of interrater consistency using a small subset (seven) of the original Francophone speakers' recordings.

The recordings were presented to the listeners via a Platronics (DSP-300) stereo headset in two randomized lists in individual listening sessions, with an equal number of listeners assigned to each list. The listeners were each given a booklet containing copies of the reading (one per recording) in which the 80 instances of English /ð/ were highlighted (marked in boldfaced capital letters). The listeners were instructed to circle what they considered to be correct renditions of English /ð/ and to cross out incorrect ones. They were encouraged to make a binary decision only (right: sounds like a good English /ð/; wrong: does not sound like an English /ð/), without distinguishing various degrees of accuracy. The experiment was self-paced, and the listeners were allowed to listen to each recording, replay its segments, and change their responses as many times as they wished. Prior to listening, the listeners scored one practice reading to familiarize themselves with the procedure. With rare exceptions, all maintained an efficient scoring pace, making accuracy decisions without frequent replaying of text segments or changing of the ratings given. The listening session lasted between 3 and 3.5 h. Interrater reliability analyses comparing accuracy ratings within the first listener group (using the original 40 recordings) and within the two listener groups combined (using a subset of 7 recordings) yielded moderate to very high indexes (α range: .70–.99), which suggests that the listeners were consistent in their judgments.

Data Analysis. There are two dependent variables in this analysis. The first variable—mean number of target English /ð/ tokens produced—is a measure of the participants' overall production accuracy. This variable is calculated by averaging, for each participant, the number of English /ð/ tokens in each phonetic context (out of a possible total of 10) that were marked as correct (targetlike) by the 10 listeners. The second variable—variability score—is used in implicational scaling. This variable, which has three discrete values (1 = consistently right, 01 = variable, 0 = consistently wrong), is derived from the first dependent variable by applying the 80% accuracy criterion (Andersen, 1978; Rickford, 2002). As such, the participant's accuracy in a given context is scored as 1 (consistently right) if the mean accuracy rate in that context is at least 80% (i.e., no fewer than 8 out of 10 English /ð/ tokens on average marked as right). The participant's accuracy in a given context is scored as 0 (consistently wrong) if the mean accuracy in that context is lower than 20% (i.e., no more than 2 out of 10 English /ð/ tokens on average marked as right).

Consequently, the participant's mean accuracy rates that range between 30% and 70% targetlike in each given context are scored as 01 (variable).

Results

Production Accuracy. The first analysis examines participants' overall production accuracy. The English NSs produced English /ð/ accurately in all contexts (96–100% targetlike). By contrast, the Francophone speakers' English /ð/ accuracy was variable both within and between participants. Between participants, English /ð/ accuracy was high for some participants (81–100% targetlike, $n = 9$) but remained at intermediate (31–80% targetlike, $n = 14$) or low (0–30% targetlike, $n = 17$) levels for others. Similarly, within participants, English /ð/ accuracy was higher in some contexts than in others. A one-way ANOVA comparing the number of targetlike English /ð/ tokens produced by the participants in each context yielded a significant context effect, $F(7, 273) = 8.83$, $p < .001$. Follow-up Bonferroni tests showed that the participants were more accurate in the sentence-initial and in the voiced fricative/affricate context than in most other contexts (Figure 1).

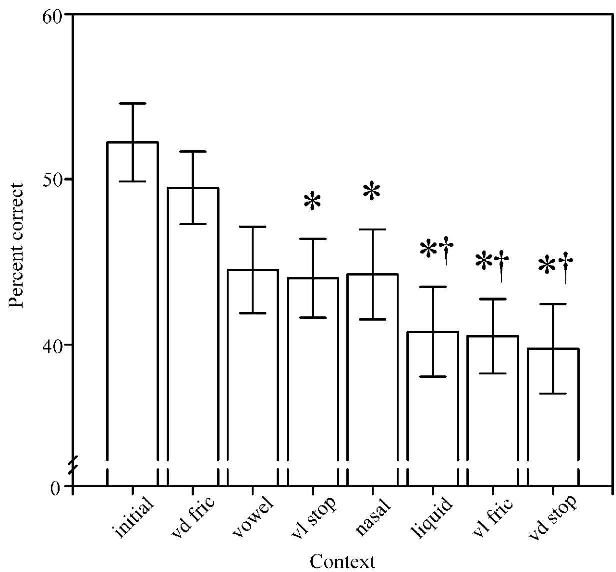

Figure 1. Francophone speakers' English /ð/ accuracy (± 1 *SE*) as a function of phonetic context. Asterisk indicates production accuracy that is significantly different from that in the sentence-initial context, $p < .001$; dagger indicates production accuracy that is significantly different from that in the voiced fricative/affricate context, $p < .001$.

Scaling Solution. The data were then subjected to implicational scaling to determine if English /ð/ production accuracy patterned in a manner predicted by the gradual diffusion framework, with phonetic contexts ordered according to the sonority hierarchy used by Gatbonton (1975, 1978) on the continuum between the sentence-initial and intervocalic contexts (+vocalic, +voice, +continuant) and the voiceless stop context (−vocalic, −voice, −continuant): sentence-initial, intervocalic, voiced fricative/affricate, liquid, nasal, voiced stop, voiceless fricative/affricate, voiceless stop.[2] The scaling solution evaluated in this analysis involves a data matrix representing 17 possible ways (represented by the 17 rows in Table 2), out of a much larger set of theoretically possible arrangements (3^8 = 6,561 to be exact), in which the target and nontarget renditions of English /ð/ can be distributed across eight phonetic contexts (Table 2). This matrix is identical to that described by Gatbonton (1975, 1978), with two exceptions. First, this matrix includes one phonetic context (sentence initial) not examined by Gatbonton. In the present analysis, this context is ranked the highest in the context hierarchy (and is

Table 2. Data matrix used in the implicational scaling analysis

Phonetic context								
1	2	3	4	5	6	7	8	Stage
0	0	0	0	0	0	0	0	1
01	0	0	0	0	0	0	0	2
01	01	0	0	0	0	0	0	3
01	01	01	0	0	0	0	0	4
01	01	01	01	0	0	0	0	5
01	01	01	01	01	0	0	0	6
01	01	01	01	01	01	0	0	7
01	01	01	01	01	01	01	0	8
01	01	01	01	01	01	01	01	9
1	01	01	01	01	01	01	01	10
1	1	01	01	01	01	01	01	11
1	1	1	01	01	01	01	01	12
1	1	1	1	01	01	01	01	13
1	1	1	1	1	01	01	01	14
1	1	1	1	1	1	01	01	15
1	1	1	1	1	1	1	01	16
1	1	1	1	1	1	1	1	17

Note. The dotted steplike lines separate cells with consistently nontarget (0), variable (01), and consistently target (1) renditions of the target segment.

therefore placed ahead of the intervocalic context) as it presumably contains the clearest, most prototypical instances of English /ð/. (See Analysis 2 for a principled way of context ranking.) Second, this matrix features two phonetic contexts (voiced fricative/affricate and nasal) previously collapsed by Gatbonton (1975, 1978) as part of two larger contexts: voiced continuant (+vocalic, +voice, +continuant) and voiced stop (−vocalic, +voice, −continuant), respectively.

Table 3 presents the implicational scaling matrix of the participants' English /ð/ accuracy data. In this matrix, each participant is assigned to the learning stage from Table 2 that is most similar to that participant's actual pattern. This analysis reveals a clear relationship between the participants' English /ð/ accuracy and the phonetic contexts in which English /ð/ was produced. Seven participants produced nontarget renditions of English /ð/ in all phonetic contexts consistently, according to the 80% rule, corresponding to learning stage 1 in the scaling solution (first row in Table 2). Some targetlike (albeit inconsistently so) renditions of English /ð/ featured in the speech of 11 other participants, first in higher ranked contexts and later in lower ranked ones (stages 2–8). Ten participants demonstrated variable English /ð/ production accuracy in all phonetic contexts (stage 9). Finally, the remaining 12 participants appeared to gradually replace variable renditions of English /ð/ with consistently targetlike ones, first in higher and then in lower ranked contexts (stages 10–17).

Although the obtained relationship between English /ð/ production accuracy and phonetic context clearly corresponds to the gradual diffusion pattern obtained by Gatbonton (1975, 1978), only 16 participants' data (40%) perfectly match the patterns predicted by the scaling solution (cf. a 78% fit in Gatbonton's original study). In fact, Guttman's (1944) index of reproducibility (IR), an estimate of implicational prediction in a scaling analysis (see also Dunn-Rankin, 1983; Torgerson, 1958), computed for this matrix (IR = 1 − total number of errors/total number of opportunities for error) yields only a moderate index of .88, which is below both the .93 value accepted as being an approximate of a .05 significance level (Dunn-Rankin, 1983; Hatch & Lazaraton, 1991) and the .93 index retrospectively obtained by analyzing the data from Gatbonton (1978).

One explanation for the discrepancy between the findings obtained here and the results reported by Gatbonton (1975, 1978) might relate to the fact that the present scaling analysis involves a larger, more refined matrix than the one used by Gatbonton. To test this possibility, the obtained data were fit into a matrix identical to that used by Gatbonton (1975, 1978); that is, this matrix excluded the sentence-initial context and combined the liquid and voiced fricative/affricate contexts as well as the voiced stop and the nasal contexts into two: voiced continuant and voiced stop contexts, respectively. The scalability of the resulting dataset improved only slightly, yielding the IR value of .91 (still short of the benchmark .93), with 26 participants' data (60%) patterning according to the framework's predictions.

Table 3. Implicational scaling of English /ð/ accuracy with context ordered according to the sonority hierarchy

	Phonetic context								
Subject	Initial	Vowel	Vd fric	Liquid	Nasal	Vd stop	Vl fric	Vl stop	Stage
31	0	0	0	0	0	0	0	0	1
18	0	0	0	0	0	0	0	0	1
29	0	0	0	0	0	0	0	0	1
4	0	0	0	0	0	0	0	0	1
20	0	0	0	0	0	0	0	0	1
21	0	0	01*	0	0	0	0	0	1
3	0	0	01*	0	0	0	0	0	1
19	01	0	01*	0	0	0	0	0	2
12	01	0	01*	0	0	0	0	0	2
14	01	0	0	0	0	0	0	01*	2
10	01	0	0	0	0	0	0	0	2
7	01	01	01	0	0	0	0	01*	4
13	01	0*	01	0	0	0	0	01*	4
6	01	0*	01	01	0	0	0	01*	5
23	01	0*	01	0*	01	0	0	0	6
32	01	0*	01	0*	01	0	0	01*	6
35	01	0*	01	01	01	01	01	0	8
1	01	01	01	0*	01	0*	01	0	8
30	01	01	01	0*	01	0*	01	01	9
37	01	01	01	0*	01	0*	01	01	9
36	01	01	01	0*	01	01	01	01	9
5	01	01	01	0*	0*	0*	01	01	9
11	01	0*	01	0*	01	0*	01	01	9
34	01	01	01	0*	01	0*	01	01	9
38	01	01	01	01	01	01	01	01	9
26	01	01	01	01	01	01	01	01	9
8	01	01	01	1*	01	01	01	01	9
2	01	01	1*	01	01	01	01	01	9
28	1	01	01	01	01	01	01	01	10
22	1	01	01	01	01	01	01	01	10
15	01*	1	01	01	01	01	01	01	11
25	1	1	1	1	1	1	01	1*	15
33	1	1	1	1	1	1	01	1*	15
17	1	1	1	1	1	1	01	1*	15
39	1	1	1	1	1	1	1	1	17
16	1	1	1	1	1	1	1	1	17
24	1	1	1	1	1	1	1	1	17
9	1	1	1	1	1	1	1	1	17
40	1	1	1	1	1	1	1	1	17
27	1	1	1	1	1	1	1	1	17

Note. Asterisks indicate cells that do not conform to the implicational patterns specified in Table 2. Vd = voiced; Vl = voiceless; Fric = fricative/affricate; Vowel = intervocalic.

Discussion

The objective of Analysis 1 was to determine if Francophone speakers' English /ð/ production accuracy would depend on the phonetic context in which English /ð/ was produced, patterning in a manner predicted by the gradual diffusion framework. Results of implicational scaling revealed that the speakers produced English /ð/ with variable degree of accuracy, as a function of phonetic context, and that target renditions of English /ð/ replaced variable and nontarget ones systematically, proceeding from the ostensibly easy sentence-initial and intervocalic contexts to the ostensibly difficult voiceless stop context.

Although the data reported in this analysis overall correspond well to the pattern reported by Gatbonton (1975, 1978), thus suggesting that Gatbonton's original data are replicable, several important discrepancies emerge between Gatbonton's studies and this one. In particular, the present analysis produced more data cells incompatible with the scaling solution and yielded a weaker goodness-of-fit measure (IR) than Gatbonton's analysis. These discrepancies might be attributed to several factors, of which methodological differences between the two studies are the most plausible. First, the participants in Gatbonton's studies were largely low-proficiency learners of English, whereas the participants in this study represented a wide range of proficiency levels, from beginning to advanced. In fact, 17 of 21 participants (81%) in Gatbonton's studies were placed in early learning stages, suggesting that what was modeled using implicational scaling represented only a portion of the possible variability space. Second, Gatbonton collected English /ð/ accuracy data in five phonetic contexts (as opposed to eight in this study), with an unequal number of tokens contributing to each participant's mean accuracy score in each context (e.g., 14 tokens in the intervocalic context vs. 4 tokens in the voiceless fricative/affricate context). A restricted range of contexts coupled with uneven token-to-context ratios might have rendered some contexts more or less difficult than they otherwise would be had a more refined selection of contexts and their token composition been utilized.

The present study sidestepped these methodological shortcomings and demonstrated, as did Gatbonton's (1975, 1978) studies, a clear and systematic patterning of variability in Francophone speakers' production of English /ð/, in line with gradual diffusion framework's predictions.[3] The obtained patterning was not perfect, however, which suggests that factors other than those related to carefully defining phonetic contexts and their token makeup might determine the degree to which English /ð/ accuracy fits the variability pattern predicted by the gradual diffusion framework.

One important consideration not given sufficient attention in either Gatbonton's (1975, 1978) original studies or in the current analysis is the nature of the implicational relations signaled by the phonetic contexts rank-ordered from easy (ostensibly, sentence-initially) to difficult (after a voiceless stop). In Gatbonton's original studies (and in the present analysis), a predominantly linguistic criterion (sonority hierarchy) was used to place contexts on a continuum from most vowellike to most consonantlike. Underlying this criterion is the assumption that English /ð/ can be perceived more clearly (and thus, ultimately, be produced more accurately) when it occurs in a vocalic context than when it is found in a consonantal context, where it is more likely to be coarticulated with preceding and following consonants (Manuel, 1995). Although this criterion is based on a well-established linguistic hierarchy (Clements, 1990), it is plausible that this hierarchy, at least to a certain extent, does not reflect the degree to which a L2 consonant is perceptible and ultimately learnable. In other words, there might not be a clear-cut correspondence between a context's ranking on the sonority continuum and the actual psycholinguistic (processing) demands on listeners exposed to English /ð/ in that context.

Thus, one way to clarify whether L2 learners' production accuracy follows the systematic patterning expected within the gradual diffusion framework would be to identify a principled, psychologically motivated criterion for defining phonetic context difficulty. The development of such a criterion need not be viewed as undermining the usefulness of the framework under investigation here. As discussed previously, the value of the framework resides in its novel use of implicational relations to describe variability in L2 phonological learning, not necessarily in its choice of sonority hierarchy as the criterion for rank-ordering phonetic contexts. Seen from this viewpoint, then, seeking new, more appropriate criteria for defining phonetic context difficulty within the gradual diffusion framework represents what Gregg (1996; see also Atkinson, 1982) termed the development of a sequence criterion (i.e., the identification and description of psychologically plausible accounts of developmental sequences in SLA). The goal of Analysis 2 is thus to develop a psycholinguistic, processing-based criterion for defining phonetic context difficulty within the framework and to examine if this new criterion for establishing implicational relations would yield a better scaling solution for the English /ð/ accuracy data. At a more general level, the goal of Analysis 2 is to add an explanatory dimension to the descriptive gradual diffusion framework, a necessary step in the development of any theoretical construct (Jordan, 2004).

ANALYSIS 2: DEVELOPING PSYCHOLINGUISTIC EXPLANATIONS

Cross-Language Similarity

The objective of Analysis 2 is to identify a processing-based criterion for describing context difficulty and to subject the accuracy data to implicational scaling using this criterion. One processing dimension that might be helpful in defining context difficulty in L2 learning is perceived cross-language similarity. Cross-language similarity refers to how perceptually similar the segments in the learner's L1 and L2 are. There is evidence that the degree of perceived dissimilarity (or similarity) between L1 and L2 segments might determine how L2 segments are perceived and produced (Baker & Trofimovich, 2005; Guion et al., 2000). For example, Japanese learners might produce English /ɹ/ more accurately than English /l/ (Flege, Takagi, & Mann, 1995) because they are more likely to perceptually differentiate English /ɹ/, but not /l/, from Japanese /ɾ/ (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004). In this situation, cross-language dissimilarity renders a L2 segment easier to learn. Japanese learners might also produce English /t/ more accurately than English /θ/ because they are more likely to perceptually equate English /t/, but not /θ/, with the similar Japanese /t/ (Guion et al., 2000). In this situation, although cross-language similarity aids in perceiving and producing a L2 segment, it might in fact render it harder to learn; that is, when a L2 segment is assimilated fully by a perceptually similar L1 segment, learners fail to perceive fine-grained phonetic detail in the L2 segment and are, therefore, unable to establish a nativelike category for this L2 segment (Flege, Schirru, & MacKay, 2003). Thus, L2 perception and production appear to depend on the perceived distance between L1 and L2 segments, such that (depending on the particular relationship between L1 and L2 segments) cross-language similarity can either help or hinder L2 perception and production.

If cross-language similarity determines how accurately L2 sounds are perceived and produced, this processing dimension can be used to develop a principled, perception-based criterion for defining the degree of a segment's contextual difficulty.[4] In the analysis that follows, estimates of cross-language similarity were obtained from the participants' judgments of perceptual similarity between tokens of English /ð/ and French consonants. The hypothesis was that instances of English /ð/ that are perceptually similar to a single French consonant would be difficult to perceive and, consequently, learn and, vice versa, that instances of English /ð/ that are perceptually dissimilar to any French consonant would be easy to perceive and learn (Guion et al., 2000; Trofimovich, Baker, & Mack, 2001; see Flege, 1995, for theoretical justifications).

Method

Participants. Cross-language similarity judgments were collected from all 40 participants described in Analysis 1; however, only the data from the participants with the lowest amount of contact with English were analyzed. This is because cross-language similarity judgments are predictive of the L1-L2 perceptual relationship only when such judgments are collected from inexperienced or, in the parlance of many researchers, naïve listeners (Best, McRoberts, & Sithole, 1988). Indeed, cross-language similarity judgments obtained from more experienced listeners (e.g., those with 10 or more years of L2 experience) likely reflect what these listeners already "know" about the L2 phonetic system rather than how they initially perceive L2 sounds with respect to the L1 phonetic system (Trofimovich et al., 2001).

Thus, included in this analysis were only the data from the 17 participants (10 female, 7 male) with the lowest amount of self-reported daily exposure to and use of English (below 25%). These participants (mean age: 41.7; range: 18.1–61.0) reported using English on average only about 5% daily (0–23%), estimating their daily use of French at a mean of 95% (80–100%). The participants' French self-ratings yielded consistently high mean proficiency scores (8.5–8.9 on 9-point scales); their English self-ratings (4.4.–5.8) and accent scores (1.8–5.2) were quite low.

Materials. The stimuli used in the cross-language identification task included 80 target and 15 control stimuli. The target stimuli were 80 English /ð/ tokens from the reading; the control stimuli were 15 French /s/ tokens from the French translation of the reading. It was hypothesized that the participants would be able to easily identify the consonant in the control stimuli as French /s/. Thus, the control stimuli were included to ensure that the participants understood the task and could perform it as intended.

The English stimuli were recorded by three male speakers of Canadian English from Montreal (mean age: 31), of whom two were balanced English-French bilinguals and one was a balanced English-Italian-French trilingual. All speakers were exposed to English from birth and spoke English natively.[5] The French stimuli were recorded by three male speakers of Quebec French from Montreal (mean age: 36); all were exposed to French from birth and spoke French natively. The speakers were asked to do the reading twice, speaking at a normal rate. A Platronics (DSP-300) head-mounted microphone was used to record the speakers directly onto a computer; *CoolEdit 2000* software (Syntrillium Software Corporation) was used to excise each stimulus from the second (more fluent) reading.

Because listeners' familiarity with particular lexical items in which the target sound is embedded might bias their cross-language similarity judgments (e.g., Pitt & Samuel, 1995), the stimuli were excised from the speech stream not as lexical items but as VC-CV, V-CV, or (in the case of sentence-initial /ð/) as CV syllables. For example, the English stimuli *dim the*, *above the*, *showed the*, *father*, and *Then* were excised as *-im_the-*, *-ove_the-*, *-owed_the-*, *-athe-*, and *the-*, respectively. Likewise, the French stimuli *à cette*, *assister*, *baisser*, and *d'essayer* were excised as *-à_ce-*, *-assi-*, *-aisse-*, and *-essa-*, respectively. The resulting 285 stimuli (95 tokens × 3 speakers) were ramped off during the first and last 15 ms to eliminate audible clicks and were normalized for peak intensity and perceived loudness.

The stimuli were organized into six balanced lists of 135 tokens, each with 15 tokens of English /ð/ in each of the eight phonetic contexts and 15 tokens of French /s/. Each of the three English speakers contributed five unique tokens of English /ð/ per context to every list. Each of the three French speakers contributed five tokens of French /s/ to every list. Across all lists, each token was spoken equally frequently by each of the three English or French speakers. The participants were assigned equally frequently to each list.

Procedure. The 135 tokens were randomly presented using speech presentation software (Smith, 1997) over stereo headphones. The participants performed a forced-choice cross-language identification and similarity rating task. They were asked to identify the consonant in each stimulus token with 1 of the 11 response alternatives (determined in pilot testing)—10 French consonants (/t/, /v/, /n/, /s/, /d/, /ʃ/, /z/, /f/, /l/, /g/) and a "not a French consonant" option—and to rate the similarity between them. The response alternatives were displayed in French orthography as *t*, *v*, *n*, *s*, *d*, *ch*, *z*, *f*, *l*, *g*, and *pas français* "not French," respectively.

The procedure was as follows. First, each participant heard a stimulus token (e.g., *-athe-*, extracted from *father*) and then selected, by clicking the appropriate button on the computer screen, the French consonant that sounded most similar to the consonant in the token (e.g., French /d/). Next, the participant heard the same stimulus token again but this time rated the similarity of the consonant it contained to the chosen French consonant on an 11-point Likert-type scale (1 = *pas du tout similaire* "not at all similar" to 11 = *très similaire* "very similar"). The participants had unlimited time to identify the consonants and to provide similarity judgments but were not permitted to change their responses. Before testing, participants were given a 10-item practice session to familiarize themselves with the procedure.

Data Analysis. The cross-language identification responses were scored by computing how many times (out of 15) each participant identified English /ð/ tokens in a given context (or, in the case of the control stimuli, tokens of French /s/) with any of the 11 response alternatives. For example, the number of times each participant identified English /ð/ in the intervocalic context with French /d/ was tabulated as well as the number of times it was identified with each of the 10 other response alternatives. The similarity rating responses on the 11-point Likert scale were scored by computing the participant's mean rating of the similarity between each English /ð/ token in a given context and each chosen response alternative. For example, a participant who chose French /d/ as the most similar consonant to *-athe-* on 9 out of the 15 presentations of *-athe-* was assigned a mean similarity rating of 8.2 on the 11-point Likert scale for these nine responses. Mean cross-language identification and similarity rating responses were computed for the 17 participants across all English /ð/ tokens in the same phonetic context.

The resulting identification rates were compared against chance performance, namely choosing 1 out of 11 possible response alternatives (9%), which is equivalent to 1.4 out of a possible total of 15 responses. Thus, considered in the final analyses were only those cross-language identification responses (and similarity ratings associated with them) that were at least two standard deviations above this chance identification rate (24%, or 3.6/15 responses). The standard deviation value used in this calculation represents the standard deviation of the proportion associated with the chance identification rate (1/11), computed for a possible total of 15 responses for English /ð/ tokens in each phonetic context. Initial analyses of the participants' responses to French /s/ tokens (the control stimuli), revealed that they identified these tokens with French /s/ a majority of the time (73%) and rated them as being similar to it (7.8). These identification rates and similarity ratings (cf. Trofimovich et al., 2001) suggest that the participants understood the task and had performed it as intended.

Results

Cross-Language Identification and Similarity Rating. Table 4 presents mean cross-language identification rates and mean similarity ratings (shown in parentheses) for English /ð/ tokens spoken in eight phonetic contexts. In response to English /ð/ in the sentence-initial, liquid, intervocalic, and voiced stop contexts, the participants chose a single French response alternative (French /d/) a majority of the time (31–61%) and rated the similarity between English /ð/ and French /d/ with an average rating of 4.5–6.2. They identified English /ð/ in the voiceless stop and voiceless fricative/affricate contexts with another single French response alternative: French /t/ (44%; mean rating: 5.1) and French /ʃ/ (33%; mean rating: 5.5), respectively. In response to English /ð/ in the nasal context, participants chose two response alternatives, French /n/ and /d/, 32% and 44% of the time, respectively, and rated these choices with a mean rating of 6.0 and 6.2, respectively. By contrast, English /ð/ in the voiced fricative/affricate context evinced no above-chance identification responses; in response to these English /ð/ tokens, the participants chose among nine different response options 1–23% of the time.

Whereas these cross-language identification patterns likely reflect the various degrees of similarity between French consonants and contextual instances of English /ð/, a more accurate measure of cross-language similarity requires combining both identification and similarity-rating judgments. This combined measure would allow for distinguishing between cases in which two contextual instances of English /ð/ are identified with the same L1 consonant yet differ in their similarity rating. In such cases, there is potentially a difference in the degree of the so-called perceptual fit between English /ð/ tokens and the chosen L1 consonant that cross-language identification and similarity judgments, considered separately, fail to capture. As Table 4 indicates, one such case would be tokens of English /ð/ in the liquid and in the sentence-initial context. These were identified relatively equally often with French /d/ (38% and 31%, respectively), yet the match between French /d/ and English /ð/ in the liquid context was rated lower (4.5) than the match between French /d/ and English /ð/ in the sentence-initial context (5.8). Therefore, to estimate more accurately the L1-L2 perceptual relationship, the fit index—a measure first used in a similar analysis by Guion et al. (2000) to combine both identification and similarity judgments—was applied to the data.

Table 4. Mean proportion of times that tokens of English /ð/ in eight phonetic contexts were identified with French response alternatives (with similarity ratings in parentheses)

English	French response alternatives										
/ð/ context	/t/	/v/	/n/	/s/	/d/	/ʃ/	/z/	/f/	/l/	/g/	Not French
Initial	.15 (3.7)	.09 (6.6)	.04 (7.2)	.01 (1.3)	**.31** (5.8)		.01 (9.0)		.15 (7.7)		.24 (1.1)
Vowel	.22 (3.4)	.02 (5.3)	.02 (2.5)	.01 (4.5)	**.45** (4.9)	.01 (5.5)	.05 (6.2)		.06 (7.4)	.01 (9.0)	.16 (1.2)
Vd fric	.07 (2.9)	.11 (7.2)		.11 (5.8)	.22 (5.4)	.06 (5.2)	.23 (6.2)		.01 (3.0)	.08 (6.0)	.11 (1.2)
Liquid	.16 (3.8)	.08 (4.9)	.01 (4.0)	.03 (5.8)	**.38** (4.5)		.11 (5.8)		.07 (4.6)	.01 (6.0)	.16 (1.1)
Nasal	.04 (3.0)	.02 (5.0)	**.44** (6.0)		**.32** (6.2)		.01 (1.0)		.02 (5.0)	.05 (5.3)	.10 (1.1)
Vd stop	.11 (3.6)	.04 (6.0)	.02 (3.5)	.01 (8.7)	**.61** (6.2)		.04 (6.3)		.01 (2.0)		.14 (1.3)
Vl fric	.22 (4.8)		.01 (2.0)	.21 (5.3)	.05 (5.2)	**.33** (5.5)	.01 (5.8)	.03 (6.6)	.02 (5.5)		.11 (1.6)
Vl stop	**.44** (5.1)	.01 (8.0)	.04 (7.2)	.06 (6.0)	.18 (5.1)	.02 (5.8)	.02 (6.5)	.02 (7.8)	.02 (3.8)	.02 (4.8)	.22 (1.4)

Note. Boldfaced values represent the identification rates that are +2 *SD* above chance performance. Vd = voiced; Vl = voiceless; fric = fricative/affricate; Vowel = intervocalic..

Table 5. Selected French consonants, identification rates, similarity ratings, perceptual fit indexes, log-based frequency estimates, and difficulty ranks for English /ð/ tokens in eight phonetic contexts

English /ð/ context	French consonants	Identification rate	Similarity rating	Fit index	Log frequency	Rank
Vd fric	—	—	—	0.00	2.0	1
Liquid	/d/	.38	4.5	1.71	2.9	3
Initial	/d/	.31	5.8	1.80	3.5	2
Vl fric	/ʃ/	.33	5.5	1.82	1.7	6
Vowel	/d/	.45	4.9	2.21	3.3	4
Vl stop	/t/	.44	5.1	2.24	2.4	5
Nasal	/n/	.44	6.0	2.64	1.6	7
	/d/	.32	6.2	1.98		
Vd stop	/d/	.61	6.2	3.78	0.6	8

Note. Vd = voiced; Vl = voiceless; fric = fricative/affricate; Vowel = intervocalic.

Fit indexes between English /ð/ in a given context and a given French consonant were computed by multiplying the mean above-chance rate with which English /ð/ was identified with this French consonant by the mean similarity rating between English /ð/ and this same French consonant. For example, the fit index between English /ð/ in the intervocalic context and French /d/ was calculated by multiplying the .45 identification rate by the 4.9 similarity rating, yielding a value of 2.21. The fit indexes for each phonetic context, ranging from 1.71 to 3.78, appear in Table 5, along with the French segments that were statistically significantly associated with each context. (The percentage selection and mean similarity ratings for these items, first reported in Table 4, are for convenience presented again in Table 5.)

Defining Context Difficulty. The objective of this analysis is to establish patterns of cross-language similarity between English /ð/ and French consonants that would permit the prediction of perceptual difficulty of English /ð/ in each context. It was hypothesized that those English /ð/ tokens that are perceptually similar to one or more French consonants would be more difficult to learn and, conversely, those English /ð/ tokens that are perceptually dissimilar to any French consonant would be easier to learn. If high fit indexes reflect a strong perceptual match and low fit indexes reflect a weak perceptual match between English /ð/ and French consonants, then it is possible to rank-order English /ð/ tokens, occurring in each context, according to their predicted ease of learning (the weaker the perceptual match, the easier to learn).

Table 5 summarizes the perceptual match data needed to make differential predictions about the relative difficulty presented by the eight different phonetic contexts. The tokens of English /ð/ in the voiced fricative/affricate context show the least perceptual match to a French segment, suggesting that this should be the context in which the learning of the target segment should be easiest. In this context, English /ð/ was not identified with any French consonant at above-chance rates. The tokens with the greatest perceptual match to French segments, expected to be the most difficult to learn, appear to be tokens of English /ð/ in the voiced stop context, followed by those in the nasal context. In the voiced stop context, English /ð/ was heard predominantly as French /d/ (61% of the time, with a fit index of 3.78), exemplifying perceptual assimilation of English /ð/ to a single L1 category. In the nasal context, English /ð/ was heard as both French /d/ and /n/, representing perceptual assimilation of English /ð/ to two L1 categories varying in degree of fit. The latter (dual-category) assimilation pattern is reflective of a more easily discriminable (and hence learnable) contextual variant of English /ð/ than the former (single-category) assimilation pattern (Best, McRoberts, & Goodell, 2001; Best et al., 1988). English /ð/ in the remaining five contexts was heard as a single French consonant (/t/, /ʃ/, or /d/), with various degrees of fit to each. These five contextual variants of English /ð/ can be ordered according to their perceptual difficulty from easy (liquid context) to difficult (voiceless stop context) in increasing value of their perceptual fit to a single French consonant. The ranking of English /ð/ tokens in eight phonetic contexts, from easy to difficult, appears in the column labeled "Fit index" in Table 5.

Scaling Solution. The English /ð/ accuracy data reported in Analysis 1 were subjected to implicational scaling, with phonetic contexts ordered according to perceived cross-language similarity (i.e., in increasing degree of the English consonant's perceptual fit to a single French consonant: voiced fricative/affricate, liquid, sentence initial, voiceless fricative/affricate, intervocalic, voiceless stop, nasal, voiced stop). Again, the goal was to see whether the accuracy data patterned in a manner predicted by the gradual diffusion framework. As in the previous analysis, the matrix evaluated here involved a 17-stage scaling solution (see Table 6).

The analysis of the fit of the participants' accuracy data into the matrix again revealed a relationship between the participants' English /ð/ production accuracy and the phonetic contexts in which English /ð/ was produced, much akin to that obtained in Analysis 1. However, as in Analysis 1, only 16 participants' data (40%) perfectly matched the patterns predicted by the gradual diffusion framework, yielding an IR index of .88, which was lower than the benchmark .93. Apart from this relatively weak fit of the data to the predicted pattern, two findings are noteworthy. First, the participants' accuracy data in the voiced fricative/affricate, nasal, and voiced stop contexts—those representing the end points of the similarity-based context hierarchy—represented a good fit to the matrix (with only 2 cases out of 34, or 5.8%, not fitting the expected pattern). Second, the participants' accuracy data in the remaining five phonetic contexts (those in the middle of the hierarchy) yielded a relatively large number of discrepancies between the obtained patterns and those

Table 6. Implicational scaling of English /ð/ accuracy with context ordered by degree of cross-language similarity

	Phonetic context								
Subject	Vd fric	Liquid	Initial	Vl fric	Vowel	Vl stop	Nasal	Vd stop	Stage
31	0	0	0	0	0	0	0	0	1
18	0	0	0	0	0	0	0	0	1
29	0	0	0	0	0	0	0	0	1
4	0	0	0	0	0	0	0	0	1
20	0	0	0	0	0	0	0	0	1
10	0	0	01*	0	0	0	0	0	1
14	0	0	01*	0	0	01*	0	0	1
23	01	0	01*	0	0	0	01*	0	2
7	01	0	01*	0	01*	01*	0	0	2
32	01	0	01*	0	0	01*	01*	0	2
21	01	0	0	0	0	0	0	0	2
3	01	0	0	0	0	0	0	0	2
12	01	0	01*	0	0	0	0	0	2
19	01	0	01*	0	0	0	0	0	2
13	01	0	01*	0	0	01*	0	0	2
6	01	01	01	0	0	01*	0	0	4
5	01	0*	01	01	01	01	0	0	7
37	01	0*	01	01	01	01	01	0	8
30	01	0*	01	01	01	01	01	0	8
1	01	0*	01	01	01	0*	01	0	8
34	01	0*	01	01	01	01	01	0	8
11	01	0*	01	01	0*	01	01	0	8
35	01	01	01	01	0*	0*	01	01	9
36	01	0*	01	01	01	01	01	01	9
38	01	01	01	01	01	01	01	01	9
26	01	01	01	01	01	01	01	01	9
22	01	01	1*	01	01	01	01	01	9
15	01	01	01	01	1*	01	01	01	9
28	01	01	1*	01	01	01	01	01	9
8	01	1*	01	01	01	01	01	01	9
2	1	01	01	01	01	01	01	01	10
25	1	1	1	01*	1	1	1	1	17
33	1	1	1	01*	1	1	1	1	17
17	1	1	1	01*	1	1	1	1	17
39	1	1	1	1	1	1	1	1	17
16	1	1	1	1	1	1	1	1	17
24	1	1	1	1	1	1	1	1	17
9	1	1	1	1	1	1	1	1	17
40	1	1	1	1	1	1	1	1	17
27	1	1	1	1	1	1	1	1	17

Note. Asterisks indicate cells that do not confirm to the implicational patterns specified in Table 2. Vd = voiced, Vl = voiceless, fric = fricative/affricate, Vowel = intervocalic.

predicted by the gradual diffusion framework. These findings suggest that the participants' similarity judgments provide a good (albeit unrefined) measure of context difficulty, in that they differentiate well only among the so-called end-point contextual instances of English /ð/: those that appear to be the easiest and the most difficult.

Although this similarity-based criterion was overall successful in discriminating among the easiest and the most difficult contextual instances of English /ð/, it ostensibly failed to distinguish fine-grained differences in context difficulty. One psycholinguistic dimension that might help refine the criterion for defining context difficulty is lexical frequency. Indeed, even the most difficult of L2 segments (in terms of cross-language similarity) might effectively be easy to perceive and produce if they are sufficiently frequent in the input available to the learner (Flege, Takagi, & Mann, 1996). The next analysis attempts to develop a criterion for defining context difficulty based on both cross-language similarity and lexical frequency.

Lexical Frequency

Lexical frequency refers to the frequency with which individual lexical items occur in spoken or written language; it is often predictive of how rapidly or accurately both L1 (e.g., Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997) and L2 (e.g., Flege et al., 1996) segments are perceived and produced. For example, Flege et al. (1996) found that adult Japanese speakers identified /ɹ/ and /l/ tokens correctly more often when they occurred in words that were more frequent (and therefore more familiar) than their minimal pairs (i.e., when the /ɹ/ in *room* was paired with low-frequency *loom*, or the /l/ in *lip* was paired with low-frequency *rip*). In another study, Bradlow and Pisoni (1999) asked adult L2 learners to identify easy and hard words. Easy words were high-frequency words with few similar-sounding lexical neighbors (e.g., *work*, *long*, *both*), whereas hard words were low-frequency words with many lexical neighbors (e.g., *hoot*, *mace*, *moan*). The learners appeared less likely to accurately identify hard words than easy ones even when word familiarity was controlled, indicating that word frequency and neighborhood density (lexical factors characterizing perceptual difficulty of spoken words) effectively modulated learners' L2 perceptual accuracy. These findings demonstrate that speakers are sensitive to lexical factors, perceiving and producing segments more accurately and rapidly in frequent (familiar) than in infrequent (unfamiliar) words or syllables (e.g., Flege et al., 1996; Vitevitch et al., 1997). These findings are also compatible with the results of recent variationist studies that have revealed an important role of lexical frequency in determining variability in L2 interlanguage (Langman & Bayley, 2002; Regan, 1996) and are consistent with frequency-based models of L1 and L2 variation (Bybee, 2002; see Bybee & Hopper, 2001, for review).

Defining Context Difficulty. If lexical frequency indeed determines how accurately L2 segments are perceived and produced, then this processing dimension can help refine the similarity-based criterion for defining the degree of contextual difficulty for English /ð/. In this analysis, estimates of lexical frequency were obtained from the British National Corpus (2000) spoken frequency counts (total number of spoken texts searched: 775,799) individually for each of the 80 English /ð/ tokens used in this study.[6] The hypothesis was that those English /ð/ tokens that are more frequent in spoken input would be easier to perceive and produce than those that are less frequent.

The obtained spoken frequency counts were first subjected to a logarithmic (log) transformation, a procedure that normalizes non-normally distributed data, which is often the case with corpus-based frequency counts. Mean log-based frequency estimates (shown in Table 5) were computed separately in each phonetic context by averaging the obtained frequency counts for the 10 tokens in that context. In order to create a continuum of contextual instances of English /ð/ ranging from easy to difficult, based on both indexes of cross-language similarity and spoken frequency, the eight phonetic contexts were rank-ordered by fit index (a measure of cross-language similarity) using log-based frequency values (estimates of lexical frequency) as the weighting variable. This was done by dividing the fit index for each phonetic context by the log-based lexical frequency value for that same context, the assumption being that a low fit index and a high log value characterize contextual instances of English /ð/ that are easy to learn. This procedure yielded a continuum of contextual instances of English /ð/ ranging from ostensibly the easiest contexts (voiced fricative/affricate), indicated by a low value on the continuum, to the most difficult contexts (voiced stop), indicated by a high value on the continuum. This ranking appears in the rightmost column of Table 5.

Scaling Solution. The participants' English /ð/ production data reported earlier were then subjected to implicational scaling using this new criterion to order phonetic contexts from easy to difficult. An initial analysis reveals that one of the eight contexts (liquid) accounted for 33% of all discrepancies between the predicted and obtained patterns, a value that is manifestly greater than the error rates observed in the remaining contexts (9.6% on average). Before subjecting the resulting scaling solution to further evaluation, the decision was made to eliminate the data from the liquid context from further analyses (see the Discussion section for additional reasons for this decision). The final matrix evaluated in this implicational scaling analysis thus involves a data matrix that represents 15 possible ways in which the target and nontarget renditions of English /ð/ can be distributed across seven phonetic contexts (from easy to difficult): voiced fricative/affricate, sentence initial, intervocalic, voiceless stop, voiceless fricative/affricate, nasal, voiced stop (Table 7).

The scaling solution obtained in this analysis represents a good fit between the pattern predicted by the gradual diffusion framework and the obtained

Table 7. Implicational Scaling of English /ð/ Accuracy with Context Ordered by Degree of Cross-Language Similarity and Lexical Frequency

	Phonetic context							
Subject	Vd fric	Initial	Vowel	Vl stop	Vl fric	Nasal	Vd stop	Stage
31	0	0	0	0	0	0	0	1
18	0	0	0	0	0	0	0	1
29	0	0	0	0	0	0	0	1
4	0	0	0	0	0	0	0	1
20	0	0	0	0	0	0	0	1
21	01	0	0	0	0	0	0	2
3	01	0	0	0	0	0	0	2
14	0*	01	0	01*	0	0	0	3
10	0*	01	0	0	0	0	0	3
19	01	01	0	0	0	0	0	3
12	01	01	0	0	0	0	0	3
23	01	01	0	0	0	01*	0	3
13	01	01	0	01*	0	0	0	3
32	01	01	0	01*	0	01*	0	3
7	01	01	01	01	0	0	0	5
6	01	01	0*	01	0	0	0	5
5	01	01	01	01	01	0	0	6
1	01	01	01	0*	01	01	0	7
11	01	01	0*	01	01	01	0	7
34	01	01	01	01	01	01	0	7
37	01	01	01	01	01	01	0	7
30	01	01	01	01	01	01	0	7
36	01	01	01	01	01	01	01	8
35	01	01	0*	0*	01	01	01	8
38	01	01	01	01	01	01	01	8
26	01	01	01	01	01	01	01	8
15	01	01	1*	01	01	01	01	8
8	01	01	01	01	01	01	01	8
2	1	01	01	01	01	01	01	9
22	01*	1	01	01	01	01	01	10
28	01*	1	01	01	01	01	01	10
25	1	1	1	1	01*	1	1	15
33	1	1	1	1	01*	1	1	15
17	1	1	1	1	01*	1	1	15
39	1	1	1	1	1	1	1	15
16	1	1	1	1	1	1	1	15
24	1	1	1	1	1	1	1	15
9	1	1	1	1	1	1	1	15
40	1	1	1	1	1	1	1	15
27	1	1	1	1	1	1	1	15

Note. Asterisks indicate cells that do not confirm to the implicational patterns specified in Table 2. Vd = voiced; Vl = voiceless; fric = fricative/affricate; Vowel = intervocalic.

accuracy pattern, with 25 of 40 participants (63%) fitting perfectly into 1 of 15 patterns predicted by the framework. The obtained IR index, estimated at .94, appears to satisfy the minimum reproducibility value (.93) used in previous evaluations of implicational data (Dunn-Rankin, 1983; Hatch & Lazaraton, 1991). These descriptive indexes suggest that the obtained accuracy data clearly correspond to the gradual diffusion pattern obtained by Gatbonton (1975, 1978), such that the accurate productions of English /ð/ first appear in easy contexts (those in which English /ð/ is frequent and which render English /ð/ dissimilar to L1 consonants) and then gradually, via a stage of variable accuracy, spread into more difficult ones (those that are less frequent and render English /ð/ more similar to L1 consonants). However, before drawing more definitive conclusions about the accuracy of the obtained scaling solution, it is important to determine that the obtained pattern indeed represents one that was nonrandom.

The obtained data matrix was tested against the null hypothesis that predicts that there is in fact no systematic patterning in the distribution of target and nontarget instances of English /ð/ in L2 learners' speech. There are 3^7 or 2,187 ways in which the three values (0, 01, and 1) can be distributed across seven phonetic contexts (in a hypothetical matrix of 2,187 rows and seven columns). Assuming that values 0, 01, and 1 could occur with equal likelihood, then the probability that a given participant's English /ð/ accuracy would match by chance alone 1 of the 15 patterns in the gradual diffusion matrix would be $15/3^7 = .0069$ (approximately 1 participant in 160). A chi-square goodness-of-fit procedure used to test for a significant difference between the observed number of participants whose accuracy data matched the predicted patterns (25 of 40) and the number of participants whose accuracy was expected to match any of the predicted patterns by chance (.28 of 40), yielded a significant value, $\chi^2(1) = 2197.78$, $p < .00001$, suggesting that the obtained match to the matrix was nonrandom.

Although convincing, this analysis might not be completely accurate. This is because the chance probability of obtaining a pattern consistent with the predictions was calculated, under the null hypothesis, by assuming equal likelihood for each event (0, 01, 1). However, most participants might have a tendency to show variable English /ð/ production (01) in every context, rather than the 0 or 1 patterns exclusively, for reasons that have nothing to do with the systematic patterning presumed to underlie the gradual diffusion framework. This would therefore effectively inflate the participants' chances of producing speech patterns matching the implicational framework evaluated here, particularly because the 01 value is relatively more frequent than the 0 or 1 values. In other words, the participants might display a bias to perform in a way that increases their chances of appearing as though they were matching that small part of the large (2,187 × 7) set of potential patterns. The chance probability that takes into account this type of bias was estimated directly from the obtained data matrix by computing and summing over the marginal probabilities for obtaining each pattern. This computation yielded the probability value of .0242 for a given pattern (approximately 1 participant in 40), considerably higher than .0069 calculated under the equal-likelihood assumption. With this new estimate of the number of participants expected to match the predicted matrix by chance (1 in 40), the chi-square goodness-of-fit calculations yielded a statistically significant result, $\chi^2(1) = 590.77$, $p < .002$, indicating that the obtained pattern departed from what would be expected by chance.

Taken together, these analyses indicate that the distribution of target and nontarget renditions of English /ð/ cannot be attributed to chance alone. Instead, these renditions represent the systematic patterning presumed to underlie the obtained implicational scaling solution. At least one criticism of this conclusion, however, might relate to the fact that the statistical evaluation of the obtained data matrix was based both on the patterns characterized by variable performance (01 patterns in stages 2–14) and those characterized by nonvariable performance (0 and 1 patterns in stages 1 and 15, respectively). From the theoretical point of view, the latter two nonvariable patterns are the least relevant to testing the gradual diffusion framework, as this framework was designed to explain variability in L2 phonological accuracy. Also, from the statistical point of view, the inclusion of these nonvariable patterns might have increased the likelihood of the participants' data matching the predicted patterns by chance alone. To address this potential criticism, the obtained data matrix was subjected to a more conservative test, one that excluded nonvariable patterns. The evaluated matrix involved 13 variable patterns, with 13 of the 25 participants' English /ð/ accuracy data fitting one of these patterns. A chi-square goodness-of-fit procedure testing for a significant difference between the observed number of participants who matched the predicted patterns (13 of 25) and the number of participants expected to match the predicted patterns by chance (.60 of 25; sum of marginal probabilities: .0238) yielded a significant value, $\chi^2(1) = 262.58$, $p < .01$, suggesting again that the obtained match to the matrix was nonrandom.

Discussion

The goal of this analysis was to develop a processing-based criterion for defining context difficulty to be applied, in place of the sonority hierarchy criterion used by Gatbonton (1975, 1978), to scaling the accuracy of Francophone speakers' production of English /ð/. Results revealed that the criterion based on two measures—perceived cross-language similarity and lexical frequency—yielded the best scaling solution for the production accuracy data obtained in Analysis 1. Taken together, these findings indicate that the Francophone speakers produced English /ð/ with variable degree of accuracy, as a function of the phonetic context in which English /ð/ occurred. In these speakers' L2 speech, the target renditions of English /ð/ replaced variable and erroneous ones systematically, proceeding from easy contexts (those in which English

/ð/ is frequent and which render English /ð/ dissimilar to L1 consonants) to difficult ones (those that are less frequent and render English /ð/ similar to L1 consonants).

One exception to the obtained pattern of results—one that did not obey the implicational relations based on a perceptual measure of cross-language similarity and corpus estimates of lexical frequency—involved accuracy data in the liquid context (English /ð/ preceded by /l/ or /ɹ/). This context accounted for one third of all discrepancies between the predicted and obtained accuracy patterns, at a rate at least three times higher than in any other context. This finding was likely caused by inaccuracies in estimating the perceptual fit between English /ð/, as spoken in the liquid context, and French consonants in the cross-language identification task. In this task, 8 of the 10 tokens of English /ð/ were preceded by English /ɹ/; however, French /l/, but not French /ʁ/, featured among the viable response alternatives given to the participants. The exclusion of French /ʁ/ from the list of viable response alternatives (based on the results of pilot testing in which French /ʁ/ appeared infrequently among the selected response alternatives) might have thus biased the obtained index of perceptual similarity (fit index) toward French /l/, thereby concealing a true perceptual similarity relationship between French consonants and English /ð/ tokens heard in the context of /l/ or /ɹ/.

Accuracy data in the liquid context notwithstanding, the analyses presented here indicate that Francophone speakers' production of English /ð/ follows the dynamic and implicational pattern predicted by the gradual diffusion framework. More importantly, this pattern is clearly evident not only in the implicational scaling of accuracy data using trinary (three-valued) scales (0, 01, 1) but also in the implicational scaling of raw production accuracy data. In fact, some researchers have recently called for the use of raw accuracy data in implicational scaling (Rickford, 2002), arguing that data fitting into binary and trinary scales involves arbitrary decisions (e.g., the use of the 80% accuracy criterion to define variable accuracy) and thus conceals "vast extremes of variability" (p. 158).[7] Figure 2A presents the 40 Francophone speakers' English /ð/ production accuracy data (i.e., the same data scaled in Table 7 using trinary accuracy values) plotted as a function of phonetic context and learning stage. In this figure, each bar represents mean English /ð/ accuracy in a given phonetic context for all speakers assigned to a particular learning stage on the basis of the implicational scaling analysis using the cross-language similarity and lexical frequency criterion (see Table 7). These data closely correspond to the ideal hypothetical distribution of raw accuracy data as predicted by the gradual diffusion framework (Figure 2B).

The observed pattern is robust, clearly seen in a strong relationship between the speakers' overall accuracy (mean accuracy across all phonetic contexts) and the learning stage to which they were assigned based on the implicational relations among the phonetic contexts. Indeed, a Spearman rank-order correlation computed between the speakers' overall accuracy and the learning stage to which they were assigned yielded a near-perfect correlation, $r(38) =$

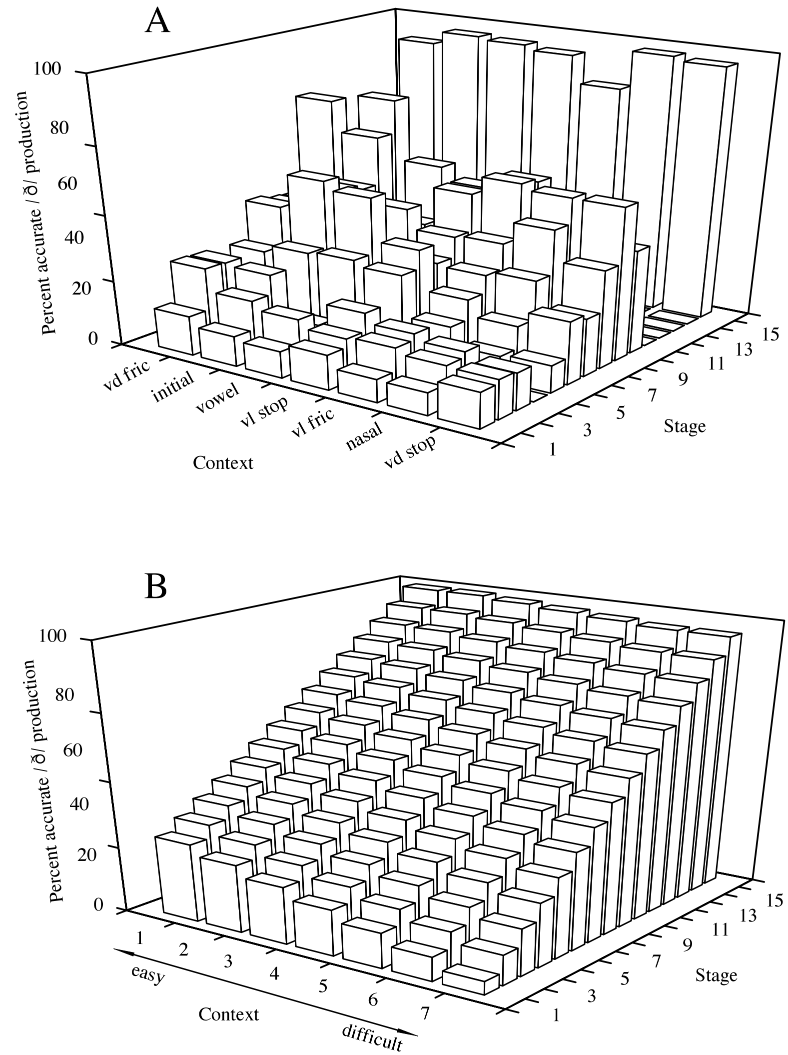

Figure 2. The 40 Francophone speakers' English /ð/ accuracy (A) and a hypothetical distribution of English /ð/ accuracy predicted by the gradual diffusion framework (B) plotted as function of phonetic context and stage.

.97, $p < .001$. This indicates that the learning stage to which the speakers were assigned based on the rank-ordering of the phonetic contexts from easy to difficult (using measures of cross-language similarity and lexical frequency) is strongly predictive of their overall English /ð/ production accuracy. Such a robust relationship would not have been possible had the speakers displayed random patterns of production accuracy in each phonetic context or, at a minimum, patterns not predicted by measures of cross-language similarity and lexical frequency.

GENERAL DISCUSSION

Conceptualized within the dynamic approach to describing language variation and change, the present study was conducted to accomplish two objectives: to test whether L2 phonological learning proceeds in a gradual and systematic way, at least at the level of phonetic segments, and to offer psycholinguistic (processing) explanations for the gradual and systematic nature of this learning. To address these objectives, the present study tested the assumptions underlying Gatbonton's (1975, 1978) gradual diffusion framework, a dynamic framework of L2 phonological learning.

The findings of the present study yield support for this framework of L2 phonological learning. They reveal that L2 learning indeed progresses in a gradual and systematic manner, and they identify at least two psycholinguistic factors—cross-language perceptual similarity and lexical frequency—that might determine its course. In particular, L2 phonological learning, defined within the framework as the diffusion of the target forms into learners' L2, appears to be predicted both by how similar L2 segments are to perceptually proximitous L1 segments and by how frequently, based on spoken corpus frequency counts, L2 segments occur in spoken language. Taken together, these factors help clarify how learners' L1 and the nature of L2 input interact to result in learning characterized either by targetlike speech or by variable accuracy.

To put evidence for the gradual diffusion framework of L2 phonological learning in perspective, it is important, of course, to test the framework in other contexts. In the present study, support for the gradual diffusion framework was found in a cross-sectional investigation of Francophone speakers' acquisition of English /ð/. Obviously, the framework should be extended to account for the acquisition of other segmental (and perhaps even suprasegmental) aspects of L2 phonology by Francophones as well as speakers from other L1 backgrounds. For example, the framework can, in principle, be extended to explain the variable nature of English /ɹ/-/l/ production by Japanese and Korean learners of English (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997) or English /t/-/d/ production by Chinese learners of English (Flege, Munro, & Skelton, 1992). The framework can also be extended to explain the variable (Type 2) production of L2 morphological elements (e.g., English *-ing*) by learners of English from various language backgrounds (e.g., Young, 1991). Furthermore, predictions based on the framework need to be tested by employing more ecologically valid speech elicitation procedures, such as role-plays or spoken narratives, by carrying out more fine-grained (acoustic and articulatory) analyses of speech production, and by using estimates of lexical (and phonological, segmental) frequency that will reflect learner input better than NS spoken corpus frequency counts. Finally, because the gradual diffusion framework addresses L2 phonological learning, it is essential to establish that its predictions are upheld in a situation in which learning occurs gradually over time. Thus, a longitudinal study will be not only crucial for the development of the framework but also relevant to the field of L2 phonological learning, in which longitudinal studies are uncommon (see Snow & Hoefnagel-Hoehle, 1978, and Aoyama et al., 2004, for rare examples of longitudinal studies).

The Gradual Diffusion Framework and Other Approaches to L2 Phonological Learning

Should further investigations corroborate the findings reported here, the gradual diffusion framework can offer an insightful view of L2 phonological learning. Based on the findings to date, this view is not incompatible with, and is clearly complementary to, other existing conceptualizations of L2 learning, particularly in the realm of phonology. For example, the framework adds a dynamic dimension to Eckman's (1991) structural conformity hypothesis, an account of L2 phonological learning that predicts (on the basis of typological cross-language comparisons) the relative difficulty of learning certain aspects of a L2 versus others. (See Archibald, 1998, and Broselow, Chen, & Wang, 1998, for studies conducted within this and related frameworks.) Using perceptual measures of cross-language similarity and corpus-based estimates of lexical frequency, the gradual diffusion framework qualifies, in processing terms, what is easy and what is difficult crosslinguistically and specifies how L2 phonological learning might progress over time.

The gradual diffusion framework is also compatible with Flege's (1995) speech learning model, a well-developed psycholinguistic view of L2 phonological learning. The framework shares with this model a focus on cross-language similarity as one factor that predicts learners' success in L2 learning, at least at the level of phonetic segments. Flege's model assumes that, in the course of L2 learning, learners make context-specific perceptual comparisons between L1 and L2 segments and that their ability to "discern at least some of the phonetic differences" between them might trigger learning (Flege, p. 239). The current framework adds a developmental dimension to this view of L2 learning by identifying the order in which context-specific instances of a L2 segment are learned and by describing the nature of variability associated with this learning.

Major's (1987, 2002) ontogeny-phylogeny model of L2 phonological learning and, more recently, Escudero and Boersma's (2004) optimality-theoretic model of L2 segmental learning come closest to the developmental dimension captured by the gradual diffusion framework. Major's model specifies three stages of L2 phonological learning: an initial stage characterized by a heavy

L1 influence, an intermediate stage of variable L2 production accuracy, and the final stage of targetlike L2 performance. It views L2 phonological learning as progressing from nontarget to target L2 forms via a period of variable performance, and it includes a discussion of possible factors (e.g., similarity) responsible for this learning path. Major's model, however, offers no developed psycholinguistic account of how learning takes places over time, nor does it specify, in processing terms, how such factors as similarity influence this learning. The gradual diffusion framework, as discussed here, thus extends Major's model with a description of two factors (perceived cross-language similarity and lexical frequency) and a demonstration of how these factors might determine the path of L2 phonological learning.

In their optimality-theoretic model of phonological categorization, Escudero and Boersma (2004) recently offered by far one of the most comprehensive accounts of L2 phonological learning. The centerpiece of this model is a set of ranked constraints and a formal learning algorithm, based on the principles of Optimality Theory and tested in a number of computer simulations. The described learning algorithm, an extension of the gradual learning algorithm (Boersma & Hayes, 2001), appears to provide a principled account of variability in the development of L2 perceptual categories (this particular research did not address production) and, in doing so, to successfully capture the effects of both psycholinguistic (processing) factors investigated in this study: cross-language similarity and lexical frequency. The gradual diffusion framework thus appears fully compatible with Escudero and Boersma's model and describes precisely the types of L2 production data that might be modeled within an optimality-theoretic framework. Additionally, however, the gradual diffusion framework identifies a particular implicational pattern according to which the newly acquired forms appear to emerge in L2 learners' speech. In future investigations, it will be interesting to determine how the implicational relations that underlie the gradual diffusion framework can be captured within an optimality-theoretic account of L2 production development and, vice versa, how the gradual diffusion framework can be applied to describe the development of L2 perceptual categories.

Developing Psycholinguistic Explanations for L2 Phonological Learning

Viewed from any theoretical standpoint, one contribution of the present study to understanding L2 phonological learning lies in its response to what Rickford (2002) termed the challenge of "seeking explanations for the implicational patterns" (p. 160); that is, a viable account of variability (and, by extension, of L2 learning as well), Rickford argued, should not only establish a systematic patterning to variability but should also attempt to explain it. As the results of the present study suggest, the observed systematicity in L2 phonological learning, at least at the level of individual segments, might be explainable through two (of course, among potentially many other) independently established psycholinguistic factors: perceived cross-language similarity and lexical frequency. Both factors and, in particular, their interaction in L2 phonological learning merit closer scrutiny.

Given that both perceived cross-language similarity (a segmental factor) and lexical frequency (a lexical factor) influence L2 phonological learning, how can this influence be explained within a single conceptualization, particularly one compatible with a dynamic framework of L2 phonological learning? One such conceptualization is exemplified by a class of two-representation connectionist models of spoken word processing and learning (Gupta & Dell, 1999; Gupta & MacWhinney, 1997; Luce, Goldinger, Auer, & Vitevitch, 2000). Although these models focus on different aspects of word processing—vocabulary learning (Gupta & MacWhinney), word production (Gupta & Dell), and recognition (Luce et al.)—they share the assumption that word processing and learning occur (at a minimum) at two levels of representation: lexical and phonological.

These models assume that, at the lexical representation level, words are organized in a network of interconnected nodes, each with its particular resting threshold of activation, which refers to the degree of a word's initial accessibility in perception or production. This threshold of activation is receptive to experience, such that frequent words (those that are perceived and produced frequently) have a higher resting threshold than infrequent words. Because frequent words are thus inherently more accessible than infrequent words, their processing should be facilitated. Research findings, in fact, support this claim, suggesting that frequent words are recognized and produced more rapidly and accurately than are infrequent words in both L1 and L2 (e.g., Luce & Pisoni, 1998).

These models also assume that, at the phonological representation level, individual segments are organized in a network of interconnected nodes as well, again each with its resting threshold of activation. At this level, however, each segment's resting threshold varies not as a function of lexical frequency but, rather, as a function of each segment's frequency in spoken language or the segment's phonotactic probability (Vitevitch & Luce, 2004). In English, for instance, /d/ would have a higher resting threshold than /ð/ because the former segment is more frequent than the latter. At this level, associations between individual segments also vary predictably, depending on how frequently each combination of segments occurs in spoken language, or depending on the segments' biphone probability (Vitevitch & Luce, 2004). Thus, for example, the connection strength between English /d/ and /ɪ/, which co-occur frequently, would be stronger than the connection strength between English /ð/ and /ɪ/, which do not. Because frequent segments and segment combinations are thus inherently more accessible than infrequent ones, their processing and learning should be facilitated. In fact, in both L1 and L2, common segment sequences are recognized and produced more rapidly and accurately than are uncommon ones (e.g., Bradlow & Pisoni, 1999; Vitevitch et al., 1997).

Two-representation connectionist models offer a viable framework for understanding L2 phonological learning in a manner that is compatible with the gradual diffusion approach. The view of L2 phonological learning as a process involving gradual learning driven by diffusion of target forms into learners' speech is clearly compatible with the assumptions that underlie computational approaches to modeling language processing and learning. In computational models, information is usually represented as patterns of activation across several processing units, and learning is often described as a gradual input-driven strengthening of associations between these units (e.g., Elman, 1990). Thus, a gradual learning-driven diffusion of target forms into learners' speech might be conceptualized as a gradual attunement of associations between the initial state of the learning network (which, in the case of L2 learning, is clearly influenced by learners' L1) and its target (L2) state.

Two-representation connectionist models also allow for understanding the role of cross-language similarity and lexical frequency in a manner that is compatible with the gradual diffusion framework presented here. Seen in the context of such models, both cross-language similarity (whose effects reside at the phonological level) and lexical frequency (whose effects reside at the lexical level) likely affect L2 segments' resting thresholds and hence their accessibility in processing. For example, L2 segments that occur in frequent words might become more readily accessible to learners and, therefore, more learnable than those that occur in infrequent words. Similarly, L2 segments that are perceptually distinct from L1 segments might become more readily accessible to L2 learners and, therefore, more learnable than L2 segments that are perceptually equated with an L1 segment. Until the exact nature of such enhanced accessibility and its consequences on learning is clarified in future research, particularly with respect to cross-language similarity effects (see Rivera-Gaxiola, Csibra, Johnson, & Karmiloff-Smith, 2000, and Guenther & Gjaja, 1996, for preliminary evidence), these claims must remain speculative.

CONCLUDING REMARKS

In summary, the goal of the present study was to reexamine and further develop a descriptive framework of adult L2 phonological learning, with a view to identifying the role of L2 input and of the learners' L1 and clarifying some of the sources of variability in L2 phonological learning. The point of departure was the descriptive framework provided by Gatbonton (1975, 1978), which offered a dynamic view of L2 phonological learning as a particular form of gradual diffusion of the newly acquired target elements throughout the learner's speech. The data presented in the research reported here pointed to cross-language similarity and lexical frequency as factors that are especially implicated in the learning process that underlies gradual diffusion. This conceptualization of L2 phonological learning aligns well with theories of L1 phonological development in children (Ferguson & Farwell, 1975; Morrisette & Gierut, 2002) and with computational modeling of language processing

(Gupta & Dell, 1999) and language change (Cangelosi & Parisi, 2002) as well as with recent views of language learning as emerging from the learner's interaction with language input (see MacWhinney, 1999, for review). In future research, it is essential to establish whether the predictions that stem from a gradual diffusion perspective that implicates cross-language similarity and lexical frequency are upheld in other learning situations (e.g., in the context of a longitudinal study) and in computational modeling. These contexts will provide environments not only for testing the predictions of dynamic models of L2 phonological learning but also for gaining insights into the processing mechanisms that underlie such learning.

(Received 11 September 2006)

NOTES

1. Although the gradual diffusion framework was originally discussed as a model, we refer to it as a descriptive framework. The term "framework" does not typically imply determining causal relationships or offering explanations (see Jordan, 2004, for an insightful discussion). Therefore, this term more closely corresponds to the original purposes of the framework: describing and predicting the acquisition of L2 segments over time.

2. As indicated by an anonymous *SSLA* reviewer, logistic regression (the procedure used to analyze linguistic variation in such software packages as VARBRUL) is a useful statistical procedure for analyzing production data. The use of logistic regression or any other similar procedure, although potentially revealing, addresses issues that go beyond the primary goal of the present paper (e.g., allocating weights to each factor known to determine variable responses), which is to analyze the dynamic nature of implicational relations in L2 production.

3. As suggested by an anonymous *SSLA* reviewer, this obtained pattern of variable /ð/ production might not involve Type 1 linguistic variation exclusively, because L1 realizations of English /ð/ are known to be variable in many varieties of North American English (e.g., Labov, 1966). It is possible, then, that the participants' variable /ð/ production might reflect, at least in part, some nonstandard L1 contextual realizations of English /ð/. In our dataset, however, we did not find support for this interpretation of our findings, as the five English NSs exhibited no contextual variability in their realizations of English /ð/. This interesting interpretation of our findings should nevertheless be explored in future research.

4. An anonymous *SSLA* reviewer suggested using fine-grained acoustic and articulatory analyses of the learners' productions as one principled way of determining context difficulty. This approach to defining phonetic context difficulty was, however, deemed inappropriate for the purposes of this study because it would involve the implicational scaling of production data based on the criterion derived by analyzing the same production data. To avoid such circularity in defining and applying the dynamic framework investigated here, a decision was made to use cross-language similarity and lexical frequency as the two psycholinguistic criteria for defining phonetic context difficulty. Well-motivated psycholinguistically, both these criteria are separable from the learners' production accuracy.

5. Because in Quebec it is virtually impossible to find unilingual English NSs having no exposure to or experience with French or another language, Quebec bilinguals for whom English is the mother tongue represent an ecologically valid population of English NSs. The variety of English spoken by these speakers is the language to which the majority of Francophone learners of English are exposed.

6. The British National Corpus is one of the largest language corpora currently available. This corpus, as opposed to, for example, the Brown Corpus, was used in this study because it allows for searching a large corpus of spoken language.

7. An anonymous *SSLA* reviewer suggested that variation in NS speech might be more systematic than variation in interlanguage. According to this reviewer, then, trinary scales (of the type used in the present study), as opposed to four- and five-valued scales or scales that use raw data, might be the most finely grained for scaling most interlanguage data. This interesting claim about the systematicity of interlanguage and the appropriate implicational scales needed to capture it should be addressed in future research.

REFERENCES

Adamson, H. D., & Regan, V. (1991). The acquisition of community speech norms by Asian immigrants learning English as a second language: A preliminary study. *Studies in Second Language Acquisition*, *13*, 1–22.

Amastae, J. (1978). The acquisition of English vowels. *Papers in Linguistics*, *11*, 423–458.

Andersen, R. W. (1978). An implicational model for second language research. *Language Learning*, *28*, 221–282.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /ɾ/ and English /l/ and /ɹ/. *Journal of Phonetics*, *32*, 233–250.

Archibald, J. (1998). Second language phonology, phonetics, and typology. *Studies in Second Language Acquisition*, *20*, 189–211.

Atkinson, M. (1982). *Explanation in the study of child language development*. New York: Cambridge University Press.

Bailey, C.-J. N. (1973a). *Variation and linguistic theory*. Arlington, VA: Center for Applied Linguistics.

Bailey, C.-J. N. (1973b). The patterning of language variation. In R. W. Bailey & J. L. Robinson (Eds.), *Varieties of present-day English* (pp. 156–189). New York: Macmillan.

Baker, W., & Trofimovich, P. (2005). Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language and Speech*, *48*, 1–27.

Bayley, R. (1999). The primacy of aspect hypothesis revisited: Evidence from language shift. *Southwest Journal of Linguistics*, *18*, 1–22.

Bayley, R., & Regan, V. (2004). Introduction: The acquisition of sociolinguistic competence. *Journal of Sociolinguistics*, *8*, 323–338.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, *109*, 775–793.

Best, C. T., McRoberts, G. W., & Sithole, N. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu clicks discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 345–360.

Bickerton, D. (1971). Inherent variability and variable rules. *Foundations of Language*, *7*, 457–492.

Bickerton, D. (1975). *Dynamics of a creole system*. New York: Cambridge University Press.

Boersma, P., & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry*, *21*, 45–86.

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America*, *106*, 2074–2085.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*, 2299–2310.

British National Corpus Consortium. (2000). *British National Corpus* (World edition). Oxford: Humanities Computing Unit of Oxford University.

Broselow, E., Chen, S.-I., & Wang, C. (1998). The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition*, *20*, 261–280.

Bybee, J. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition*, *24*, 215–222.

Bybee, J., & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.

Cangelosi, A., & Parisi, D. (Eds.). (2002). *Simulating the evolution of language*. New York: Springer-Verlag.

Cardoso, W. (2007). The variable development of English word-final stops by Brazilian Portuguese speakers: A stochastic optimality theoretic account. *Language Variation and Change*, *19*, 218–249.

Chen, M. Y., & Wang, W. S.-Y. (1975). Sound change: Actuation and implementation. *Language*, *51*, 255–281.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology I* (pp. 283–333). New York: Cambridge University Press.

DeCamp, D. (1971). Toward a generative analysis of a post-creole speech continuum. In D. Hymes (Ed.), *Pidginization and creolization of languages* (pp. 349–370). New York: Cambridge University Press.

Dickerson, W. B. (1976). The psycholinguistic unity of language learning and language change. *Language Learning*, *26*, 215–231.

Dinnsen, D. A. (1984). Methods and empirical issues in analyzing functional misarticulation. In M. Elbert, D. A. Dinnsen, & G. Weismer (Eds.), *Phonological theory and the misarticulating child* (ASHA Monographs No. 22). Rockville, MD: ASHA.

Dunn-Rankin, P. (1983). *Scaling methods*. Mahwah, NJ: Erlbaum.

Eckman, F. R. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition*, *13*, 23–41.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*, 551–585.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, *51*, 419–439.

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229–273). Timonium, MD: York Press.

Flege, J. E., Munro, M. J., & Skelton, L. (1992). Production of the word-final English /t/-/d/ contrast by native speakers of English, Mandarin, and Spanish. *Journal of the Acoustical Society of America*, *92*, 128–143.

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication, 40*, 467–491.

Flege, J. E., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English /r/ and /l/ accurately. *Language and Speech*, *38*, 25–55.

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. *Journal of the Acoustical Society of America*, *99*, 1161–1173.

Fougeron, C., & Smith, C. L. (1999). French. In *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet* (pp. 78–81). New York: Cambridge University Press.

Gatbonton, E. (1975). *Systematic variations in second language speech: A sociolinguistic study*. Unpublished doctoral dissertation, McGill University, Montreal, Canada.

Gatbonton, E. (1978). Patterned phonetic variability in second-language speech: A gradual diffusion model. *Canadian Modern Language Review*, *34*, 335–347.

Gregg, K. R. (1996). The logical and developmental problems of second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 49–81). San Diego: Academic Press.

Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America*, *100*, 1111–1121.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America*, *107*, 2711–2724.

Gupta, P., & Dell, G. S. (1999). The emergence of language from serial order and procedural memory. In B. MacWhinney (Ed.), *The emergence of language* (pp. 447–481). Mahwah, NJ: Erlbaum.

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language*, *59*, 267–333.

Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, *9*, 139–150.

Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Rowley, MA: Newbury House.

Hyltenstam, K. (1977). Implicational patterns in interlanguage syntax variation. *Language Learning*, *27*, 383–411.

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/- /θ/ contrast by Francophones. *Perception & Psychophysics*, *40*, 205–215.

Jespersen, O. (1912). *Lehrbuch der Phonetik*. Leipzig: Teubner.

Jordan, G. (2004). *Theory construction in second language acquisition*. Amsterdam: Benjamins.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*, 201–258.

Labov, W. (1966). *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Labov, W. (1971). Methodology. In W. O. Dingwall (Ed.), *A survey of linguistic science* (pp. 412–491). College Park: The University of Maryland.

Langman, J., & Bayley, R. (2002). The acquisition of verbal morphology by Chinese learners of Hungarian. *Language Variation and Change*, *14*, 55–77.

Leonard, L. B., Newhoff, M., & Mesalam, L. (1980). Individual differences in early child phonology. *Applied Psycholinguistics*, *1*, 7–30.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception and Psychophysics*, *62*, 615–625.
Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*, 1–36.
MacWhinney, B. (Ed.). (1999). *The emergence of language*. Mahwah, NJ: Erlbaum.
Major, R. C. (1987). A model for interlanguage phonology. In G. Ioup & S. H. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 101–125). New York: Newbury House/Harper & Row.
Major, R. C. (2002). The phonology of the L2 user. In V. Cook (Ed.), *Portraits of the L2 user* (pp. 67–92). Clevedon, UK: Multilingual Matters.
Manuel, S. Y. (1995). Speakers nasalize /ð/ after /n/, but listeners still hear /ð/. *Journal of Phonetics*, *23*, 453–476.
Meisel, J. M., Clahsen, H., & Pienemann, M. (1981). On determining developmental stages in natural second language acquisition. *Studies in Second Language Acquisition*, *3*, 109–135.
Morrisette, M. L., & Gierut, J. A. (2002). Lexical organization and phonological change in treatment. *Journal of Speech, Language, and Hearing Research*, *45*, 143–159.
Mougeon, R., Rehner, K., & Nadashi, T. (2004). The learning of spoken French variation by immersion students from Toronto, Canada. *Journal of Sociolinguistics*, *8*, 408–432.
Moyer, A. (1999). Ultimate achievement in L2 phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition*, *21*, 81–108.
Nagy, N., Moisset, C., & Sankoff, G. (1996). On the acquisition of variable phonology in L2. *University of Pennsylvania Working Papers in Linguistics*, *3*, 111–126.
Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: Benjamins.
Pienemann, M., & Mackey, A. (1993). An empirical study of children's ESL development and Rapid Profile. In P. McKay (Ed.), *ESL development and language literacy in schools* (pp. 115–259). Canberra: Commonwealth of Australia and National Languages and Literacy Institute of Australia.
Pitt, M. A., & Samuel, A. G. (1995). Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology*, *29*, 149–188.
Politzer, R. L. (1976). The implicational relation paradigm in language acquisition. In A. Juilland (Ed.), *Linguistic studies offered to Joseph Greenberg* (pp. 123–135). Saratoga, CA: Anma Libri.
Preston, D. R. (1996). Variationist perspectives on second language acquisition. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 1–45). Amsterdam: Benjamins.
Preston, D. R. (2000). Three kinds of sociolinguistics and SLA: A psycholinguistic perspective. In B. Swierzbin, F. Morris, M. E. Anderson, C. A. Klee, & E. Tarone (Eds.), *Social and cognitive factors in second language acquisition: Selected proceedings of the 1999 Second Language Research Forum* (pp. 3–30). Somerville, MA: Cascadilla Press.
Regan, V. (1996). Variation in French interlanguage: A longitudinal study of sociolinguistic competence. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 177–202). Amsterdam: Benjamins.
Rickford, J. R. (2002). Implicational scales. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 142–167). Oxford: Blackwell.
Rivera-Gaxiola, M., Csibra, G., Johnson, M. H., & Karmiloff-Smith, A. (2000). Electrophysiological correlates of cross-linguistic speech perception in native English speakers. *Behavioural Brain Research*, *111*, 13–23.
Smith, S. C. (1997). *UAB Software*. Birmingham: Department of Rehabilitation Sciences, University of Alabama at Birmingham.
Snow, C. E., & Hoefnagel-Hoehle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development*, *49*, 1114–1128.
Tarone, E. (1988). *Variation in interlanguage*. London: Edward Arnold.
Tarone, E. (2002). Frequency effects, noticing, and creativity. *Studies in Second Language Acquisition*, *24*, 287–296.
Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
Trofimovich, P., Baker, W., & Mack, M. (2001). Context- and experience-based effects on the learning of vowels in a second language. *Studies in the Linguistic Sciences*, *31*, 167–186.
Trubetzkoy, N. S. (1969). *Principles of phonology* (A. M. Baltaxe, Trans.). Los Angeles: University of California Press. (Original work published 1939).

Trudgill, P. (1986). *Dialects in contact*. Oxford: Blackwell.
Vitevitch, M. S., Luce, P., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, *40*, 47–62.
Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, *36*, 481–487.
Wenk, B. J. (1979). Articulatory setting and de-fossilization. *Interlanguage Studies Bulletin*, *4*, 202–220.
Young, R. F. (1991). *Variation in interlanguage morphology*. Bern: Peter Lang.

APPENDIX A

THE READING USED IN THE PRESENT STUDY: THE FAMOUS AUTHOR

A famous author had just finished another play and preparations were underway to stage the play at the local theatre. My father, who was at that time directing the play, thought he should seek the author's advice on the scenery, costumes, or the lights. So he invited the author to help set the stage. The writer was happy to be asked. In the beginning, he came to the studio only once in a while. After that, he came more frequently. Soon he was there every day, carefully observing the crew working to finish the set. At first, he offered his ideas only when my father asked him to, but before long he was giving advice without consulting anyone. Then he began supervising the crew himself, and it was clear that he was hard to please. In fact, he bothered everyone.

He had definite ideas about everything. For example, he wanted the scene where the main characters hold hands while watching the sunset to be spectacular. So he spared nothing to achieve this effect. He instructed the crew about what to do all the time. They worked hard to produce the effect he desired. They had to replace the curtains several times to choose the right color background for the sunset scene. He would tell the lighting technicians to try different lighting combinations and would show them how to do it. At his request, these workers took the red lights from the high ceiling in order to attach them to the wall. They projected the lights from the seating area and from beside the stage. They shone the lights directly above the stage and beneath the curtains. Sometimes he directed the crew to dim the lights. At other times, he had to order them to flash the lights full blast. On his instructions, the crew took off the light covers to wash them. They wrapped the lights in cloth or hung them bare over the stage. They flooded the whole theatre with a soft light. They shed the brightest lights from under the stage. But nothing satisfied the author. The effect he wanted was not there.

A month later, during an unusually hectic rehearsal, he suddenly saw the effect he had dreamed of.

"Hold that!" he shouted to the men behind the stage. "Leave the lights alone. Don't move them. Don't touch them. Don't change them till I get there."

"I'm sorry, sir"—shouted the stage manager, running up the stage—"but this is impossible! We can't do that!"

"Why not?" asked the author." "Is there a problem?"

"Because the theatre is on fire, sir. That's the effect you're seeing now."

APPENDIX B

TARGET ENGLISH /ð/ TOKENS IN EIGHT PHONETIC CONTEXTS

Intervocalic	Voiceless stop	Voiced stop	Voiceless fricative/affricate	Nasal	Liquid	Sentence initial	Voiced fricative/affricate
another	at the	invited the	finish the	directing the	or the	The. . .	move them
father	set the	flooded the	wash them	observing the	after that	The. . .	stage the
without	asked the	wanted the	produce the	supervising the	clear that	Then. . .	choose the
bothered	not there	shed the	off the	watching the	under the	They. . .	achieve this
other	took the	instructed the	replace the	hung the	where the	They. . .	is there
show them	up the	satisfied the	beneath the	dim the	tell the	They. . .	leave the
to the	wrapped the	projected the	attach the	on the	all the	They. . .	above the
with a	but this	shouted the	touch them	in the	over the	They. . .	change them
do that	get the	directed the	flash the	from the	for the	They. . .	was there
saw the	seek the	beside the	that’s the	shone the	order them	They. . .	because the