

THE EMERGENCE OF THE UNMARKED IN SECOND LANGUAGE PHONOLOGY

Ellen Broselow, Su-I Chen, and Chilin Wang

*State University of New York at Stony Brook,
National Tsing-Hua University, and Tunghai University*

This paper discusses the simplification of forms ending in obstruents by native speakers of Mandarin, in particular two effects that are not obviously motivated by either the native- or the target-language grammars: a tendency to devoice final voiced obstruents and a tendency to maximize the number of bisyllabic forms in the output. These patterns are accounted for within Optimality Theory, which describes a grammar as a set of universal, ranked constraints. It is argued that the devoicing and bisyllabicity effects result from universal markedness constraints that are present in all grammars but that are masked in the learner's native-language grammar by the effects of higher ranking constraints.

Much research in second-language acquisition has uncovered patterns that appear to be independent of both the native-language grammar and the target-language grammar. Such patterns frequently reveal a preference for less marked structures. Simplification in the direction of less marked structures is generally described as an effect of universal principles of markedness, often conceived of as part of the innate endowment provided by Universal Grammar (see Epstein, Flynn, & Martohardjono, 1996, for a review of relevant second-language acquisition literature).

What has been missing from many analyses is an explicit account of precisely what role these universal principles play in the grammars of language learners. Virtually all work in generative phonology, beginning with *The sound pattern of English* (Chomsky & Halle, 1968), has been predicated on the model of a

We are grateful to two anonymous reviewers for helpful comments. Address correspondence to Ellen Broselow, Department of Linguistics, SUNY, Stony Brook, NY 11794-4376; e-mail: ellen.broselow@sunysb.edu.

phonological grammar as a set of learned rules that serve to convert underlying representations into surface representations. Although assumptions concerning the nature of phonological representations have changed considerably during the decades following the publication of *SPE* (see Kenstowicz, 1994), until very recently most research still assumed a model in which the learner of a language posited a set of rules based on the surface patterns of that language. Such a model provides no obvious place either for the appearance in interlanguage grammars of rules that are not motivated by surface alternations, or for the effects of markedness constraints in shaping the language learner's grammar.

We will argue that a more recent model of phonology, the framework of Optimality Theory (McCarthy & Prince, 1993, 1995; Prince & Smolensky, 1993), provides a potential solution to this problem. The conception of a grammar in this framework is substantially different from that in earlier generative frameworks: A grammar consists not of a set of rules, but rather of a set of ranked constraints defining the optimal output corresponding to any input string. This set of constraints is presumed to be innate and universal. What the language learner must induce from the data is the rankings of these universal constraints, rather than the constraints themselves. Optimality Theory therefore differs crucially from standard generative rule-based phonology, in which it is generally assumed that the rules of a grammar are learned by the speaker, based on data encountered in the language-learning situation,¹ and that languages differ by virtue of having different rules. In Optimality Theory, languages differ not in their set of constraints, but in the rankings of those constraints, where ranking determines the strength of a particular constraint in a particular language. Lower ranked constraints may normally have no visible effects in a grammar, but they are still assumed to be present in the grammar. Much work in Optimality Theory has focused on the circumstances in which the effects of low-ranked markedness constraints become visible, a situation described as "the emergence of the unmarked" by McCarthy and Prince (1994). We will argue that the markedness effects that are often visible in second-language acquisition represent this sort of situation.

Our data involve the simplification of English syllable codas by native speakers of Mandarin Chinese. Mandarin is far more restrictive in its range of coda consonants than English, and learners' simplifications of English codas have been the subject of a number of studies (Eckman, 1981; Heyer, 1986; Weinberger, 1988; Yin, 1984). We will focus on the findings of Wang (1995), who carried out a carefully controlled study of the pronunciation of English syllable codas by native speakers of both Mandarin and Taiwanese. Whereas English permits both voiced and voiceless obstruents (e.g., /b/ and /p/) in syllable coda position, Taiwanese permits only voiceless obstruents in the coda and Mandarin permits no coda obstruents of either type. In this paper we restrict our discussion to patterns of the Mandarin speakers because Mandarin's extremely restrictive range of coda types affords the most occasions for syllable simplification. We argue that describing the interlanguage grammars of these speakers as a set

of ranked, universal constraints accounts for both the origin of the simplification strategies in the grammars and the learners' choice of simplification strategy.

THE PROBLEM: MANDARIN ERROR PATTERNS

One problem faced by speakers of Mandarin learning English is the mastery of English codas, which permit a far wider range of consonant types than attested in their native language. As shown in (1), Mandarin does not allow the nasal [m], any liquids, or any obstruents in coda position.

- (1) Possible Syllable Codas
- a. Mandarin: glide, nasal (n, ŋ)
 - b. English: glide, liquid, nasal (m, n, ŋ), voiceless obstruent, voiced obstruent

We would expect Mandarin speakers to have difficulty with obstruent codas because obstruent codas do not occur in their native language, and the presence of obstruent codas is more marked than their absence (i.e., no language has syllables ending in obstruents but not vowels).

Data: Wang 1995

Wang examined the pronunciation of English codas by 10 Mandarin speakers aged 23–30, each of whom had had 6–7 years of EFL instruction in their home country and had been in an English-speaking country for less than a year. The instrument used was a version of Broselow and Finer's (1991) so-called vocabulary learning test, designed to deflect subjects' attention from pronunciation. Subjects were asked to listen to and memorize a list of nonce words constructed by the experimenter, with definitions. Subjects were then given definitions and asked to choose from three possible nonce words fitting the definition. Because all three choices ended in the same coda (and were otherwise prosodically similar), any of the three answers served equally well as a probe into the learner's pronunciation of that coda. Possible answers were presented both on tape and in written form, simultaneously, to ensure that learners were aware of the identity of the coda consonant. Subjects' responses were spoken into a tape recorder and then transcribed phonetically.

In the discussion below we focus on the subjects' treatment of final stops, both voiceless stops /p t k/ and voiced stops /b d g/, neither of which is possible in Mandarin codas. Wang (1995) found that subjects did indeed have considerable difficulty with those codas that do not occur in Mandarin—of a total of 180 final stop tokens, 81% of the voiceless stops and 98% of the voiced stops were produced incorrectly. The incorrect productions involved epenthesis of a vowel after the coda stop (i.e., target [vig] pronounced as [vi.gə], where “.” indicates a syllable boundary), deletion of the coda stop ([vig] pronounced as [vi]), or devoicing of a final voiced stop ([vig] pronounced as [vik]). The percentages and numbers of the different error types are shown in Table 1.

Table 1. Mandarin error types (Wang, 1995)

	Voiceless Stops (<i>n</i> = 90)	Voiced Stops (<i>n</i> = 90)
Correct:	19% (17)	2% (2)
Epenthesis:	36% (32)	36% (32)
Deletion:	46% (41)	43% (39)
Devoicing:	—	19% (17)

The data in Table 1 raise three major questions, which we will address in turn.

The Source of Interlanguage Rules. One question raised by these data is: Just where do the strategies of epenthesis, deletion, and devoicing originate? Native Mandarin forms provide no evidence for underlying representations with illegal codas, so there are no alternations providing evidence for a rule adding a vowel after an illegal coda consonant. Similarly, the data available to the Mandarin speaker provide no evidence for a productive rule deleting illegal codas. Nor can these rules be imported from the target-language grammar: English has no regular, productive rule of vowel epenthesis or deletion for illegal codas, and although English does provide for schwa epenthesis after some unsyllabifiable consonants (most noticeably in onset position, as in the pronunciation of Evil Knievel's last name as [kə'nivəl]), this is but one of several strategies used for rendering onsets pronounceable (including segment reordering, as in the common pronunciation of Tbilisi as [tɪblɪzi]). Certainly epenthesis and deletion are not the sorts of productive processes in English that we would expect L2 learners to have mastered on the basis of exposure to English. And final devoicing is clearly not a rule of Mandarin because neither voiced nor voiceless final obstruents are permitted; neither is it a phonological rule of English. (Although there is a tendency for final voiced obstruents in English to be less fully voiced than their counterparts in initial and final positions, this is not the sort of full phonological neutralization found in, for example, German and Russian.) It is therefore not clear how epenthesis, deletion, and devoicing appear in the learner's interlanguage grammar, given the assumption that these phenomena are the effects of rules that are learned from exposure to alternations in surface representations.

The Choice of Simplification Strategy. The second issue we will consider is the factors determining the choice of simplification strategy. Most of the subjects in this study employed both epenthesis and deletion, and some employed devoicing as well, to transform the target-language structures into structures that conform to native-language syllable coda types or at least into less marked structures. Wang (1995) argues that the choice of strategy is connected to a preference for bisyllabic forms: Epenthesis is favored in monosyllables

because the output is a bisyllabic word, whereas other strategies, which do not add a syllable, are favored in bisyllabic forms. Because Mandarin does allow native words of one syllable, how is the preference for bisyllabic forms encoded in the interlanguage grammar?

The Effect of Markedness. The third issue we will address concerns the asymmetry between voiced and voiceless stops. Because neither voiced nor voiceless stops are permitted in Mandarin codas, we would expect, under a simple transfer account, that both types should be equally difficult for Mandarin speakers. However, whereas 19% of voiceless stops were produced correctly, only 2% of the voiced stops were produced correctly. Furthermore, 19% of the incorrect productions of voiced stops involved their production as the corresponding voiceless stop—in other words, one impossible native-language segment type was converted to another impossible native-language segment type. Wang (1995) pointed out that, because voiceless stops are generally considered less marked than voiced stops in coda position, the Markedness Differential Hypothesis of Eckman (1977) predicts that the less marked voiceless stops should be easier for learners than the more marked voiced stops. But just how do markedness considerations of this type shape an interlanguage grammar?

THE SOURCE OF INTERLANGUAGE “RULES”

We've shown that in Wang's study the Mandarin speakers' mispronunciations of English forms containing stops in coda position took three different forms: Speakers added a vowel after the coda stop, deleted the coda stop, or devoiced a voiced coda stop. The motivation for the first two strategies is clear: They have the effect of transforming a form that would be an impossible syllable structure in the native language to one that is a legal native-language syllable structure. For example, [vɪg] has [g] in its syllable coda, and [g] is not a possible coda segment in Mandarin. Both [vɪ.gə] and [vɪ] avoid this illicit coda structure. The motivation for devoicing is less clear, however, because [vɪk] is also not a legal Mandarin syllable. For the present, we will focus on the strategies of epenthesis and deletion, postponing discussion of devoicing.²

Phonological rules have generally been assumed to provide a way of deriving different surface forms from a single underlying representation: For example, a rule converting intervocalic [t] to a flap before an unstressed vowel accounts for the different realizations of the morpheme *wet* in the words [wɛt] *wet* and [wɛDɪŋ] *wetting*. The absence of a particular surface configuration (e.g., obstruents in coda position) may have one of two sources: a phonological rule that converts such configurations into something else, or a constraint banning such configurations in underlying representations.

We can illustrate the rule-based account of a lack of coda obstruents by positing a hypothetical language with no surface obstruents in syllable coda position but with alternations such as the following:

(2) Hypothetical Language X Forms

	Singular	Plural	Proposed underlying representation
a.	vɪ	vɪgɔs	/vɪg/
b.	vɪ	vɪɔs	/vɪ/
c.	vɪn	vɪnɔs	/vɪn/

Most phonologists would describe the contrast between (2a) and (2b) by positing an underlying morpheme-final stop in (2a). The failure of these stops to surface when they would occur in syllable coda position (as in the singular underlying representation /vɪg/) is ascribed to a rule deleting obstruents in coda position.

However, the lack of obstruent codas in Mandarin would in most derivational accounts have a different source. Because Mandarin does not provide evidence from alternations forcing us to posit final stops in underlying representations, the Mandarin pattern would be presumed to result from constraints on underlying representations (i.e., constraints prohibiting stops in morpheme-final position). The lack of syllable-final stops would not result from a rule removing stops in this position because there are no alternations that force us to posit the illegal underlying representations. Thus, the evidence of a stop-deletion rule in Mandarin speakers' production of English forms requires us to posit a new rule, one that is not part of the native-language grammar.

In contrast to rule-based theories, in which constraints on surface configurations have no theoretical status, constraint-based grammars (such as Optimality Theory) take these surface constraints as the components of which grammars are made. In this approach, the lack of coda obstruents in both Language X and in Mandarin have the same source: a constraint barring obstruents from syllable coda position in surface forms. This constraint is a constraint on surface representations—in this framework, underlying representations are entirely unconstrained (see Yip, 1996b, for discussion of this point). The fact that Mandarin provides no evidence for final obstruents in underlying representation, whereas Language X does, is therefore of no theoretical import: Should either Language X or Mandarin be faced with any underlying representations with final stops, the surface constraint will prevent these stops from appearing in syllable coda position in surface forms.³ The presence of this constraint in the grammar ensures that, should a Mandarin speaker posit an underlying representation like /vɪg/ (most likely, in the course of learning another language), the surface constraint will rule out [vɪg] as a possible surface correspondent of that underlying representation—at least until the point at which, under pressure from such forms, the learner constructs an alternative grammar (or perhaps a subgrammar specific to foreign words; see Ito & Mester, 1995). The use of surface constraints is one aspect of Optimality Theory that makes it particularly well suited to the description of L2 phonology. The second language may introduce underlying representations that are not motivated by the facts of the first language, but because all underlying representations will ultimately be forced

to conform to the surface constraints of the learner's grammar, the foreign underlying representations will come out looking like the surface forms of the native language (until the stage at which the learner begins to develop different grammars for the native and target languages).

The precise mechanism by which Optimality Theory chooses the best surface correspondent to an underlying representation is as follows. For each underlying representation, generally called the input, a function called GEN generates a set of possible surface correspondents (i.e., ways in which this underlying representation could conceivably be pronounced). This set of candidates for surface pronunciation includes a form corresponding exactly to the input, as well as all other possible permutations and alternations of the input form. For example, the most likely candidates for surface pronunciation of /vig/ include [vig], [vi.gə], [vi], and [vik], among others. GEN's function is therefore to answer the question: Onto which surface representations might a speaker conceivably map this underlying representation? The task of the surface constraints is to answer the question: Which of these possible surface representations is best, as defined by the constraint ranking of this language?

Note that in this model the set of different surface forms (those showing effects of epenthesis, deletion, devoicing, and all other conceivable phonological modifications) is given by the universal function GEN. Thus, in order to transform underlying /vig/ to [vigə], for example, there is no need for a learner's grammar to include a rule of vowel epenthesis. Rather, the learner's grammar must simply rank the constraints so that [vigə] is identified as the best of the possible surface correspondents of underlying /vig/.

We can exemplify this process by which the optimal surface candidate is selected by examining the performance of the Mandarin speakers in Wang's (1995) study. We assume that for these subjects, as indeed for all speakers of all languages, the constraint set contains the following three constraints:

- (3) Constraints
- a. NO OBS CODA: Syllable codas may not contain obstruents.
 - b. MAX (C): Maximize the consonants in the input (do not delete consonants).
 - c. DEP (V): The vowels in the output should be dependent on the input (do not add vowels).

Constraint (3a) is a markedness constraint, which penalizes more marked forms (this constraint reflects the fact that a number of languages allow sonorant codas but not obstruent codas). Constraints (3b) and (3c) belong to the set of faithfulness constraints, which enforce close correspondence between the input and the output. Essentially, these constraints ensure that an input form is not altered unless there is a specific reason for this alteration (such as transforming the input into a less marked output). For an input containing a final stop (such as /vig/), there is no way to satisfy constraint (3a) without being unfaithful to the input: (3a) will disfavor the surface candidate ([vig]) that corresponds perfectly to the input. But candidates that avoid violating (3a) do so at the expense of faithfulness to the input. One plausible means of avoiding a violation

of (3a) is by choosing a candidate that simply lacks the final stop; that is, [vɪ]. This avoidance comes at a cost, however, because this form violates (3b). Alternatively, we could choose the form that maintains the final stop but removes it from final position by adding a vowel after it, [vɪ.gə], but again with a cost: This candidate violates (3c). We consider for the moment just these three candidate surface forms corresponding to /vig/: [vig], which violates (3a); [vɪ], which violates (3b); and [vɪ.gə], which violates (3c). Which form a speaker chooses will depend on which constraint is stronger, or more highly ranked, in the speaker's grammar. A speaker of English, who chooses the form that violates constraint (3a), ranks (3b) and (3c) higher than (3a); Mandarin speakers who choose [vɪ] rank both (3a) and (3c) higher than (3b); although Mandarin speakers who choose [vɪ.gə] rank (3a) and (3b) higher than (3c).

The mechanism for evaluation of surface candidates against constraints is illustrated in (4).

(4) Analyses of /vig/
a. English

input: vig	MAX (C)	DEP (V)	NO OBS CODA
a. v ig			*
b. vɪ	*!		
c. vɪ.gə		*!	

b. Mandarin Subjects Favoring Deletion

input: /vig/	NO OBS CODA	DEP (V)	MAX (C)
a. vig	*!		
b. v ɪ			*
c. vɪ.gə		*!	

c. Mandarin Subjects Favoring Epenthesis

input: /vig/	NO OBS CODA	MAX (C)	DEP (V)
a. vig	*!		
b. vɪ		*!	
c. v ɪ.gə			*

Each candidate is evaluated against each constraint in turn, beginning with the highest ranked constraint, with constraints arranged from left to right in order of highest- to lowest-ranked constraint. A form that violates a particular constraint receives an asterisk (*) in the box below that constraint. Those forms that violate the highest ranked constraint are eliminated (with each fatal violation indicated as *!), and surviving candidates go on to be evaluated against the next highest ranked constraint. At the point at which all candidates but one

have been eliminated, that sole survivor is identified as optimal (the optimal form is indicated by ☞). Although in each case in (4) the optimal form violates some constraint, the surviving form is nonetheless defined as optimal because all other forms violate some higher ranked constraint. The shading in the right-hand columns in (4) indicates that, at that point, because all other candidates have been eliminated, the constraint over the shaded portion does not play a role in choosing the optimal candidate.

This analysis provides a somewhat idealized picture of the Mandarin learners' grammars because, in fact, many of the same subjects chose deletion in some cases and epenthesis in others. Below, we will examine the factors affecting their choice and give a fuller picture of the interlanguage grammar. Before turning to the question of what influences the choice of epenthesis versus deletion, however, we need to determine whether this picture of the grammar answers the question of how interlanguage grammars come to incorporate processes (such as epenthesis and deletion) that are not rules of either the native- or target-language grammars (and possibly of no language, as Eckman, 1981, argues; although see Broselow, 1988, for arguments against this claim). Optimality Theory grammars do not contain language-specific rules: Instead, GEN supplies a set of all possible surface forms corresponding to each input representation. Thus, speakers of Mandarin (as for every other language) will be supplied by GEN with a set of possible surface forms showing the effects of vowel insertion, consonant deletion, and so forth. The speaker's job is to determine which of these forms best satisfies the set of universal constraints, as these constraints are ranked in the speaker's language. This account does not, therefore, require us to posit anything in the interlanguage grammar that is not in the native- or target-language grammars, both of which contain GEN along with the set of universal constraints. A Mandarin learner of English is therefore not required to induce a rule of vowel insertion or consonant deletion but merely to check the set of forms supplied by GEN (which include forms with inserted vowels and with deleted consonants) against the highly ranked NO OBS CODA constraint.

The model of second-language phonology we are assuming, then, is one in which the learner evaluates the set of candidate surface representations corresponding to a given input against the surface constraints of the L1. Initially, the ranking of these constraints will be as in the native language, although as the learner becomes more proficient, an interlanguage grammar will develop in which the rankings of constraints more closely approximate the target-language ranking. Thus, in the grammar of the Mandarin speakers who come to pronounce coda stops (i.e., who achieve a grammar closer to that of the target language), constraint (3a) NO OBS CODA must be demoted from its high ranking in the native-language grammar. Demotion of NO OBS CODA to a position below the faithfulness constraints (which results from the learner's recognition that obstruent codas are permitted in the target language) will make deletion of a consonant or insertion of a vowel less desirable than maintenance of a stop consonant in coda position. The initial and target rankings are shown in (5).

- (5) Rankings
- a. Initial Mandarin Ranking: NO OBS CODA \gg MAX (C), DEP (V)
 - b. Target-language (English) Ranking: MAX (C), DEP (V) \gg NO OBS CODA

In (5a), we show no ranking between MAX (C) and DEP (V), and in fact, the native language does not provide clear evidence for any such ranking. As discussed above, Wang's (1995) subjects did employ both vowel epenthesis and consonant deletion to remove stops from syllable coda position. In the next section, we consider how this model of a grammar might help to account for their choice of strategy.

THE CHOICE OF SIMPLIFICATION STRATEGY: EPENTHESIS VS. DELETION

Having presented a model in which epenthesis and deletion reflect the universal function GEN, plus language-specific constraint rankings, we now turn to the question of whether the learners' choice of strategies is predictable and, if so, how this is reflected in their grammars.

Data

Wang's study (1995) revealed variation in strategies for producing surface forms without stop codas, even among the same subjects. Wang argued that the major predictor of strategy is number of syllables, with subjects showing a clear preference for bisyllabic over monosyllabic or trisyllabic forms (see also Heyer, 1986). This means that in monosyllabic input forms epenthesis should be preferred over deletion of the final consonant because epenthesis will transform a monosyllabic form [vɪg] to a bisyllabic form [vi.gə], whereas correct production, deletion of the final consonant, and devoicing of the final consonant all yield a monosyllabic output. Table 2 shows the output realization, in terms of number of syllables, of monosyllabic input forms. The preferred pronunciation of monosyllabic forms was as a bisyllabic output form with the vowel of the second syllable supplied by epenthesis, chosen in 72% of the cases.

The production of input bisyllabic forms contrasts sharply. As shown in Table 3, epenthesis was employed in only 18% of the bisyllabic forms, whereas 83% of the forms were produced as bisyllabic.

Thus the choice between a pronunciation that adds a syllable to the input (via epenthesis) and a pronunciation that maintains the original syllable count (via correct production, deletion of the final stop, or devoicing of the final stop) seems clearly related to a preference for bisyllabic forms. An input monosyllabic form is most likely to add a syllable, via epenthesis, whereas an input bisyllabic form is most likely to resist the addition of a syllable.

At this point we must consider the possibility that the preference for epenthesis of a vowel after a coda stop over deletion, devoicing, or correct production of the coda stop is related to stress. In monosyllables, for which epenthesis is

Table 2. Monosyllabic input

Input σ ($n = 60$)	Correct	Deletion	Devoicing	Epenthesis
a. Output σ : 28% ($n = 17$)	6	5	6	—
b. Output $\sigma\sigma$: 72% ($n = 43$)	—	—	—	43

Table 3. Bisyllabic input

Input $\sigma\sigma$ ($n = 120$)	Correct	Deletion	Devoicing	Epenthesis
a. Output $\sigma\sigma$: 83% ($n = 99$)	13	75	11	—
b. Output $\sigma\sigma\sigma$: 18% ($n = 21$)	—	—	—	21

Table 4. Bisyllabic input by stress

Input $\sigma\sigma'$ ($n = 60$)	Correct	Deletion	Devoicing	Epenthesis
a. Output $\sigma\sigma'$: 70% ($n = 42$)	3	32	7	—
b. Output $\sigma\sigma'\sigma$: 30% ($n = 18$)	—	—	—	18

the preferred option, the monosyllable is also stressed. Could it be that consonants at the end of a stressed syllable are more salient and therefore less likely to undergo deletion or devoicing?

Wang's study (1995) provides an answer to this question because her study also controlled for stress. Half the input bisyllabic forms had stress on their first syllable, and half on the second. If coda stops in stressed syllables are more likely to undergo epenthesis than coda stops in unstressed syllables, we should see the same pattern in monosyllabic forms and in bisyllabic forms with final stress. However, the latter pattern, shown in Table 4, is quite different from the pattern for monosyllables shown in Table 2. Thus, whereas monosyllables with final stress undergo epenthesis in 72% of the cases, bisyllables with final stress undergo epenthesis in only 30% of the cases. This suggests that bisyllabicity is the major factor affecting choice of strategy.⁴ We will now consider how the preference for bisyllabicity is manifested in the interlanguage grammar.

Choice of Strategy

A number of languages show a preference for surface forms that are minimally bisyllabic (see McCarthy & Prince, 1993; Prince & Smolensky, 1993). In most Optimality Theory accounts, this is the cumulative effect of several constraints variously requiring that each major lexical category word (noun, verb, adjective, or adverb) contain a stress foot, that feet be binary, and that all syllables be parsed into feet. A bisyllabic word is optimal because every syllable can be parsed into a binary foot. In the following discussion, we will use the single

constraint WD BIN (Word binarity), with the understanding that it is doing duty for a set of related constraints that combine to favor bisyllabic words:

(6) WD BIN: Words should consist of two syllables.

Whether Mandarin words are subject to this constraint is not entirely clear. Wang (1995) argued that the preference for bisyllabic words is evidenced in native-language forms by the occurrence of semantically empty filler syllables serving only to augment monosyllabic morphemes. Thus in (7), the syllable /zi/ is attached to the monosyllabic morpheme meaning table, but disappears when this morpheme is compounded with another syllable.

- (7) Mandarin filler syllables
- | | | |
|----|------------------------|-------------------|
| a. | <i>zwo-zi</i> | table |
| b. | <i>fan-zwo</i> | dining room table |
| c. | (* <i>fan-zwo-zi</i>) | |

However, unaugmented monosyllabic words certainly do occur in Mandarin:

- (8) *wo xi-huan kan hu*
 I like see flowers
 "I like to see flowers."

The preference for bisyllabic words cannot therefore be an absolute requirement of Mandarin; at best, we can describe it as a tendency. How, then, does this come to play a role in second-language acquisition?

We argue that the preference for bisyllabic forms in the interlanguage of the Mandarin speakers is a case of the emergence of the unmarked. The constraints we collapse as WD BIN (constraints requiring that each lexical word minimally contain a foot and that feet are optimally binary) are well-established markedness constraints (i.e., constraints enforcing a preference for particular structures) in the grammar of a number of languages.⁵ According to the assumptions of Optimality Theory, the presence of WD BIN in the grammar of one language implies that it is present in the grammar of all languages, although its effects may be masked by higher ranking constraints—specifically, constraints requiring faithfulness to the input. Therefore, WD BIN is presumably present in the Mandarin grammar, although ranked below faithfulness constraints. Monosyllabic input forms are not altered in the direction of bisyllabicity because to do so (i.e., to make these forms satisfy WD BIN) would make these forms unfaithful to their inputs, violating higher ranking faithfulness constraints.

However, input forms that cannot be realized faithfully under any circumstances—that is, forms that must be altered to avoid violating some higher ranked constraint, such as NO OBS CODA—will be altered in the direction of satisfying WD BIN. We can illustrate this with tableaux (9), (10), and (11). We assume that, for at least one group of Mandarin learners, the two faithfulness constraints DEP (V) and MAX (C) have the same ranking: Both are ranked below

NO OBS CODA, and above WD BIN, but are not ranked with respect to each other. In tableau (9), the native form /kan/ “see” satisfies the highest ranked constraint, NO OBS CODA, as do its competitors, with epenthesis and deletion, respectively. We then evaluate the three candidates with respect to the faithfulness constraints DEP (V) (do not insert a vowel) and MAX (C) (do not delete a consonant). Only the faithful form [kan] satisfies these constraints, and so the competitors (10b) and (10c) are eliminated at this stage. Because evaluation proceeds step by step, with the losers in each round being eliminated from further consideration, it is irrelevant that [kan] violates lower ranked WD BIN because [kanə] has already been disqualified:

(9) Mandarin Native Form

input: kan	NO OBS CODA	DEP (V), MAX (C)	WD BIN
a. \emptyset kan			*
b. ka.n \emptyset		*!	
c. ka		*!	*

The situation is different with an input like the test form /vig/, because this form cannot satisfy NO OBS CODA without a violation of faithfulness: Either epenthesis or deletion is required to move the stop out of coda position. We see how this works in (10).

(10) Interlanguage Form: Monosyllabic (vig)

input: /vi/	NO OBS CODA	DEP (V), MAX (C)	WD BIN
a. vig	*!		*
b. \emptyset vi.g \emptyset		*	
c. vi		*	*!

The coda stop in (10a) disqualifies the most faithful candidate in the first round. We then move to the next co-ranked constraints forbidding vowel insertion and consonant deletion. Because (10b) and (10c) tie at this stage, both survive, leaving the choice up to the next constraint. Candidate (10b), being bisyllabic, is the winner at this stage.

For the sake of completeness, we now consider the bisyllabic nonce form *fealig* /filig/, also from Wang, 1995:

(11) Interlanguage Form: Bisyllabic (*fealig*)

input: /filig/	NO OBS CODA	DEP (V), MAX (C)	WD BIN
a. fi.lig	*!		*
b. fi.lig \emptyset		*	*!
c. \emptyset fi.li		*	

The evaluation of this form proceeds much as with the monosyllabic form. The faithful form is eliminated by NO OBS CODA, and both (11b), with an epenthetic vowel, and (11c), with deleted consonant, violate the faithfulness constraints. The choice then falls to WD BIN, which in this case will choose the form with deletion, rather than the form with epenthesis, because that form best satisfies the requirement that words be bisyllabic.

In sum, Optimality Theory gives us a mechanism for understanding the role of factors such as the preference for bisyllabicity in the learner's developing grammar. In the native language, the role of the markedness constraint WD BIN is masked by higher ranked faithfulness constraints, which demand correspondence between input and output forms: WD BIN is not strong enough on its own to force the learner to choose the form with epenthesis or deletion just to get a bisyllabic form. However, target-language forms ending in stops cannot be accepted because they violate the highest ranked NO OBS CODA. Given an input that violates this constraint, the learner has no choice but to violate the faithfulness constraints. It is when faithfulness constraints must be violated that the lower ranked constraint becomes visible. Given a choice between two forms, both of which are unfaithful to the input, the learner chooses the form that best satisfies the markedness constraint WD BIN.

We have once again presented a somewhat idealized picture. The grammar developed in this section predicts that speakers will always choose epenthesis in monosyllabic forms with final stops and deletion in bisyllabic forms with final stops. Of course, the picture is more complicated. For one thing, some subjects produced some forms correctly; some devoiced final stops rather than deleting them, and the correlation between epenthesis versus deletion or devoicing and syllable number was not perfect.⁶ We should note, however, that this view of a grammar potentially provides for a wider range of possibilities by providing for different rankings. Because the movement from the native-language grammar to a grammar that more closely approximates that of target-language speakers involves reranking of constraints, we would expect the rankings of these constraints to be in flux.⁷

In the following section, we explore the question of what sort of grammar would allow for the correct production of some of the target-language forms—specifically, how the interlanguage grammar must differ from the native-language grammar in order to permit learners to produce final stops. We also consider the question of how this model can account for the learners' preference for the unmarked voiceless stop codas over the more marked voiced codas.

THE EFFECT OF MARKEDNESS: FINAL DEVOICING

Although both epenthesis and deletion are employed to transform English codas to fit the syllable structures of Mandarin, some Mandarin speakers also employ another strategy—devoicing. Because CVC(stop) syllables are not allowed in

Mandarin, the production of CVC(stop) cannot be motivated by the native language. Nor is the devoicing strategy motivated by the grammar of the target language because both voiced and voiceless obstruents are permitted in English codas. Furthermore, because Mandarin learners of English are able to pronounce voiced obstruents in onset position, devoicing of the coda obstruents cannot be viewed as a simple mapping between English voiced consonants and Mandarin voiceless unaspirated consonants. The choice of devoicing as a mechanism to fix undesirable English codas by Mandarin speakers is therefore independent of the native language and the target language. We argue that devoicing of English coda stops by Mandarin speakers provides a clear case of the emergence of the unmarked.

The Preference for Voiceless Stops

Universally, voiceless codas are less marked than voiced codas (see Major & Faudree, 1996, for evidence of this effect in interlanguages). Markedness of coda types can be formalized by a markedness constraint NO VOICED OBS CODA that penalizes the more marked voiced obstruents in the coda.

- (12) NO VOICED OBS CODA: Syllable codas may not contain voiced obstruents.

By hypothesis, NO VOICED OBS CODA is present in every language. Its effects are clearly visible in languages like German and Russian, which allow voiceless but not voiced obstruents in codas. However, the effects of NO VOICED OBS CODA are not visible in every language: English, which allows both types of coda, shows no visible effect of NO VOICED OBS CODA. The contrast between German and English is attributed to the relative ranking of NO VOICED OBS CODA in each language: NO VOICED OBS CODA is ranked high in German, but in English it is ranked below faithfulness constraints MAX (C), DEP (V), and IDENT (VOI), a constraint that mandates preservation of input voicing:

- (13) IDENT (VOI): An output segment should be identical in voicing to the corresponding input segment.

In Mandarin, the effects of NO VOICED OBS CODA are not visible because Mandarin forms are subject to the more general constraint NO OBS CODA. NO OBS CODA bans all obstruent codas, whereas NO VOICED OBS CODA bans only a portion of obstruent codas. We argued above that the ranking of constraints in the native-language grammar of Mandarin speakers is NO OBS CODA \gg MAX (C), DEP (V). The ranking of NO VOICED OBS CODA is indeterminate: as long as it is equal in ranking or lower ranked than NO OBS CODA, we will not see its effects. We assume the following rankings for languages exemplifying three coda patterns:

- (14) Constraint Rankings
- a. No coda obstruents (Mandarin): NO OBS CODA, NO VOICED OBS CODA \gg MAX (C), DEP (V), IDENT (VOI)
(*vlg, *vIk)
 - b. Only voiceless coda obstruents (German): NO VOICED OBS CODA \gg MAX (C), DEP (V), IDENT (VOI) \gg NO OBS CODA
(*vlg, vIk)
 - c. Both voiced and voiceless coda obstruents (English): MAX (C), DEP (V), IDENT (VOI) \gg NO VOICED OBS CODA, NO OBS CODA
(vlg, vIk)

However, some Mandarin subjects in Wang's (1995) study did choose devoicing, which indicates that these speakers have developed an interlanguage grammar that differs from both the native-language and the target-language grammars (but is similar to the grammars of German or Russian speakers). Note, however, that we are not forced to say that learners have acquired a rule that is present in neither the native- nor target-language grammars. Instead, we describe this as a reranking of constraints that are already present in the learner's native-language grammar. We assume that the Mandarin subjects who employ devoicing move NO OBS CODA down in the ranking, below both NO VOICED OBS CODA and IDENT (VOI). Once NO VOICED OBS CODA is ranked higher than NO OBS CODA, the possibility of voiceless obstruent codas is introduced. This is illustrated in (15).

(15) Mandarin Subjects Favoring Devoicing

input: /vlg/	NO VOICED OBS CODA	MAX (C), DEP (V)	IDENT (VOI)	NO OBS CODA
a. vIk			*	*
b. vlg	*!			*
c. vI		*!		
d. $\text{vI.g}\text{ə}$		*!		

NO VOICED OBS CODA rules out candidate (15b), and MAX(C) and DEP(V) rule out candidates (15c) and (15d). Because these subjects have departed from their native-language grammar by demoting NO OBS CODA, they choose the candidate with voiceless final coda, (15a), as optimal.

The grammar that predicts devoicing of English voiced codas also predicts an asymmetry between voiced codas and voiceless codas. Once a learner has demoted NO OBS CODA below the faithfulness constraints but maintained high ranking of NO VOICED OBS CODA, final voiceless stops can be correctly pronounced, as shown in (16).

(16) Mandarin Subjects Favoring Devoicing

input: mUk	NO VOICED OBS CODA	MAX (C), DEP (V)	IDENT (VOI)	NO OBS CODA
a. [☞] mUk				*
b. mUg	*!		*	*
c. mU		*!		
d. mU.kə		*!		

Eventually, learners who become proficient in English pronunciation will demote NO VOICED OBS CODA below the faithfulness constraints, thus allowing the full range of English coda consonants. None of Wang's (1995) subjects consistently produced all final voiceless stops, presumably due to the instability of the constraint rankings in the developing interlanguage grammars.

The asymmetry in voiced and voiceless coda stops in Wang's (1995) subjects is clearly an effect of the greater markedness of voiced obstruents, as compared to voiceless obstruents, in syllable coda. This effect has long been recognized as playing a role in second-language phonology (Eckman, 1977), but the precise mechanism by which this markedness effect shapes the learner's grammar has been unclear. We propose that the constraint NO VOICED OBS CODA is present in the grammar of both Mandarin and English speakers. However, in Mandarin this constraint is ranked below NO OBS CODA, and its effects are therefore invisible. To move toward a grammar more closely approximating English, in which NO OBS CODA is very low ranked, Mandarin speakers must demote NO OBS CODA. Some of these speakers demote NO OBS CODA but leave NO VOICED OBS CODA highly ranked. This ranking permits final voiceless obstruents but not final voiced ones.⁸

Devoicing versus Epenthesis

Earlier we described the grammar of Mandarin subjects who favor epenthesis for monosyllabic inputs and deletion for bisyllabic inputs. We now consider subjects who also favor bisyllabic output forms but devoice a final stop coda in bisyllabic forms rather than deleting it. We described the epenthesis in monosyllables and deletion in bisyllables pattern as resulting from the ranking of syllable-structure constraints (e.g., NO OBS CODA, NO VOICED OBS CODA) above faithfulness constraints (e.g., MAX (C), DEP (V), IDENT (VOI)), which are in turn ranked above WD BIN, which favors bisyllabic words. We can describe the epenthesis in monosyllables and devoicing in bisyllables with the ranking shown in (17), in which NO OBS CODA is demoted below WD BIN and MAX (C), DEP (V).

(17) Mandarin Subjects Favoring Devoicing: Bisyllabic Output

input: filg	NO VOICED OBS CODA	WD BIN	MAX, DEP	IDENT (VOI) NO OBS CODA
a. [☞] fi.lɪk				*
b. fi.lɪg	*!			
c. fi.lɪ.gə		*!	*	
d. fi.lɪ			*!	

The optimal bisyllabic form, according to this ranking, is one that avoids violating NO VOICED OBS CODA through violation of the low-ranked IDENT (VOI). With a monosyllabic form, however, WD BIN will favor the form with epenthesis because that is the only bisyllabic candidate, as shown in (18).

(18) Mandarin Subjects Favoring Devoicing: Monosyllabic Output

input: /vɪg/	NO VOICED OBS CODA	WD BIN	MAX, DEP	IDENT (VOI) NO OBS CODA
a. [☞] vɪ.gə			*	
b. vɪg	*!	*		
c. vɪ		*!	*	
d. vɪk		*!		*

CONCLUSION

When research in second-language phonology was carried out in a framework in which all phonological modifications were viewed as the effect of language-specific rules, one persistent question arose: How does the language learner get rules that are not motivated by either the target language or the native language? Often, the surface forms produced by these rules were less marked than the target-language forms, and these facts raised a related question: How do markedness constraints shape the speaker's grammar? Optimality Theory dispenses with language-specific rules, assuming instead that the universal function GEN provides speakers with all possible modifications of an underlying representation. Universal Grammar also provides speakers with a set of universal constraints (many of which encode markedness effects by penalizing marked structures). Language learners' modifications of underlying representations (via epenthesis, deletion, devoicing, etc.) and learners' tendency to favor less marked structures in interlanguage grammars can be seen simply as effects of universal constraints. In some cases, a constraint that plays a role in the interlanguage

grammar may be invisible in the native-language grammar for one of two reasons: Either target-language inputs do not provide the range necessary to allow this low-ranked constraint to have any visible effects, or the constraint may be ranked below constraints that always mask its effects. The sudden visibility of these constraints in second-language acquisition may result from the richness of the target-language inputs or from the development of an interlanguage grammar that differs from the native-language grammar in its constraint rankings. For example, a Mandarin speaker confronted with English forms ending in obstruents has clear evidence that NO OBS CODA is not a dominant constraint in the target language. Thus, the model of second-language acquisition proposed here is one in which learners, under pressure from interlanguage data, begin to construct an interlanguage grammar in which the rankings of constraints may differ from the native-language ranking. In this case, markedness effects that are not visible in either the native or the target language may become visible in the interlanguage data.

NOTES

1. But see Stampe (1979) for the proposal that some rules are innate. See also Hayes (1996) for the proposal that some constraints within Optimality Theory are induced from the data.

2. We note that place of articulation had no effect on choice of strategy. The difference in error rate for different places of articulation (e.g., /p/ vs. /t/) was not significant (the greatest difference in number of errors between any two segments within the categories of voiced or voiceless stops had a chi-square value of 2.9, which at the level $p = .05$ was not significant). Voicing also had no effect on the choice of epenthesis versus deletion.

3. The relationship between the underlying representations posited by learners and by native speakers is a thorny issue. As one reviewer pointed out, strong evidence exists suggesting that native-language phonotactic constraints may affect the perception of English forms (Flege & Wang, 1989; Silverman, 1992; Yip, 1996a; though see Broselow & Park, 1995, for a case in which learners appear to perceive target-language contrasts that they cannot produce). Wang's (1995) experiment, in which target forms were presented both orally and in writing, was designed to bias the subjects toward observing the native-language contrasts, though no test was done to determine whether subjects who produced voiceless final stops in place of target-language voiced stops actually did perceive voicing in the target-language forms.

4. Stress pattern is, however, a significant factor: Wang found that the difference between deletion rates of the final consonant in initially stressed bisyllables (72% of 60 tokens) and finally stressed bisyllables (53% of 60 tokens) was significant at the level of $p < .05$ (χ^2 3.91). The difference between epenthesis rates in initially stressed bisyllables (5% of 60 tokens) and finally stressed bisyllables (30% of 60 tokens) was significant at the level of $p < .005$ (χ^2 13). For the purposes of clarity, we have collapsed a number of constraints into (Word Binarity), but WD BIN actually abbreviates several different constraints, including one constraint identifying iambic feet as optimal and another identifying trochaic feet as optimal. We could account for the effect of stress in certain subjects' productions as an effect of particular rankings of these constraints on foot type.

5. See for example Prince and Smolensky (1993) and McCarthy and Prince (1993); see also Yip (1990) on a preference for iambic binary feet in Cantonese loanwords.

6. Wang (1995) notes that stress was a significant factor in the prediction of epenthesis versus deletion, though considerably weaker than syllable number. The set of constraints alluded to earlier, collapsed as WD BIN, would allow for the possibility of favoring different arrangements of stressed versus unstressed syllables, but the full explication of this is beyond the range of this paper.

7. See Davidson (1997) for proposals concerning an Optimality Theory account of second-language acquisition.

8. We can compare this account to a model of second-language learning as parameter resetting. In the parameter-resetting account, we could assume that the learners who devolve the English voiced obstruent codas have reset their parameters from the setting allowing no obstruent codas to the setting allowing only voiceless obstruent codas. However, if we assume that learners do perceive the

contrast between voiced and voiceless codas in English, this provides no mechanism by which the learners transform target-language voiced codas to the corresponding voiceless stop. To do so would still require the addition of a devoicing rule to the grammar.

REFERENCES

- Broselow, E. (1988). Second language acquisition. In F. Newmeyer (Ed.), *Linguistics: The Cambridge survey. Volume III: Language: Psychological and biological aspects* (pp. 194–209). New York: Cambridge University Press.
- Broselow, E., & Finer, D. (1991). Parameter setting in second language acquisition. *Second Language Research*, 7, 35–59.
- Broselow, E., & Park, H. (1995). Mora conservation in second language prosody. In J. Archibald (Ed.), *Phonological acquisition and phonological theory* (pp. 151–168). Hillsdale, NJ: Erlbaum.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.
- Davidson, L. (1997). An optimality theoretic approach to second language acquisition. Unpublished honors thesis, Brown University, Providence, RI.
- Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, 27, 315–330.
- Eckman, F. (1981). On the naturalness of interlanguage phonological rules. *Language Learning*, 31, 195–216.
- Epstein, S., Flynn, S., & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences*, 19, 677–758.
- Flege, J., & Wang, C. (1989). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/-/d/ contrast. *Journal of Phonetics*, 17, 299–315.
- Hayes, B. (1996). Phonetically driven phonology: The role of Optimality Theory and inductive grounding. Paper posted on Rutgers Optimality Archive, ROA-158-1196.
- Heyer, S. (1986). English final consonants and the Chinese learner. Unpublished master's thesis, Southern Illinois University, Edwardsville.
- Ito, J., & Mester, A. (1995). The core-periphery structure of the lexicon and constraints on reranking. In J. Beckman, L. Dickey, & S. Urbanczyk (Eds.), *Papers in Optimality Theory, University of Massachusetts Occasional Papers*, 18, 181–210.
- Kenstowicz, M. (1994). *Phonology in generative grammar*. Cambridge, MA: Oxford.
- Major, R., & Faudree, M. (1996). Markedness universals and the acquisition of voicing contrasts by Korean speakers of English. *Studies in Second Language Acquisition*, 18, 69–90.
- McCarthy, J., & Prince, A. (1993). Prosodic morphology I: Constraint interaction and satisfaction. Unpublished manuscript, University of Massachusetts at Amherst and Rutgers University, New Brunswick, NJ.
- McCarthy, J., & Prince, A. (1994). The emergence of the unmarked: Optimality in prosodic morphology. *Proceedings of the North East Linguistics Society*, 24, 333–379.
- McCarthy, J., & Prince, A. (1995). Faithfulness and reduplicative identity. In J. Beckman, L. Dickey, & S. Urbanczyk (Eds.), *Papers in Optimality Theory, University of Massachusetts Occasional Papers*, 18, 249–384.
- Prince, A., & Smolensky, P. (1993). Optimality Theory: Constraint interaction in generative grammar. Unpublished manuscript, Rutgers University, New Brunswick, NJ.
- Silverman, D. (1992). Multiple scansion in loanword phonology: Evidence from Cantonese. *Phonology*, 9, 289–328.
- Stampe, D. (1979). *A dissertation in natural phonology*. New York: Garland.
- Wang, C. (1995). The acquisition of English word-final obstruents by Chinese speakers. Unpublished doctoral dissertation, State University of New York at Stony Brook, NY.
- Weinberger, S. (1988). Theoretical foundations of second language phonology. Unpublished doctoral dissertation, University of Washington, Seattle, WA.
- Yin, Y. M. (1984). A comparative study of the interlanguage phonology of two Chinese dialect groups. Unpublished manuscript, The University of Texas at Austin, TX.
- Yip, M. (1990). The phonology of Cantonese loanwords: Evidence for unmarked settings for prosodic parameters. Paper presented at the Second Northeastern Conference on Chinese Linguistics, University of Pennsylvania, Philadelphia, PA.
- Yip, M. (1996a). Cantonese loanword phonology and Optimality Theory. *Journal of East Asian Linguistics*, 2, 261–291.
- Yip, M. (1996b). Lexicon optimization in languages without alternations. In J. Durand & B. Laks (Eds.), *Current trends in phonology: Models and methods* (pp. 757–786). Paris: University of Salford Publications.