[published 2002 in Journal of Chinese Phonology 11:163-190 with different pagination]

AN ANALOGICAL APPROACH TO THE MANDARIN SYLLABARY*

James Myers
National Chung Cheng University

The standard argument for the psychological reality of phonological grammar, namely the ability of speakers to make judgments on nonwords, does not eliminate the possibility that speakers generalize from exemplars by analogy. This paper makes a case for such an analogical approach, demonstrates how analogy can be formalized in various ways, and then applies analogy to native-speaker judgments of Mandarin syllables. The major empirical findings (judgments about real vs. nonreal syllables are much sharper than about systematic vs. accidental gaps, and there is no significant difference between violations of apparent universal constraints and of ad hoc language-specific constraints) are argued to conform better to an analogical than to a grammatical approach to phonology, though challenges remain in choosing the formal analaogical model that best handles the data.

KEYWORDS: analogy, Mandarin, syllable

1. Introduction.

A standard argument against what Kenstowicz and Kisseberth (1979) call the "null hypothesis" (i.e. that phonology doesn't exist) is the ability of native speakers to make judgments on nonwords. For example, English speakers know that *blick* is a possible word but *bnick* is not, and since neither are real words, this cannot be explained by memorization alone. However, such arguments do not eliminate an extended version of the null hypothesis, which allows that speakers can generalize from memorized forms by analogy. Thus since there are many real English words beginning with /bl/ but none beginning with /bn/, it could be that it is this, rather than grammar per se, that speakers base their judgments on. This extended null hypothesis might be termed an analogical approach to acceptability judgments, as opposed to a grammatical approach, which claims that acceptability judgments derive from native speakers' on-line application of grammatical rules, principles, and/or constraints.

This paper describes an exploration of the analogical approach in the domain of the Mandarin syllabary (i.e. the set of syllables that are found in actual Mandarin words). There are at least two reasons why this is a worthwhile exercise. First, the Mandarin syllabary is almost ideally suited to basic research on phonological analogy because of its small size (even if tone is taken into account, Mandarin has far fewer than 1500 distinct syllables; Ho 1976) and lack of morphological influence (in this paper, syllables formed via r-suffixation will be ignored, since they are virtually absent in Taiwan Mandarin). Since analogical models generalize patterns from a set of lexical exemplars, these properties make it easier to define and process this set. Second, analogical models of the Mandarin syllabary can easily be

* The author would like to thank 蔡素娟, 程涂媛, 李家銘, 黃婷鈺, and 林清淵 for much help with the design of the experiments and collection of data, and the audience at the Seventh International and Nineteenth National Conferences on Chinese Phonology for comments and discussion.

tested against actual native speaker judgments, since data on such judgments already exist; Wang (1998) reports on just such a study, and this paper provides further data that corroborates and supplements Wang's.

The remainder of this paper is organized as follows. Section 2 attempts to deal with common objections to the notion of analogy, and then provides some evidence for it in synchronic phonology. Section 3 explains how analogy can be formalized in explicit models. Here special attention is given to a formalization that relies on nothing but devices already used in the generative literature on Optimality Theory (OT). Section 4 turns to the empirical problems posed by the Mandarin syllabary, describing experimentally collected native speaker judgments on existent and nonexistent syllables. Finally, section 5 compares these judgments with the predictions made by analogical models implemented on a computer. The results, while somewhat promising, have not proven as successful as hoped, but as will be seen, there are many different ways that analogy can be formalized; searching for the right one (if any exists) is a difficult, though often quite enlightening, task.

2. Phonological analogy.

This section gives a brief overview of the notion of analogy, first discussing two of the most common objections to the use of the analogical approach to synchronic phonology and showing why they may not be as serious as is commonly thought. Then some empirical arguments are given for the relevance of analogy to the description of synchronic phonology, both in English and in Mandarin.

2.1 Objections to analogy.

The first reason why the analogical approach to synchronic phonology has not been popular in linguistics, especially generative linguistics, is that analogy can only work to help spread a pattern if there is already something of a pattern there to start with. But where do linguistic patterns come from in the first place? The traditional answer in generative linguistics has been that they are in some sense innate, deriving from "natural" and "universal" principles (e.g. in Optimality Theory terminology, they are a consequence of the set of universal Structure constraints encoding unmarkedness).

There is another answer to this question, however, and that is to suppose that phonological patterns originate outside of phonology proper, as the result of systematic forces working beyond the confines of a single human brain (e.g. physiology, psychoacoustics, and the vagaries of history). Over the past few decades, there has been growing acknowledgement of this alternative answer in generative circles. The most common way this has been expressed has been to use OT Structure constraints that are explicitly physical in nature (e.g. Jun 1995, Silverman 1996, Kirchner 1997, Myers 1997). However, a more appropriate analytical solution may be to separate the causes of phonological patterns from their processing by individual speakers. For example, while it may be true that the [k]~[s] alternation in *electric-electricity* is phonetically natural in some sense (e.g. as a kind of lenition), there is no evidence suggesting that this naturalness plays any active role in the minds of modern-day speakers. As a matter of fact, this [k]~[s] alternation is not very systematic, in spite of its naturalness; it is restricted to a small set of non-Germanic suffixes, and as Myers (1993) has shown, almost exclusively to the suffix *-ic*. Moreover, speakers don't readily generalize it to novel words (see e.g. Ohala 1974). For reasons like these (and others to follow), a modular approach to phonological theory seems called for, where separate

subtheories handle the ontogenesis (e.g. physics) and maintenance (e.g. analogy) of phonological patterns.

Another major argument (more properly, set of arguments) against analogy has been that it doesn't seem able to make any specific predictions about which analogies are possible and which are impossible. Part of the problem is the structure-dependence of language: analogies often don't work correctly if one assumes the wrong sort of representations. To use a famous example of Chomsky (1975), if one doesn't assume any syntactic structure, it's possible to generalize from sentence pairs like "The boy is here"/"Is the boy here?" to "The boy who is here is sad"/*"Is the boy who here is sad?" (i.e. by positing that the pattern involves moving the first "is", rather than the first matrix-clause "is"). There are two responses to this criticism that analogy advocates can take. First, one can say that an analogical approach does not claim that representations are necessarily without structure; there may indeed be innate restrictions on the sorts of representations that the human brain can store and process. By reducing the generative apparatus to a single mechanism (analogy), linguists can actually be more clear about what restrictions on the representations must be like. There is also a second response: perhaps the necessity of structured representations has been overrated. Indeed, amazing successes have been achieved by analogical models that assume entirely "flat" representations, e.g. Elman's (1992) model of syntax, which learns about long-distance dependencies without explicitly being taught about phrase structure, and the model of Dell, Juliano, and Govindjee (1993), which mimics the way speech errors respect syllable structure, in spite of the fact that the model is not explicitly trained about syllable structure. In both cases, the models learn about the structures inherent in the input themselves, so there is no need to make knowledge of such structure innate.

However, even if analogy is assumed to work with structured representations, it seems to many linguists to be impossible to decide systematically which exemplars are relevant and which are not. Kiparsky (1988) gives an amusing example of this, pointing out that *ears hear*, so why can't one say, by analogy, that *eyes *heye*? To respond to this sort of problem linguists need explicit models of how analogy operates. True exemplar-driven analogy has traditionally been difficult to model because it refers to memorized exemplars, and there may be thousands upon thousands of them. Analogical modeling has only really begun to blossom with the development of computers. For example, both the models of Elman (1992) and of Dell et al. (1993) are computer models of analogy (connectionism). Explaining how formal models of analogy work is the purpose of section 3.

2.2 Evidence for analogy in English.

This section briefly sketches out evidence for analogy in English. Irregular inflection is a safe place to start, since even scholars who have argued forcefully against analogy in general (in the form of connectionism) admit its usefulness in an analysis of irregular inflection (e.g. Pinker and Prince 1992). One reason for thinking that analogy is relevant to alternations like *ride-rode* and *write-wrote* is the fact that they show family resemblances (Bybee and Slobin 1982): similarities between pairs of category members that create cohesiveness even in the absence of a single feature or set of features shared by all members. By definition, family resemblances cannot be handled with general rules or constraints, but they are expected with analogy. For example, the subset of irregular English verbs that show the alternation [aj]~[o] form a coherent set because for every member one can find another member that is closely similar to it; thus the coda of *drive* differs from that of *rise* only in Place, that of *rise* from that of *ride* only in [continuant], and that of *ride* from that of *write*

only in [voice]. Irregular inflection also shows what might be termed nondirectionality: morphologically derived forms behave as if they are listed, since they can also be subject to analogies or trigger analogies in base forms. In the English irregular verbs, nondirectionality can be seen in the fact that strong family resemblances are found among past forms ostensibly derived from quite different base forms (e.g. *buy-bought*, *bring-brought*, *seek-sought*, *catch-caught*).

Even beyond irregular inflection, phonology can behave like analogy. To take just one example from Myers (2000), it has been claimed that English phonology includes a rule of s-voicing (Chomsky and Halle 1968), which voices /s/ intervocalically, as in the pairs *Pari**s**-Pari**s**ian*, and ***s**ign-de**s**ign*. However, this is actually just part of a larger set of word pairs showing family resemblances (***s**olve-di**ss**olve*, *in**s**ist-per**s**ist*, *Mar**s**-Mar**t**ian*) and nondirectionality (*diver**t**-diver**s**ion*, *submer**g**e-submer**s**ion*, *cohe**r**e-cohe**s**ion*, *equa**t**e-equa**t**ion* [cf. *rela**t**e-rela**t**ion*]). Patterns like these are beginning to emerge from the footnotes to take center stage in generative analyses. For example, Hayes (1999) observes that linguists often pronounce *consonantal* as *conson[ɛ]ntal* instead of *conson[æ]ntal*, arguing that this probably occurs by analogy with words like *accidental* and *continental* (note that in the unsuffixed forms all of these contain a schwa in the final syllable).

Analogy is not restricted to patterns involving morphology, but affects the phonotactics of monomorphemic forms as well (e.g. *blick* vs. *bnick*). An important source of evidence for this comes from the phenomenon of lexical diffusion (e.g. Wang 1969, 1977, Labov 1981, Bybee 2000), in which a sound change spreads through a lexicon word by word (or word class by word class), rather than all at once in Neogrammarian fashion; morphology need not be involved. Many cases have been described in the quantitative sociolinguistic literature. For example, Labov (1981) describes a case in Philadelphia English in which /æ/ is changing from a more lax to a more tense articulation; some speakers pronounce *bad*, *mad*, and *glad* with the newer tense pronunciation, but *sad* and *dad* still have the older lax form. While this synchronic pattern could be handled simply with distinct underlying representations, the forces that give rise to it surely include psychological processes of interest to the study of synchronic phonology (and sociolinguists would perhaps add that the division between "diachronic" and "synchronic" is not as sharp as Saussureans have claimed). Moreover, lexical diffusion often gives rise to patterns of definite interest to generativists. For example, Labov (1981) reports that some Philadelphians preserve the lax vowels in irregular verbs *ran*, *swam*, and *began*, but tense the vowels in words like *man*, *Dan*, and *slam*. A similar pattern is found in a variety of Scottish English spoken in Glenoe, Ireland: the base forms of irregular verbs (e.g. *rise*, *drive*) are exceptions to a rule that normally lengthens vowels before voiced continuants (Gregg 1973). Note that both of these cases show nondirectionality: *base* forms are marked as phonological exceptions because their *derived* forms are irregular. In addition, Robinson (1977) gives examples from German dialects showing how lexical diffusion caught mid-way can cause rule ordering paradoxes in sychronic phonology. Given the word-by-word nature of lexical diffusion, it's not surprising that the best models of the phenomenon involve analogy (e.g. the interesting mathematical model of Shen 1990).

There is also psycholinguistic evidence for the role of analogy in making judgments about phonotactics. Several researchers have found that English speakers are sensitive to the mere statistical frequency of particular combinations of phonemes, as found in their mental lexicon, even if there are no violations of what linguists would consider true phonotactic constraints. For example, Ohala and Ohala (1986) asked adult English speakers to compare the acceptability of nonwords like /klæb/ and /klɛb/. The onset /kl/ and the rimes /æb/ and

/ɛb/ are all found in English words (e.g. *climb*, *lab*, *ebb*), yet it so happens that /æb/ appears in more real words also containing /k/ or /l/ than does /ɛb/. An analogical approach to phonological acceptability judgments thus predicts that /klæb/ should be considered more acceptable than /klɛb/, and that is indeed what Ohala and Ohala found. Moreover, a traditional phonotactic analysis (or its modern OT equivalent) does not seem possible here, since as just noted, no pairing of phonemes in either nonword is categorically disallowed (as /bn/ is in *bnick*). Similar results have been found in experiments by other researchers using a variety of techniques, including on-line reaction-time measures (e.g. Vitevich, Luce, Pisoni, and Auer 1999). Based on such evidence, it appears that speakers can indeed derive acceptability judgments from memorized exemplars directly, without the use of grammar.

A final argument rests on the fact that the analogical approach distinguishes naturalness from systematicity (e.g. generality within a language), whereas the generative grammar approach conflates the two notions. It is of course true that most phonotactic constraints are both systematic and natural (e.g. the */bn/ constraint follows from a universal constraint against clusters with consonants too close in sonority, a constraint that also disallows */tf/ and */nr/). Nevertheless, systematicity and naturalness do not always cooccur. In such cases, an analogical approach predicts that systematicity alone can account for productivity to novel forms, while the grammatical approach might predict that naturalness should be more important (e.g. via the OT notion of "emergence of the unmarked"; McCarthy and Prince 1995b). While apparently no one has explored these predictions carefully, some simple cases seem to weigh in on the side of the analogical approach. For example, English has a constraint against *sC_1VC_2, where C_1 and C_2 have the same Place (except Coronal): *spap, *smap, *skak, *skag; cf *state.* As Davis (1991) has shown, this constraint is phonologically natural; it is essentially a variation of the Obligatory Contour Principle (OCP), with the exceptional behavior of Coronal due to underspecification (perhaps universal). Nevertheless, there is no sign that this constraint has any psychological force on native speakers at all; for example, it didn't prevent the relatively new words *Spam* and *skunk* from entering the lexicon. Presumably this is because the pattern is not systematic enough in the lexicon for strong analogies to form.

2.3 Evidence for analogy in the Mandarin syllabary.

As is often noted, most of Mandarin phonology is phonotactics. The arguments for analogy here run quite parallel to those for English. The argument from lexical diffusion doesn't need elaboration; this phenomenon was in fact first discovered in the study of diachronic Chinese phonology (Wang 1969). Statistical phonotactics (of the /klæb/ vs. /klɛb/ sort) have not been explicitly studied in Mandarin, but some of the results of Wang (1998) can be understood this way. As will be discussed more fully in section 4, Wang (1998) found that native Mandarin speakers judged nonwords created by changing the tone (e.g. [xeǐ^{51}]) as more acceptable than phonotactically legal nonwords created by changing a phoneme (e.g. [pou^{55}]; cf. [p^{h}ou^{55}] 剖). By definition, then, neither kind of nonlexical form violates Mandarin phonotactics, so the difference in judgment needs to be explained some other way. One way would be to suppose that tone is somehow perceptually less salient than segmental information, in spite of its phonemic status in Mandarin; this would be consistent with other studies that have found that Mandarin tone is less important in psycholinguistic processing than was once thought (e.g. Taft and Chen 1992, Ye and Connine 1999). But there is also an analogical explanation: a given tone cooccurs with a wider variety of segments than segments do with each other. For example, Tone 4 appears in many lexical

syllables with [x] or [ei], but [p] doesn't combine with as many words with [o] or [u].

The Mandarin syllabary, moreover, makes the comparison of naturalness and systematicity quite convenient, since these two properties can be doubly dissociated. As an example of a natural constraint that is not systematic, consider the following curious gap: syllables both beginning and ending with Coronal segments are disfavored when there is a single intervening high front vowel or an underlyingly unspecified vowel (i.e. schwa).

(1) *[tʰin], *[tin], [nin] (1 lexical item: **您**), [lin] (OK)
*[tʰyn], *[tyn], *[nyn], *[lyn]
*[tʰən], *[tən], [nən] (1 lexical item: **嫩**), *[lən]
*[tʰei], [tei] (1 lexical item: **得**), [nei] (3 lexical items: **那**, **內**, **哪**), [lei] (OK)

As Myers (1995) has argued, this appears to be a long-distance OCP effect similar to the English *sCVC constraint noted earlier (though now of course Coronal cannot be unspecified). The relevance of the vowels makes sense because an unspecified vowel is incapable of blocking OCP effects, and front vowels can themselves be analyzed as Coronal (e.g. Clements and Hume 1995). In spite of this naturalness, however, there is absolutely no evidence that native speakers consider the above marked items to be more unacceptable than any purely "accidental" gap. For example, the putative constraint didn't prevent **您** from entering the lexicon nor **嫩** from changing its pronunciation from the older [nun^{51}]. Native speakers seem to value systematicity over naturalness, and this pattern simply isn't systematic enough.

For an example of a systematic but apparently unnatural constraint, consider another curious gap: Mandarin is very stingy in its use of the rime /ia/. It is acceptable alone (e.g. **鴨**) and with palatals (e.g. **家**, although such syllables are often analyzed as syllables of the form C^{j}V, with a palatalized onset; see e.g. Duanmu 1990). But it is not found with labial and alveolar onsets, onsets that normally allow /i/, and which (making things even more mysterious) do allow /ia/ if this string of phonemes is followed by another segment (e.g. **點**, **兩**, **標**), and also allow /i/ as a medial if followed by another vowel (e.g. **貼**, **丟**). There is no obvious "natural" explanation for why this should be so, nor does there appear to be restrictions quite like this in any other language. This strange constraint thus makes a sharp contrast with a superficially very similar one, namely the disallowing of /ua/ after labial onsets (which can be, and has been, analyzed as an OCP effect; see e.g. Duanmu 1990). However, they are different primarily in naturalness; in terms of systematicity they are both roughly equivalent, since both account for the absence of relatively large sets of unattested syllables. A nonanalogical model which claims that naturalness is a psychologically active notion therefore predicts a sharp difference in speaker judgments of syllables like *pia* and *pua*. As will be shown in section 4, there is as yet no evidence that such a difference exists.

A further argument that can be added comes from lexicality judgments. From a generative perspective, it is something of an accident whether or not a grammatically well-formed item is in fact a genuine lexical item (see e.g. Di Sciullo and Williams 1987 for disparaging remarks about mere "listemes"). Nevertheless, it is well known that speakers can easily judge the difference between real and non-real words. In Mandarin, anecdotal evidence suggests that such judgments are particularly sharp; asking even trained Mandarin-speaking linguists to decide whether a given item is "possible" or "impossible" is often a fruitless task, since they only feel really certain about whether or not it is a real word.

Moreover, while nonexistent syllables are invented rather freely in English (e.g. as the names of new products), this does not seem to be the case in Mandarin. A lexicality effect makes sense in an analogical approach, which says that judgments derive from the real words themselves, and the difference between English and Mandarin in the strength of the lexicality effect also makes sense: Mandarin speakers develop their "acceptability metric" in a relatively homogeneous lexical space, and so they become less sensitive to small differences among nonwords than do English speakers.

3. Formal models of analogy.

There are several kinds of analogical models that have been proposed and studied in recent decades. While they differ in technical details, they all share the same goal: to derive productive linguistic behavior from memorized exemplars, rather than from general grammatical rules, principles or constraints. They are also all mathematically and/or computationally explicit; contrary to what was often the case when Neogrammarians invoked analogy, with these models no hand-waving is allowed. This section describes two simple models that have been used to generate predictions for comparison with native speaker judgment data, and shows how (perhaps surprisingly) they can be expressed in an Optimality-Theoretic framework.

3.1 A simple nearest-neighbor model.

One of the oldest and most intuitive ways to formalize analogy, a nearest-neighbor model generates responses to an item that is not in the lexicon by how similar it is to an item or set of items that is in the lexicon (e.g. Daelemans 1998). Parameters may be varied within this general framework, including representational codes (e.g. phonetic features vs. phonemes), measures of distance (e.g. the minimum number of features by which the item differs from a potential lexical neighbor), and the use of frequency (e.g. it may count more to be near a common word than a rare one). Later section 5 will describe a simple nearest-neighbor model that merely counts the number of shared features between target items and lexical items.

3.2 The Greenberg and Jenkins model.

The experiment of Ohala and Ohala (1986) mentioned above was designed to test the more sophisticated nearest-neighbor model of Greenberg and Jenkins (1964). According to this model, a target item's neighbors are defined as real words that can be made by changing one or more phonemes of the target item; an "acceptability" rating is then calculated by counting the number of such substitutions that do not create a real word. For example, /klæb/ can be transformed into a real English word by changing any one or more of its phonemes (e.g. the first can be changed to form *slab*, the last two can be changed to form *clip*, etc). By contrast, there are fewer substitutions for /klɛb/ that will result in a real word; e.g. there is no way to change the first phoneme to make a real word (**sleb*, **pleb*, etc). A major advantage of this model over the simpler one is that it is capable of handling true phonotactics, not the mere sharing of features. For example, both *bnick* and *blick* are one phoneme away from a real word, *brick*. To understand why *bnick* is nevertheless worse, it is necessary to have a way of recognizing that it contains both /b/ and /n/ in the onset at the same time. The Greenberg and Jenkins algorithm can express this, because *blick* will have many neighbors sharing both /b/ and /l/ whereas *bnick* will only have neighbors with neither /b/ or /n/, or just one of them, but never both. However, when implementing the Greenberg-Jenkins model on

a computer, a serious deficiency becomes obvious: the required processing speed and memory capacity increases by a factor of 2^n, where n is the number of features (because the number of subsets of a set with n members is 2^n). Further details on this model will be given in section 5.

3.3 Optimality-Theoretic models of exemplar-driven analogy.

This section first discusses some of the properties that make Optimality Theory (OT) more similar to an analogical approach than previous generative models and show how explicitly analogical analyses are becoming more common in the OT literature. This discussion will then lead to a description of a fully analogical OT model, or more precisely, pair of models. The first is designed to mimic the simple nearest-neighbor described in 3.1, and the second to mimic the Greenberg and Jenkins algorithm described in 3.2.

3.3.1 The basics of OT analogy.

The most obvious properties that make OT analogy-friendly are that it is non-derivational and surface-based. These properties result from OT's being a descendent both of generative theories of linguistic constraints and of constraint-satisfaction connectionist networks (see Prince and Smolensky 1997). Another important property of OT is that it posits Faithfulness constraints, whose sole job it is to require forms to be "faithful" to themselves or to other forms, that is, to prevent the Structure constraints from being effective.

Central to an analogical approach in OT is to use Faithfulness constraints that are parochial (using the terminology of Hammond 1995), that is, universal constraints that are parameterized by lexical item. Such constraints are entirely mainstream in the OT literature (e.g. the constraint EDGEMOST(Left, *um*), crucial to McCarthy and Prince's (1993) famous analysis of Tagalog *um* infixation). Of course, input-output faithfulness (e.g. IDENT-IO) is not sufficient to handle analogy, which involves relations between lexical exemplars. However, starting with McCarthy and Prince (1995a), Faithfulness has been reformalized in correspondence theory, which allows Faithfulness constraints to refer to two parts of a single output (stem and reduplicant in reduplicated forms), or even to pairs of morphologically related output forms (e.g. Kenstowicz 1995, Benua 1997). Output-output (OO) correspondence allows for analyses that are strikingly similar to traditional theories of analogy. For example, a blatant use of paradigm leveling forms the basis of Benua's (1997) analysis of morphophonology (e.g. the deletion of /n/ in *condemn* and *condemnable* but not in *condemnation*). Other applications of OO-constraints to similar "analogical" problems may be found in Steriade (2000) and Kenstowicz (1995), among many other places.

The logical next step is to try to build an OT model of true exemplar-driven analogy. This would be a model using only Faithfulness constraints, with most of the work done by OO-constraints in particular. Myers (2001) describes how such a model can be built to handle analogy involving morphology, such as that required for proportions like *drive* : *drove* :: *dive* : *dove.* To get to this to work, the crucial step is to posit parochial OO-constraints of the form IDENT-OO(W_i,W_j;F_k), where W_i and W_j are words (or word tokens) and F_k is some feature (and as noted earlier, determining how representations are encoded in an analogical model is crucial). Constraints of this form will cause words to become similar to one another; they are counteracted by IO-constraints of the form IDENT-IO(W_i;F_k) that cause words to resist change. Thus to deal with the notoriously capricious nature of analogy (which often fails to apply in one language in precisely the environment where it readily

applies in another), the model simply exploits OT's principle of extrinsic ranking. Paradoxically, extrinsic constraint ranking also turns out to provide a neat account of universal properties of analogy, such as family resemblances. The reason for this is that OT constraints can in principle be extrinsically ranked in every possible way cross-linguistically. Examining the quantitative predictions of the completely random ranking of analogical conjoined constraints shows that the probability that a given form will be changed by a given analogy is determined entirely by the number of triggering analogical constraints. For example, the more similar a target form is to an analogical trigger, the more features they will share, and thus the more analogical constraints there will be that are parochial with respect to those words (i.e. one such constraint per shared feature). For mathematical demonstrations of this and other interesting properties of the model, the reader is referred to Myers (2001); for example, it is shown there that the OT model just described is equivalent to a simple kind of connectionist network.

3.3.2 A simple nearest-neighbor model in OT.

Applying such an OT analogical framework to the problem of phonotactic judgments is rather straightforward. The goal is to determine how likely a form *a* is to be marked as optimal (i.e. grammatical, or a possible word) given a set of parochial OO-constraints of the form IDENT-OO(*a*,*x*;F), where *x* represents a real word and F a feature. Some such constraints will favor *a* (if the real word *x* happens to share feature F with it), and some will disfavor *a*. Given completely free constraint ranking, the probability that *a* will be allowed as grammatical is simply the sum of all the constraints that favor it divided by the number of all constraints that distinguish among candidates (see Myers 2001). For example, consider the nonwords *blick* and *bnick*. For simplicity suppose that they are represented with "features" that are actually position-specific phonemes; that is, the /l/ in *blick* might be represented as $/l/_2$, since it is in the second segmental slot of the word. Now consider the following two OT tableaux, which show variably ranked OO-constraints relevant to *blick* and *bnick* respectively ("¬blɪk" represents all outputs that differ from *blick* in the relevant respect). Since there are many words that have /l/ in second position but none that have /n/ in that position, the number of constraints that favor *blick* will outnumber those that favor *bnick*, and the result will therefore be that the probability that *blick* will be allowed is higher than that for *bnick*.

(2) a. The acceptability of *blick*

. . . (constraints favoring *blick*) . . . (disfavoring constraints)

blɪk	IDENT-OO (*blick*,*bliss*; $/b/_1$)	IDENT-OO (*blick*,*bliss*; $/l/_2$)	...	IDENT-OO (*blick*,*brick*; $/r/_2$)	...
blɪk				*	*
¬blɪk	*	*	*		

b. The acceptability of *bnick*

(favoring constraints) (disfavoring constraints)

bnɪk	IDENT-OO (*bnick*,*bliss*; $/b/_1$)	...	IDENT-OO (*bnick*,*brick*; $/r/_2$)	...
bnɪk			*	*
¬ bnɪk	*	*		

3.3.3 The Greenberg and Jenkins algorithm in OT.

Modeling the Greenberg and Jenkins algorithm in OT is a bit trickier, but still within the powers of OT devices. Among the innovations added to OT since its original incarnation is the notion of constraint conjunction, which creates new constraints with Boolean operators, in particular AND (∧) (Smolensky 1995, Crowhurst and Hewitt 1997, Balari, Marí n, and Vallverdú 2000 illustrate nonanalogical applications). Myers (2001) shows that to handle four-part proportional analogies like *drive* : *drove* :: *dive* : *dove*, the necessary constraint has the form given below, which is obeyed if and only if both component constraints are obeyed.

(3) IDENT-OO(*a*,*c*;F) ∧ IDENT-OO(*b*,*d*;G) [models the analogy *a* : *b* :: *c* : *d*]

To handle phonotactics using the Greenberg and Jenkins algorithm, conjoined constraints are again necessary, but this time ones whose component constraints refer to the same word pairs while varying the features being matched. For example, for *blick* (and using phonemes as "features") there would be constraints like the favoring ones listed below (out of a total set of $16 = 2^4$). If there is no appropriate real word (e.g. there is of course none that matches *blick* in all four of its phonemes), the constraint would be a disfavoring constraint.

(4) IDENT-OO(*blick*,*brood*;/b/)
IDENT-OO(*blick*,*slam*;/l/)
IDENT-OO(*blick*,*bluff*;/b/) ∧ IDENT-OO(*blick*,*bluff*;/l/)
IDENT-OO(*blick*,*brim*;/b/) ∧ IDENT-OO(*blick*,*brim*;/ɪ/)
IDENT-OO(*blick*,*bliss*;/b/) ∧ IDENT-OO(*blick*,*bliss*;/l/) ∧ IDENT-OO(*blick*,*bliss*;/ɪ/)
...

The resulting OT model can thus handle phonotactics. Because the combination of /b/ and /l/ occurs in many English words, while that of /b/ and /n/ does not, *blick* will have more favoring constraints and fewer disfavoring constraints than will *bnick*. Again, this means that under completely free constraint ranking, the acceptability of an item will be the proportion of its favoring constraints out of the total number of favoring and disfavoring constraints (constraints that are violated no matter what the representation of the target item, e.g. IDENT-OO(*blick*,*bliss*;/f/), have no effect under free ranking and can be ignored). This in turn makes quantitative predictions parallel to those tested by Ohala and Ohala (1986).

4. The Mandarin syllabary.

This section first summarizes the judgments on Mandarin syllables that were collected by Wang (1998), then turns to new results obtained using similar methods, as well as results from a pilot study which used on-line reaction-time measures to supplement the off-line judgment data. The discussion of the results from these studies will emphasize their relevance to the question of grammatical vs. analogical approaches to the Mandarin syllabary, focusing particularly on the arguments from lexicality and from naturalness vs. systematicity.

4.1 Judgment data from Wang (1998).

In Wang's (1998) study, native Mandarin speakers (college students in Taiwan) were played a tape recording with a set of 90 syllables, and they were asked to judge the distance of each from real Mandarin syllables (0 = closest to real, 10 = furthest from real). The syllables

were categorized into four sets ahead of time: real syllables; tonotactic accidental gaps (TAG syllables), which matched real words in every respect except tone, e.g. [xeɾ^{51}]; phonotactic accidental gaps (PAG syllables), which were strings of phonemes unattested regardless of tone, but were nevertheless considered by the experimenter as not forming part of a general pattern, e.g. [faɾ^{51}]; and systematic gaps (SG syllables). SG syllables included both "natural" cases like [mun^{51}], which violates the *Labial-Labial OCP constraint noted earlier, and "unnatural" cases like [fi^{213}], which violates a historically sensible but synchronically ad hoc constraint in Mandarin against the phoneme sequence /fi/ (this division into "natural" and "unnatural" was not attempted by Wang).

In order to carry out this study, then, some important decisions had to be made. First, as just observed, it was decided to distinguish between accidental and systematic gaps in the design of the materials (this discussion will be discussed further below). Second, the task was an off-line perception task: subjects heard the items, and then had some time to think about how to respond. This sort of task seems like the most obvious one to do, but it poses risks. For example, it is difficult for Taiwan Mandarin speakers to pronounce the syllable [si] (e.g. the letter "C" is typically pronounced [ɕi]); hence the "illegal" syllable /si/ will automatically become nativized before it even reaches the subjects' ears. Moreover, it is possible that acceptability has different sorts of effects on perception vs. production vs. memory. A third decision concerned the representation implicitly assumed in the selection of materials: nonword items were generated by combining Zhuyin Fuhao symbols (**注音符號**) in ways that obey general Mandarin syllable structure (e.g. onset symbols were only used in onset position, etc). This not only meant that the domain of nonwords was predefined as finite (e.g. strings like [ŋeup] were not even considered), but also that the experiment couldn't test violations of the Zhuyin Fuhao pronunciation rules (e.g. pronouncing ㄊㄧㄢ as [tʰian] rather than [tʰiɛn]). The question of representation is of course also crucial if the results are to be modeled with analogy.

Turning now to Wang's (1998) results, the overall findings were unsurprising: real words were given the best acceptability ratings (mean = 1.62), followed by TAG (4.66), then PAG (6.16), and finally SG (7.19); these differences were statistically significant. This general ranking is consistent with both a grammatical and an analogical approach (though as noted earlier some account must be given for why TAG syllables are considered better than PAG syllables, since neither violate grammatical constraints). However, the analogical approach does seem to account better for the lexicality effect found in the data: speakers had far sharper judgments about the distinction between real vs. nonreal syllables than they did about any subtype of nonreal syllables (the average score for real syllables was almost three times further from the next closest score than any score for a nonword category was from its next closest score). As discussed earlier, this result doesn't follow automatically from a grammatical approach to phonology, though one might argue that these results are due to performance factors unique to an experiment that implicitly asks subjects about lexicality as well as acceptability.

Regarding the contrast between the *Labial-Labial OCP gap and the */fi/ gap, a grammatical approach using naturalness does seem to make a correct prediction; further analysis of Wang's data shows that syllables violating the *Labial-Labial OCP constraint (mean score of 7.76, based on five items) were judged worse than those that violated the ad hoc */fi/ constraint (mean score of 5.73, based on two items). This difference could also be due to systematicity alone, though, since the *Labial-Labial constraint accounts for a larger

set of unattested syllables (e.g. those with four different onsets, not merely /f/). A better test is to compare the *Labial-Labial violations with those that violate a more systematic ad hoc gap in Mandarin. Just as /ia/ is restricted in Mandarin for whatever historical reason, so is /ua/, which tends to dislike not only Labial onsets, but Coronal onsets as well (e.g. *[tua], *[nua]), for no obvious natural reason. The mean score (based on four items) that Wang found for violations of the ad hoc *Coronal-/ua/ constraint was 7.79, virtually identical to that for violations of the natural *Labial-Labial constraint.

One final observation should be made about Wang's results: not all real words are created equal. That is, even though subjects were explicitly asked to compare test items with real words, judgments about real words were not always particularly close to the "0" end of the scale. For example, the real syllable [t^hiɛn^{213}] (e.g. 忝) received a score of 0.84, while the real syllable [mou^{213}] (e.g. 某) received 3.14. There are several possible explanations, all of which are more consistent with an analogical approach. Thus one could say that [t^hiɛn^{213}] is similar to words that are quite common in speech (e.g. 天), while [mou^{213}] may not be, or that [mou^{213}] itself represents but one morpheme, whereas [t^hiɛn^{213}] represents at least three (i.e. 忝, 殄, 舔). Another type of explanation would point out that since [mou^{213}] contains a labial onset and round vowels, it is dangerously close to falling into the *Labial-Labial gap. This isn't an argument for the grammatical approach, since [mou^{213}] does not in fact violate the *Labial-Labial constraint, and a miss is as good as a mile when one is dealing with grammatical constraints; by contrast, the fuzzy ad hoc constraints that emerge from analogizing from an existing lexicon are expected to be sensitive to just such near-misses.

4.2 New judgment data.

As valuable as Wang's (1998) data are, there are reasons for wanting to collect more. First, it doesn't seem appropriate to explicitly ask subjects to judge the distance of items from real Mandarin syllables, since this seems to bias the study in favor of an analogical approach from the start rather than test it objectively. Secondly, Wang's use of prespecified categories (i.e. TAG, PAG, SG) raises concerns. Not only did some of these categories prove to be quite misleading (for example in the categorization of syllables like [t^hiou] as PAG, since they were in fact judged to be as unacceptable as most of the SG), but there are also conceptual objections to be made against prespecifying "possible" vs. "impossible" words. If one avoids referring to naturalness, the only justifiable way to distinguish "possible" from "impossible" words is to see what generalizations a specific analogical model makes from the lexicon; otherwise one is just working in circles (i.e. this pattern is not found in real words, therefore it is impossible, therefore there is a grammatical constraint against it, therefore this constraint blocks the pattern from appearing in real words). Since one of the goals of this study was to test real judgment data against various analogical models, it was decided to leave distinctions among nonwords undefined, and let subjects simply give their judgments.

Thus rather than using preset categories, two hundred syllables were selected at random from a space including both real and "logically possible" syllables. This was done by using the chart of Mandarin syllables given in Tung (1972) (in which the vertical axis represents onsets, and the horizontal axis represents rimes according to an analysis consistent with Zhuyin Fuhao symbols). For each of the four tones, 50 syllables were randomly selected from this chart. Syllables judged to be impossible to pronounce unambiguously (e.g. if the selection procedure gave /tɕa/ rather than /tɕia/) were replaced with other randomly selected syllables. The result was a set of 200 syllables, 81 of which were judged to be real. All

items were written in Zhuyin Fuhao for a female speaker to record. Following the methods used by Wang (1998), a list of randomly ordered items was created (list A), and a second list of the same items in reverse order (list B). This was done due to technical limitations; since items could not be presented in a different random order for each subject, the two lists, one the reverse of the other, controlled for any order effects. As will shortly be seen, having two lists proved useful for other reasons as well.

Each list was played to a separate group of Mandarin speakers (female high school students in Tainan, aged 16-17, whose most commonly used language is Mandarin rather than Taiwan Southern Min). Group A, who heard list A, had 35 subjects, while Group B, who heard list B, had 33 subjects. Subjects were asked to decide whether the syllables 「**聽起來像不像國語**」 (sound like Chinese); they were given an answer sheet where they were to circle "5" to represent 「**最像**」(most like), "1" to represent 「**最不像**」(most unlike), or any intervening number. A smaller scale than Wang (1998) was used since it is generally accepted that subjects tend not to be able to make reliable or valid judgments on a scale with greater than 7 items, and the order of the scale was reversed so that a higher value represented greater acceptability. The vague wording 「**像國語**」 was used to avoid implying that subjects should compare items with specific real words when making their judgments.

Turning now to the results, the most fundamental observation was that whatever psycholinguistic phenomenon was being measured, the measurements were reliable: there was a very large correlation between scores from groups A and B ($r = 0.902$). Unsurprisingly, there was more agreement across the groups on what were the best and worst syllables than on the syllables in between (this result appears to have no bearing on the grammar vs. analogy issue). As it happened, only four of the new items were identical (entirely or but for tone) with those in Wang (1998), but the new judgments seemed to correlate quite well with his, with one exception: the syllable [$k^{h}iɛ^{55}$], which is obviously illegal, was not given a particularly poor score in the new study (3.40, which is closer to the "most similar" end of the 5-point scale). Apparently this was due to purely acoustic factors: the recording of this syllable (and others containing velars and /i/) was unfortunately acoustically quite similar to [$t^{h}iɛ^{55}$], which of course is perfectly acceptable. An alternative explanation, namely that the subjects were basing their judgments on Southern Min (which does allow /ki/), doesn't seem correct, given that Wang (1998) did not encounter this phenomenon, in spite of having subjects with linguistic backgrounds similar to the subjects of the new study.

Although items were not placed into categories in the design of the experiment, they were divided up in accordance with Wang's (1998) criteria after the fact. This resulted in 81 real words, 29 TAG, 5 PAG, and 80 SG (five items were too difficult to categorizae and so were left out of the analysis: [$miou^{51}$], [$mɔŋ^{213}$], [$nɤ^{55}$], [$p^{h}ia^{55}$], and [$t^{h}ɤ^{55}$]). Note how rare PAG syllables are in this random sample. This highlights another pitfall in the study of phonotactics: the proportion of items of a given type used in an experiment. It has been found that such proportions can exert dramatic effects on subject strategies (see Rubin, Becker, and Freeman 1979 for an example in the study of morphological processing). Since PAG syllables are relatively rare in the space of "logically possible" syllables, one might legitimately worry that this may cause speakers to process them in atypical ways if their proportion is too high in an experiment, for example by finding them more natural-sounding than they would otherwise.

Mean scores for the other categories fell into a pattern quite similar to that found by Wang (1998): real syllables were judged best (4.52), followed by TAG (3.17), and then finally SG (2.60). Again, the gap between the real syllable score and the closest nonreal score was more than twice as large than that between the nonreal scores, emphasizing the importance of the lexicality factor. The mean score for PAG was 3.45, higher than for TAG, but this may be an artifact of the small size of the PAG set. The results also replicated Wang's (1998) finding that real syllables are not created equal. Two sets of judgments (A and B) were used to test the reliability of this observation by comparing the set of real syllables that received a score of 4 or higher on both. These 69 items were thus considered to be very good examples of real words; nevertheless, groups A and B tended to agree that some sounded better than others. The correlation between the two was r = 0.48, which is not only statistically significant, but rather high as correlations go. As with Wang's data, however, the variation does not seem to be related in general to lexical frequency. Using the syllable frequencies given in Ho (1976), the correlation between frequency and real syllable scores was only r = 0.25, which, though statistically significant, is not very large (more than 93% of the variation must be caused by something else). This phenomenon will be mentioned again in the discussion of the modeling results in section 5.

Finally, interesting results were found in the comparison between gaps that appear to be both systematic and phonetically natural, and those that are merely systematic. Through the luck of the draw, most of the "systematic gaps" in this experiment involved front vowels appearing after onsets that do not allow front vowels. Unfortunately, as noted earlier, the subjects tended to find such forms surprisingly acceptable, most likely because of acoustic properties of the stimuli. However, there were also seven items that violated the *Labial-Labial constraint, which are not as prone to acoustic problems; they had a mean score of 2.95. By comparison, the mean score for the 20 items categorized as systematic but not phonetically motivated gaps (such as [tia^{213}]) was 2.21. This difference is not significant, though it gets close ($t(25)=1.84$, $p=0.08$); in any event, the difference goes the wrong way from that predicted by a grammatical approach (i.e. violating the "ad hoc" constraint is worse). Thus this comparison reconfirms the observation made earlier about Wang's results, namely that there is no evidence that phonetically natural constraints are more psychologically real than ad hoc language-specific gaps. Just for the record, the mean score for the 51 phonetically motivated gaps of the /ki/ type was also very close to these other two, namely 2.71; the larger sample sizes in this case now allow this difference to be statistically significant from the "ad hoc" set ($t(70)=2.35$, $p=0.022$). Again the difference goes the wrong way from the perspective of a grammatical approach, but due to the acoustic problems with these items, this result is harder to interpret.

4.3 Reaction-time measures of acceptability.

Since the above judgment data were meant to represent competence, that is, native speaker knowledge, further sources of performance data were sought. Thus the same set of 200 syllables was presented to a new group of subjects in a completely different way, who were asked to perform a completely different task. Specifically, the syllables were presented visually (in Zhuyin Fuhao) and subjects were asked to pronounce them. The crucial measure was reaction time, on the assumption that the more unacceptable a syllable was, the slower subjects would be to respond.

Twenty-three native speakers of Mandarin (undergraduates at National Chung Cheng University) were divided into two groups of approximately the same number of subjects (11

and 12), and each subject was presented with a randomly ordered list containing half of the 200-syllable set, written in Zhuyin Fuhao on a computer screen. A voice key (a microphone hooked up in such a way that a signal like a button-press is sent to a computer the instant any sound is registered) was used to record reaction time (i.e. from the onset of the visual stimulus to the onset of speech). It was expected that this task would be difficult (that's why only half of the 200-syllable list was presented to each subject), but it proved to be even more difficult than anticipated: on the average, a syllable was pronounced by fewer than 9 subjects (out of the 11 or 12 total subjects) within the window given by the experimental control program (three seconds), and the average response time for all syllables was a very long 1.2 seconds (and this was the time needed just to initiate speech, not to produce the whole syllable).

Given these difficulties, it is perhaps not surprising that the results were not fully satisfying. While there was a correlation between the judgment scores from the off-line experiment and the reaction times that was significant and went in the right direction ($r=-0.36$, meaning longer reaction times for syllables with lower judgment scores), this is apparently due solely to the lexicality effect: no significant correlation was found between the two performance measures when only the nonlexical syllables are examined ($r=-0.08$). This failure could have any number of causes. For example, for technical reasons, the Zhuyin Fuhao was presented horizontally, not vertically as this system is always used, and this may have confused subjects enough to hide any effects. Another problem was that the voice key was more sensitive to syllables beginning with noisy consonants (e.g. aspirated plosives), the result being that syllables with such onsets tended to have shorter reaction times, regardless of their acceptability. Whatever the cause of these null results, there is no point in discussing this experiment further, other than to note that it further underlines the methodological difficulties in testing hypotheses about linguistic competence in the real world.

5. Modeling the judgments.

This section describes the computer implementations of the two analogical models discussed earlier (a simple nearest-neighbor model and a model based on the Greenberg and Jenkins algorithm), as applied to the Wang (1998) Mandarin syllabary judgement data. Both models were implemented in QBasic programs running in DOS.

5.1 A simple nearest-neighbor model.

For every target item, this model searches for a lexical item that matches it in the greatest number of features. The model then takes this number of matching features as its "judgment" of the acceptability of the target item. This is roughly the model described above in section 3.1 and mimicked by the OT formalism in 3.3.2. The results described here used representations encoded with the following 23 features. The set of real syllables that the model used to form analogies was taken from Ho (1976).

(5)
1. Onset: 0, 1 (i.e. absent or present)
2. Place: L(ab), C(or), D(or), R(etro), P(al)
3. Aspiration: +, -
4-5. [cont]: ++ (fric.), -- (stop), -+ (affr.)
6. Sonorant: +, -
7. Nasal: +, -
8. Lateral: +, -
13. High: +, -
14. Front: +, -
15. Round: +, -
16. Tense: +, -
17. Final: 0, 1 (i.e. absent or present)
18. Front: +, -
19. Round: +, -

9. Medial: 0, 1 (i.e. absent or present)
10. Front: +, -
11. Round: +, -
12. Low: +, -
20. Place: C, D, R, V(owel)
21. Nasal: +, -
22-23. Tone: HH, LH, LL, HL

The results are illustrated in Figures 1 and 2 (see Appendix). Figure 1 is a scatterplot showing the correlation between the scores output by the model and the real judgment scores from Wang (1998). The correlation is high (r =-0.78) and goes in the correct direction (negative, since the model produces higher numbers for better syllables, opposite to Wang's scoring system). However, as can be seen, a lot of this effect results solely from the model's ability to distinguish real syllables from nonlexical syllables, which is not the most interesting issue in the study of analogy. Moreover, a large amount of variation among the judgment scores is left completely unaccounted for, since the model only recognizes three categories: matches of 21, 22 or 23 features.

Figure 2 shows the model's mean judgment values for Wang's four categories. The results are roughly consistent with Wang's (again noting that the model's and Wang's scoring systems run in opposite directions), with Real syllables judged the best, SG syllables judged the worst, and TAG and PAG in between. However, the only significant difference is that between the Real category and the others (e.g. Real vs. TAG: $t(22)=22$, $p<0.001$).

5.2 The Greenberg and Jenkins model.

The key step of the Greenberg and Jenkins algorithm is to compare each target item with all lexical items in all subsets of features. For example, for target item *blick* and lexical item *bliss*, and using phonemes as "features", the program would compare {/b.../. /.l../. /..ɪ./, /...s/, /bl../, /b.ɪ./, /b..s/, /.lɪ./, /.l.s/, /..ɪs., /blɪ./, /bl.s/, /b.ɪs/, /.lɪs/, /blɪs/, ∅} (the null subset means that a target item gets a vacuous "match" when no features match at all). Whenever the target and lexical item match in all features in a given subset, the target item earns a point; the sum of points for a target item is the model's prediction of its acceptability. This total can be quite high, since it may represent the sum of several matches with a given lexical item, plus the sum of several matches with another lexical item, and so forth.

As noted earlier, the number of sets of comparisons that must be made for each target/lexical pair is 2^n, where n is the number of features; thus using the same 23 features as the last model was deemed impractical (2^{23} = 8,388,608). Instead only five features were used, representing four syllabic slots plus tone. Sample representations are shown below ("I" = absent initial, "M" = absent medial, "V" = unspecified nuclear vowel, "F" = absent final).

(6)	mMaF3	[mɑ^{213}]	mMai2	[mai^{35}]	qiau4	[tɕ^{h}iau^{51}]
	liVn2	[lin^{35}]	qiEF1	[tɕ^{h}iɛ^{55}]	ZMVF4	[tʂɪ^{51}]

The results of the modeling were quite different from those found with the simple nearest-neighbor model, as shown in Figures 3-6. Figure 3 shows the relation between the model's predicted values and the actual values from Wang (1998). The correlation (r=-0.23) goes in the correct direction (the model's scale is again opposite to that of Wang) and is statistically significant ($p<0.05$), but it's not very large; approximately 95% of the variation is still unaccounted for. Moreover, it's not clear from this fact alone whether the model is

sensitive to more than just lexicality. Figure 4 shows the relation between the model and Wang's values just for the real syllables. Now the correlation is larger (r=-0.43), though in the smaller sample it's still only marginally significant ($p<0.05$); the model is only capturing approximately 18% of the real variation. Nevertheless, this result shows that the model is sensitive to more than just the difference between real vs. nonlexical syllables. In addition, as noted earlier, the systematic variation in native speaker judgments of real syllables was captured much worse by lexical frequency, so it may well be that an analogical approach to this mystery is a better approach to take. Figure 5, however, shows that far more work must be done: there is no relation at all between the model's and Wang's values for the nonlexical syllables (r=-0.01). Since the primary purpose of the analogical model is to account for judgments of novel forms, this obviously comes as a rather serious disappointment. Figure 6 shows where the problem lies: the model's judgments of the SG category is far too high. While the trend for the categories Real, TAG, and PAG are consistent with Wang (albeit these differences are not significantly different), the values for SG are not significantly different from those for Real syllables ($t(38)=0.699$, $p=0.48$).

6. Concluding remarks.

In spite of continuing modeling difficulties, there are at least three reasons why analogy should interest generative phonologists. First, analogical phenomena do seem to be psychologically real, even in synchronic phonology. If the job of phonologists is to describe speakers' knowledge of the sound system of their language, then such phenomena cry out for attention. Second, as this paper has tried to show, for the first time in the history of generative linguistics a point of contact can be made between two traditionally warring camps: Optimality Theory finally makes it possible for generativists to formalize true exemplar-driven analogy (though of course whether they want to use this power is another question). Finally, thinking about analogy highlights the inherent logical distinction between systematicity and naturalness. One shouldn't simply accept their conflation as a matter of dogma; it is actually an empirical question. If one wishes to test this question, one needs a formal framework for systematicity that operates independently of naturalness considerations, and that is precisely what analogical models provide.

REFERENCES.

Balari, S., R. Marin, and T. Vallverdu. 2000. Implicational constraints, defaults and markedness. Universitat Autònoma de Barcelona ms.

Benua, L. 1997. Transderivational identity: phonological relations between words. Amherst, MA: University of Massachusetts dissertation.

Bybee, J. L. 2000. The phonology of the lexicon: evidence from lexical diffusion. In M. Barlow and S. Kemmer (eds.) Usage-based models of language, 65-85. CSLI.

Bybee, J. L., and D. I. Slobin. 1982. Rules and schemas in the development and use of the English past tense. Language 58.265-289.

Chomsky, N. and M. Halle. 1968. The sound pattern of English. Harper and Row.

Chomsky, N. 1975. Reflections on language. Pantheon.

Clements, G. N. and E. V. Hume. 1995. The internal organization of speech sounds. In J. A. Goldsmith (ed.) The handbook of phonological theory, 245-306. Blackwell.

Crowhurst, M. and M. Hewitt. 1997. Boolean operations and constraint interactions in Optimality Theory. UNC at Chapel Hill and Brandeis University ms (ROA).

Daelemans, W. 1998. Toward an exemplar-based computational model for cognitive grammar. In J. van der Auwera, F. Durieux, and L. Lejeune (eds.) English as a human language: to

honour Louis Goossens, 73-82. LINCOM Europa.
Davis, S. 1991. Coronals and phonotactics of nonadjacent consonants in English. In C. Paradis and J-F. Prunet (eds.) The special status of coronals, 49-60. Academic Press.
Dell, G. S., C. Juliano, and A.Govindjee1993.Structure and content in language production: a theory of frame constraints in phonological speech errors. Cognitive Science17.149-95.
Di Sciullo, A. M. and E. Williams. 1987. On defining the word. Cambridge, MA: MIT Press.
Duanmu, S. 1990. A formal study of syllable, tone, stress and domain in Chinese languages. Cambridge, MA: MIT dissertation.
Elman, J. L. 1992. Grammatical structure and distributed representation. In S. Davis (ed.) Connectionism: theory and practice, 138-178. Oxford University Press.
Greenberg, J. H., and J. J. Jenkins. 1964. Studies in the psychological correlates of the sound system of American English. Word 20.157-177.
Gregg, R. J. 1973. The diphthongs ai and ae in Scottish, Scotch-Irish and Canadian English. Canadian Journal of Linguistics 18.136-145.
Hammond, M. 1995. There is no lexicon! University of Arizona ms (ROA).
Hayes, B. 1999. On the richness of paradigms, and the insufficiency of underlying representations in accounting for them. Presented at Senshu University, Japan, July.
Ho, K.-C. 1976. A study of the relative frequency distribution of syllabic components in Mandarin Chinese. Journal of the Institute of Chinese Studies 8.275-352.
Jun, J. 1995. Perceptual and articulatory factors in place assimilation: an Optimality Theoretic approach. Los Angeles, CA: UCLA dissertation.
Kenstowicz, M. 1995. Cyclic vs. non-cyclic constraint evaluation. Phonology 12.397-436.
Kenstowicz, M. and C. Kisseberth. 1979. Generative phonology: description and theory. Academic Press.
Kiparsky, P. 1988. Phonological change. In F. Newmeyer (ed.) Cambridge survey of linguistics Vol. 1. Cambridge University Press.
Kirchner, R. 1997. Contrastiveness and faithfulness. Phonology 14.83-111.
Labov, W. 1981. Resolving the Neogrammarian controversy. Language 57.267-308.
McCarthy, J. and A. Prince. 1993. Prosodic morphology I: constraint interaction and satisfaction. University of Massachusetts at Amherst and Rutgers University ms.
McCarthy, J. J. and A. Prince. 1995a. Faithfulness and reduplicative identity. University of Massachusetts Occasional Papers in Linguistics 18, 249-384.
McCarthy, J. J. and A. Prince. 1995b. The emergence of the unmarked: optimality in prosodic morphology. U. of Massachusetts at Amherst and Rutgers University ms.
Myers, J. 1993. A processing model of phonological rule application. Tucson, AZ: University of Arizona dissertation.
Myers, J. 1995. Nonlocal dissimilation in Mandarin syllables. Presented at ICCL4/NACCL7, University of Wisconsin at Madison, June.
Myers, J. 1997. Canadian Raising and the representation of gradient timing relations. Studies in the Linguistic Sciences 27.169-184.
Myers, J. 2000. Analogy and Optimality. National Chung Cheng University ms.
Myers, J. 2001. Exemplar-driven analogy in Optimality Theory. To appear in R. Skousen (ed.) John Benjamins.
Ohala, J. J. 1974. Experimental historical phonology. In J. M. Anderson and C. Jones (eds) Historical Linguistics II, 353-387. Amsterdam: North-Holland.
Ohala, J. J. and M. Ohala. 1986. Testing hypotheses regarding the psychological manifestation of morpheme structure constraints. In J. J. Ohala and J. J. Jaeger (eds) Experimental phonology, 239-252. Academic Press.
Pinker, S., and A. Prince. 1992. Regular and irregular morphology and the psychological status of rules of grammar. BLS 17.230-251.

Prince, A. and P. Smolensky. 1997. Optimality: from neural networks to universal grammar. Science 275.1604-1610.
Robinson, O. W. 1977. Rule reordering and lexical diffusion. In Wang 1977, 69-85.
Rubin, G. S., C. A. Becker, and R. H. Freeman. 1979. Morphological structure and its effect on visual word recognition. J. of Verbal Learning and Verbal Behavior 18.757-67.
Shen, Z. 1990. Lexical diffusion: a population perspective and a numerical model. Journal of Chinese Linguistics 18.159-201.
Silverman, D. 1996. Voiceless nasals in auditory phonology. BLS 22.364-74.
Smolensky, P. 1995. On the internal structure of the constraint component *Con* of UG. Paper presented at UCLA.
Steriade, D. 2000. Paradigm uniformity and the phonetics-phonology boundary. In M. B. Broe and J. B. Pierrehumbert (eds.) Papers in laboratory phonology V: acquisition and the lexicon, 313-334. Cambridge University Press.
Taft, M. and H-C. Chen. 1992. Judging homophony in Chinese: the influence of tones. In H-C. Chen and O. J. L. Tzeng (eds.) Language processing in Chinese, 151-172. Amsterdam: North-Holland.
Tung, T-H. 1972. Hanyu yinyunxue. [Chinese phonology.] Taipei: Student Books.
Vitevich, M. S., P. A. Luce, D. B. Pisoni, and E. T. Auer. 1999. Phonotactics, neighborhood activation, and lexical access for spoken words. Brain and Language 68.306-311.
Wang, S. H. 1998. An experimental study on the phonotactic constraints of Mandarin Chinese. In B. K. T'sou (ed.) Studia Linguistica Serica, 259-268. Language Information Sciences Research Center, City University of Hong Kong.
Wang, W. S-Y. 1969. Competing changes as cause of residue. Language 45.9-25.
Wang, W. S-Y. (ed) 1977. The lexicon in phonological change. The Hague: Mouton.
Ye, Y. and C. M. Connine. 1999. Processing spoken Chinese: the role of tone information. Language and Cognitive Processes 14.609-630.

APPENDIX.

Figure 1.

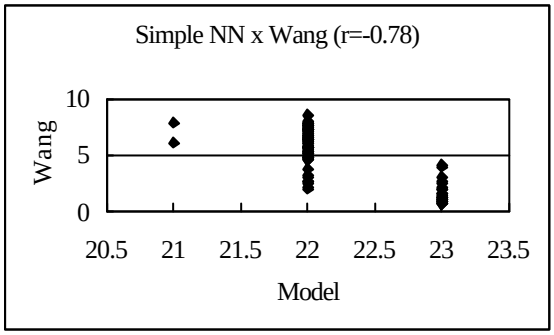

Figure 2.

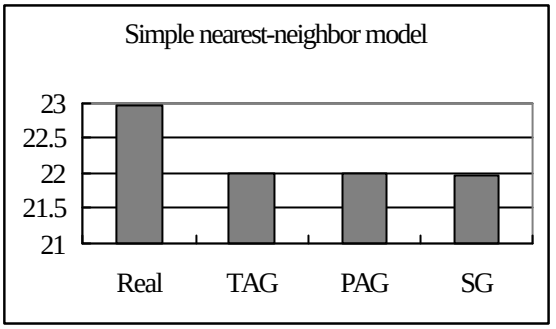

Figure 3.

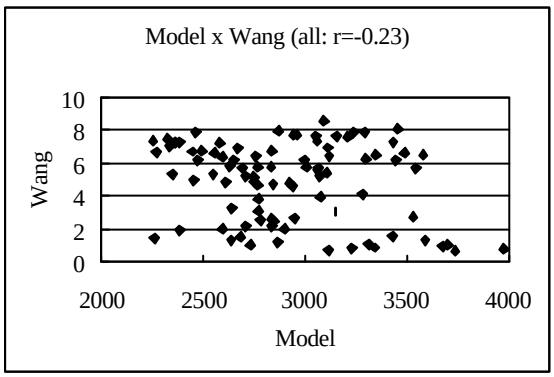

Figure 4.

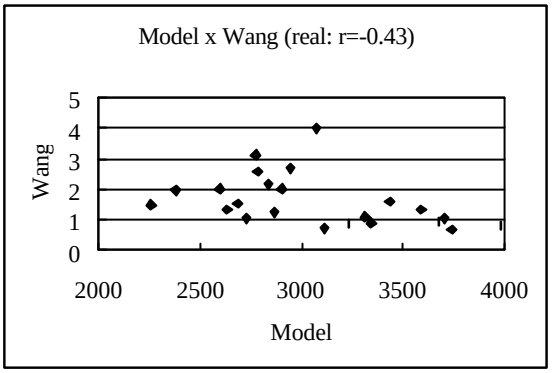

Figure 5.

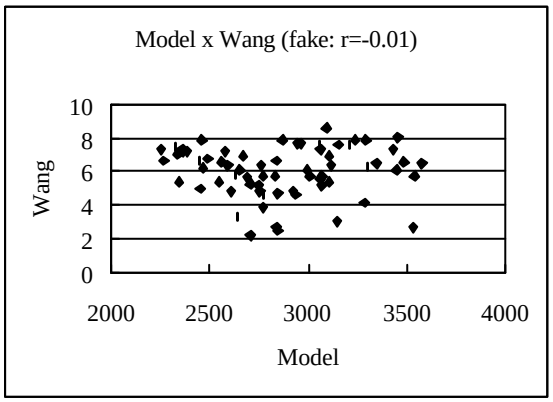

Figure 6.