

Linguistics as cognitive psychology*

James Myers

National Chung Cheng University

IACL-14/IsCLL-10 (May 2006)

The relation of linguistics to psychology is [...] implied in the basic position of the latter among the mental sciences. [...] As language is in its forms the least deliberate of human activities, the one in which rationalizing explanations are most grossly out of place, linguistics is, of all the mental sciences, most in need of guidance at every step by the best psychologic insight available.

Leonard Bloomfield (1914:322-3)

In the division of scientific labor, the linguist deals only with the speech signal [...]; he is not competent to deal with problems of physiology or psychology. [...] The findings of the linguist [...] will be all the more valuable for the psychologist if they are not distorted by any prepossessions about psychology.

Leonard Bloomfield (1933:32)

1. Introduction

Nothing, perhaps, symbolizes the fundamental dilemma of linguistics better than the paradox of the two Bloomfields. At the time he wrote his first book on language, he was a devoted follower of German psychologist Wilhelm Wundt, who pioneered the use of introspection as a source of information about the inner workings of the mind. By the time of his second book, Bloomfield had seen the introspective approach demolished in the rise of behaviorist psychology. The two Bloomfields still define the range of attitudes that linguists have about the role of psychology. On one side are the Old Bloomfieldians, who assume that psychology, as the logically more fundamental science, must provide the framework within which linguists do their work; functionalist or cognitive linguists might be put into this category. On the other side are the New Bloomfieldians, who insist that language must be studied for itself, regardless of what psychological evidence might show; formal linguists, including both Chomsky and the later Bloomfield, would go in this category.

However, neither of these two extreme positions captures the relationship between linguistics and psychology in the most productive way. In contrast to the attitude taken by the New Bloomfieldians, it's irrational to pretend that grammar and mind float around independently of each other, especially while proclaiming simultaneously that linguistics is a "cognitive science." But in contrast to the Old Bloomfieldians, it is equally irrational to hitch the fate of linguistics to whatever psychological notions one happens to find appealing (as Bloomfield himself learned from the downfall of introspectionism); moreover, language does have its unique characteristics which are missed by general-purpose psychology.

A more productive solution to the Bloomfield paradox, I argue, is to unite the content of linguistics with the structure of psychology. The content of linguistics consists of its facts, primarily corpus data (including dictionaries) and native-speaker judgments, and its hypotheses about mental grammar. The structure offered by (cognitive) psychology is what I will call linking models. In this conceptual framework, the world of observable facts is relevant to the world of hypotheses about the black box of the mind only within the context of a model that links them in an explicit and testable way. A linguistic methodology that

* Research for this paper was supported in part by National Science Council (Taiwan) grants NSC-93-2411-H-194-003 and NSC 94-2411-H-194-018.

incorporates linking models is not so much a reluctant compromise between the two Bloomfieldian views as a synthesis that maintains the key insights of both, building on "the best psychologic(al) insight available" (linking models) without being "distorted by any prepossessions about psychology": linguistic linking models link corpus data, judgment data, and grammar, all of which linguists know much more about than do psychologists. The notion of the linguistic linking model is hardly a radical proposal: the competence-performance distinction has been central to generative linguistics since Chomsky (1965), and so linguists are well aware that all linguistic observations are shaped not only by grammar but by extra-grammatical forces as well. Yet in actual practice, the competence-performance distinction is ignored, and linguistic data are treated as direct evidence for grammar. My suggestion is simply that linguists should see the competence-performance distinction as psychologists would: as a methodological framework for justifying abstract claims with concrete evidence.

I begin the discussion by showing the benefits that linking models have brought to psychology, which is demonstrably a more mature science than linguistics. I then show how linking models are intimately tied to three other characteristics that psychology shares with mature science: skepticism, quantitative methodology, and the never-ending search for better data sources. The remainder of the paper illustrates linguistic linking models in four case studies, all relating to the phonology of languages in Taiwan: the Southern Min tone circle, epenthesis and vowel harmony in the Formosan language Pazih, a comparison between the handshape inventories of Taiwan Sign Language and American Sign Language, and acceptability judgments of Mandarin syllables.

2. Psychology is a more mature science than linguistics

There is no denying that linguistics is a science, since even philosophers of science don't know what science is. Certainly linguistics tests rationally derived hypotheses with empirical observations. Yet linguists clearly spend more energy, in comparison with psychologists, on the rationalist side than on the empiricist side. As the pioneering cognitive psychologist George Miller (1990:321) puts it: "Linguists tend to accept simplifications as explanations [...] For a psychologist, on the other hand, an explanation is something phrased in terms of cause and effect [...]" This is because rationalists reason deductively, while empiricists reason inductively. This makes psychologists obsessed with methodology, a characteristic linguists tend to find a bit pathetic. When Chomsky (2002:102) notes that "the only field that has methodology courses, to my knowledge, is psychology," he doesn't mean this as a compliment. Similarly, when the linguists Phillips and Lasnik (2003:61) correctly note that collecting linguistic judgments is a kind of experimental procedure, they seem particularly proud of the fact that this procedure is "trivially simple".

However, just as an overly empiricist science wastes time on complex tests of trivial hypotheses, an overly rationalist science wastes time spinning fantasies out of bad data. Data problems particularly likely when working in the messy world of human behavior, as both psychologists and linguists do. Syntactic data is notoriously bad, as Schütze (1996) demonstrates, but phonology is not immune from the problem. Take Halle (1962), who argues for extrinsic rule ordering on the basis of the two "dialects" of Canadian English described by Joos (1942). However, by the time Chambers (1973) was conducting fieldwork, the school-age speakers of one of Joos's dialects had mysteriously disappeared, leading Kaye (1990) to ask whether this dialect ever really existed. Similarly, a whole stratum in Halle and Mohanan's (1985) model of English lexical phonology is devoted a rule of /l/ resyllabification which they exemplify as in (1) (after their figure (20), p. 65). Using instrumental phonetics, however, Sproat and Fujimura (1993) showed that /l/ resyllabification involves not a binary

palatalized/velarized /l/ contrast but a complex interplay between prosody and two distinct subgestures (lowering the tongue body and raising the tongue blade).

- (1) a. palatalized /l/: a whale edition the seal office
 b. velarized /l/: the whale and the shark the seal offered a doughnut

In a sense, the empirical challenges are even greater for phonologists than syntacticians, and not only because of phonetic misdescription. A single sentence may be capable of making or breaking a syntactic claim, at least in principle, but a phonologist needs to cite a patterned set of words, since any individual word could be a memorized lexical exception. A striking example of the problem of detecting phonological patterns is the confusion over tone spreading in Mende. Gussenhoven and Jacobs (2005:132) want to show that Mende tone consistently spreads left to right, and so they choose examples from Leben (1978) like (2ab), with surface [HLL] and [LHH] patterns. Zoll (2003:230-1) wants to show that the direction of tone spread in Mende depends on tone quality, and so she chooses examples from Leben (1978) like (3ab), with surface [HLL] and [LLH] patterns. Presumably all of the examples are factual; the unresolved question is which sorts of example are more *typical* of Mende.

- (2) a. fèlàrà "junction"
 b. ndàvùlá "sling"

- (3) a. fèlàrà "junction"
 b. lèlèrà "mantis"

Even worse, it's not enough that a phonological pattern be statistically significant; it must also be a "linguistically significant generalization." For example, Inkelas, Orgun, and Zoll (1997) point out that in Slave (Rice 1989), word-initial /l/ is always followed by /i/, as in (4). Is this a discovery about the mental grammar of Slave speakers? Inkelas et al. say no, since it turns out that all such words are borrowed French nouns, where /li/ is derived from the French article *le*; the pattern is merely a historical accident, and should not be considered theoretically relevant. This seems quite reasonable, but to apply this criterion consistently we would have to develop a theory of "historical accidents," which would force us to give up the generative version of New Bloomfieldianism, which says we must consider a synchronic grammar for itself: if Slave speakers know the /li/ pattern, they just know it, even if they don't know it's an accident. This counterargument also seems quite reasonable, and so we're stuck.

- (4) a. líbahdú "barge" (from French *le bateau*)
 b. lígarí "cards" (from French *les cartes*)

The lackadaisical attitude taken towards empirical methodology has led to a standard criticism of linguists, perhaps expressed most biting by Ohala (1986:4-5): "The history of disciplines that do not have an effective means of evaluating their claims often resembles a Brownian movement through the space of possible theories." Edelman and Christiansen (2003:60) make the same point when they complain that "the meter in the cab of the generative theory of grammar is running" (p. 60); even after all the detours it's taken, it still seems far from its promised destination of a cognitive science of language. Similarly, Pinker and Jackendoff (2005) charge that Minimalist syntactic theory requires that "most of the technical accomplishments of the preceding 25 years of research in the Chomskyan paradigm must be torn down, and proposals from long-abandoned 1950s-era formulations and from long-criticized 1970s-era rivals must be rehabilitated" (pp. 220-1). Similarly, Optimality

Theory (Prince and Smolensky 1993), which for some reason emerged at about the same time as Minimalism, was significantly worse at handling certain well-studied issues (opacity in particular). In the decade since then, both Minimalism and Optimality Theory have become increasingly ornate in the struggle to replicate the empirical success of their predecessors.

The "revolutionary" year of 1993 was simply one lurch among many in the zigzag history of generative linguistics. By comparison, progress in psycholinguistics has been both slower and steadier. This difference is demonstrated in Table 1 below, which sketches the histories of the two disciplines' dominant theories of the phonological forms of words (respectively, theories of word form knowledge and theories of word form production), spanning roughly 1970 to 2000. Major theoretical developments are indicated by brief descriptions of the claims distinguishing each theory from its predecessor, the primary data sources used to test these claims, and key references in the literature. The claims are numbered to make it easier to track how they are modified over time. As shorthand for the nature of the theoretical development, I use "+" to indicate that a change in a claim involved building on the corresponding claim in the previous theory (progress) and "-" to indicate that the change involved rejecting the corresponding claim in the previous theory (retrogression). Note that in the 1980s psycholinguists studying word production split into two competing camps (separated in the table by a dotted line): serial models vs. parallel models. As might be expected of a more empiricist science, this theoretical split corresponded with a methodological difference. Nevertheless, by the late 1990s the rift had been at least partially mended, with serial advocates admitting that some parallelism is necessary.

There are two key differences between the left and right sides of Table 1 that seem to be noncontroversial. First, the phonologists have never seriously explored data sources other than the traditional ones of dictionaries and native-speaker judgments (and the latter only rarely). The advice of Kenstowicz and Kisseberth (1979) and Ohala (1986) urging phonologists to explore new data sources has been almost entirely ignored. By contrast, the psycholinguists expanded their scope from corpora of naturally occurring speech errors to experimentally induced speech errors, then to other types of production experiments, and now they have begun to use neuroimaging. Psycholinguists interested in language production, both serialists and parallelists, have also taken the trouble to make their linking models as explicit as possible, by implementing them on computers.

This search for better data and better linking models seems natural from a psychologist's perspective, where the challenge of breaking into the black box of the mind is ever-present. By contrast, linguists have apparently stuck with the same data sources over the years because they persist in the repeatedly debunked illusion that traditional linguistic data (dictionaries and judgments) magically provide direct evidence for grammar. The competence-performance distinction, and the linking models that I argue should be developed from it, play virtually no role in actual linguistic practice.

A second difference between the two sides of Table 1 is the much greater probability that linguistic history will show retrogression rather than progress, a difference I believe is directly related to the linguists' neglect of methodological concerns. By contrast, psycholinguistic history is somewhat boring; it has never experienced anything as dramatic as the wholesale shift from a rule-based theory to a constraint-based theory, or the fickle treatment of phonemes (explicitly rejected, then given a starring role, then ignored). While Chomsky and Halle (1968) and McCarthy (1999) present essentially incompatible views, Levelt et al. (1999) is clearly just an advanced version of Fromkin (1971): a highly evolved descendent rather than an invader from another planet.

Compared with psychology, then, linguistics is a pre-paradigmatic science (Kuhn 1970), without a generally accepted framework for theorizing and data collection.

Table 1. Mainstream theories of word phonology in generative linguistics and of word production in psycholinguistics over the past three decades

Linguistics				Psycholinguistics			
Theory	Data sources	References	Theory	Data sources	References		
① Abstract underlying representations. ② Representations must be fully specified. ③ Rules are ordered extrinsically. ④ No phonemic level. ⑤ Morphology before phonology. ⑥ No syllables.	Dictionaries, supplemented occasionally by native speaker judgments	Chomsky and Halle (1968)	① Serial model. ② No mention of morphological frames. ③ Strings of segments divided by syllables.	Natural speech errors	Fromkin (1971)		
① +Restrictions on the abstractness of representations. ② –Representations may be underspecified. ③ +Some intrinsic rule ordering (Elsewhere Condition, lexical level before postlexical level) ④ –Phonemic level returns. ⑤ +Morphology and phonology interleaved. ⑥ –Syllables return.	Same	Kiparsky (1982)	① +More detailed serial model. ② +Morphological frames. ③ +More evidence for strings of segments divided by syllables. ----- ① –Parallel model. ② +Morphological frames. ③ +Strings of segments in syllabic frames.	Natural speech errors ----- Experimentally induced speech errors	Garrett (1980), Garrett (1988) ----- Dell (1989)		
① +Much more concrete underlying representations. ② –No underspecification. ③ –No rules, ordered or otherwise. ④ –Phonemic level ignored. ⑤ –Morphology and phonology in parallel.	Same	Prince and Smolensky (1993)	① +Main levels are serial, but parallelism within levels. ② +Morphological/metrical frames. ③ +Strings of segments mapped into frames from left to right. ④ Computer version of model developed.	Natural speech errors, production experiments	Levelt (1989), Levelt et al. (1999)		
③④⑤ –Competing attempts to handle facts that originally motivated rule ordering.	Same	McCarthy (1999), Kiparsky (2000), McCarthy (2006)	① +Partially overlapping serial stages associated with different brain areas. ④ +Computer model upgraded.	Production experiments, neuroimaging	Levelt & Inderfey (2000), Abdel Rahman et al. (2003)		

3. Linking models

The organizing framework used in psychology (at least in cognitive psychology) is, of course, the linking model. It is this notion that makes psychology a cognitive science. It's true that we can't directly observe mental entities; this incontrovertible fact is what led

behaviorists to deny the mind any scientific status at all. Nevertheless, if we can establish a chain of relationships linking hypothesized mental entities with observed behaviors, we have just as much reason to believe in those entities as we do any other natural phenomena that can only be observed indirectly, like subatomic particles or the sun's core. Importantly, however, this logic only works if the linking model is explicit enough to test; explicit theories about the abstract entities themselves aren't enough. Without a linking model, the behaviorists have every right to complain; a mind unlinked to behavior, or linked only in vague, commonsensical terms, is not something that science is capable of saying anything about.

Adopting linking models immediately leads to three other defining characteristics of psychology, characteristics it shares with other mature sciences. We have already seen one: the never-ending search for better data to test ever-more sophisticated linking models. The second characteristic, which underlies all of the others, is skepticism. A skeptic knows that appearances are deceptive (observed behavior is not the same as inner mental activity) and that wishful thinking can easily lead us astray (exciting hypotheses are not inherently more likely to be true than boring hypotheses). The motto of the skeptic is "Extraordinary claims require extraordinary evidence".¹ Suppose, for example, that someone claims that a quick glance at a page in a book creates a highly detailed visual copy in her mind, though the image immediately fades. Sounds crazy, and how could all this detail be proven if it fades too quickly to be fully described? Sperling (1960) found a way: flash complex visual displays to subjects, but only test them on random spots of the displays. No matter what spot he tested, his subjects gave accurate reports; therefore, the whole image must have been active in the mind before it disappeared. Through Sperling's methodological ingenuity, even the most hard-nosed skeptic was forced to accept this apparently bizarre claim about the mind's black box. Unfortunately, few linguists, whether functionalists nor formalists, are natural skeptics, most preferring instead the so-called "Galilean move towards discarding recalcitrant phenomena if you're achieving insights by doing so" (Chomsky 2002:102).²

The second characteristic of the psychologist's approach to science is the use of quantitative methods, including statistical analysis. Most linguists seem content to maintain the non-quantitative traditions that they inherited from the humanities, but traces of quantification can be detected even in current linguistic methodology: counting forms or languages to establish that such-and-such a rule or constraint is "productive" or "unmarked", or applying a sort of intuitive statistical analysis to select the most "typical" examples for a paper or presentation. Hopefully linguists will eventually overcome their math phobia, since quantification is essential if a linking model is to go beyond metaphor to working tool.

In its essence, a linking model is a statement about the correlation between observed data and a theoretical claim. Thus a schematic competence-performance linking model can be sketched as in (5). Conventionally, the element on the left side of the equation is called the dependent variable or dependent measure, and the elements on the right are called the independent variables, factors, or predictors.

(5) Observed performance = Competence + Known performance factors + Unknown

The picture in (5) needs to be complicated somewhat, since difference types of performance are influenced by different performance factors. In particular, corpus data are records of language production (output), and so are influenced by production processes (e.g. the conversion of pragmatic and semantic goals into phonological forms), whereas judgment data involve perception and comprehension processes (input), and so have different

¹ The idea goes back to David Hume, but the wording is often attributed to Carl Sagan.

² As Botha (1982) shows, it's unfair to blame Galileo for this misuse of his name.

influences (e.g. the parsability of sentences and the acoustic confusability of words). Thus the schema in (5) actually represents the family of schemas in (6).

- (6) Performance measure A = Competence + Performance factors A1, A2, ... + Unknown
 Performance measure B = Competence + Performance factors B1, B2, ... + Unknown
 ...

Importantly, if performance and competence are defined quantitatively, we can interpret every symbol in (6) literally, in their ordinary mathematical senses. For example, if the performance measure is an acceptability judgment on a gradient scale, we could code competence as a binary variable, where 1 = grammatical and -1 = ungrammatical, and code known extra-grammatical performance influences on judgments (e.g. acoustic confusability) in terms of some continuous variable. Then it makes both intuitive and mathematical sense that the judgment score by an individual speaker on an individual linguistic form really is 1 or -1 plus the performance variable plus other unknown variables, with each of the hypothesized predictors multiplied by constants (coefficients) representing their importance. If the performance measure is instead the probability that a linguistic form is actually attested in a corpus, the equation can be transformed to convert multiplicative probabilities into additive factors; this is what is done in VARBRUL, the Labovian sociolinguist's standard tool (Mendoza-Denton, Hay, and Jannedy, 2003).

At this point a skeptic should ask: What about that "unknown" component? Couldn't we manipulate it to give any result we want? And couldn't we also manipulate those coefficients to the same nefarious end? This is where statistics comes in. Applying statistics may be technically intimidating, but the purpose is very simple: to estimate, in a fully automatic and objective way, the sizes of the coefficients and of the "unknown" (random) component. Only if the unknown component is sufficiently small are we justified in claiming that the linking model as a whole really "explains" the dependent variable. Only if the coefficient associated with some specific factor is sufficiently large (far from zero) are we justified in claiming that this factor really plays a nonrandom role in the linking model. A statistical model is the ultimate skeptic: it represents the null hypothesis, the claim that there is no pattern at all. This is why psychology papers are littered with *p* values, which represent the probability that the null hypothesis is true; the lower it is, the more reasonable it is to think that something real might be going on (by convention, $p < .05$ is usually taken as statistically significant).

I begin the case studies in the next section with a very simple example of what happens when we stop to ask for the *p* value associated with an apparent linguistic pattern, rather than simply assuming that the pattern must be real because it seems interesting.

4. The Southern Min tone circle

As first noticed by Bodman (1955), Southern Min tone sandhi gives rise to the system of alternations among the lexical long tones shown in (7) (using the phonetic tone values found in Chiayi county in southern Taiwan, expressed in the 5-point IPA scale). Given its shape, this system is often called the Southern Min Tone Circle (though the Southern Min Tone Lollipop would be a more accurate name, given the "stick" where two tones neutralize into one).

- (7) [55] → [33] ← [24]
 ↑ ↓
 [51] ← [21]

A great deal of effort has been expended on accounting for this system in an elegant

generative grammar. Wang (1967) used a set of interlocking Greek-variable rules; Yip (1980) used rules where the input [33] and output [33] had different phonological representations; Tsay (1994) proposed an analysis that reversed the traditionally assumed direction. Only recently, after realizing there is really no way to choose among such analyses, have generative phonologists entertained more fundamental hypotheses. One skeptical alternative is that the inputs and outputs are not really phonetically identical as assumed in (7). However, after reviewing the phonetic literature (see also Myers and Tsay 2001), Moreton (2004) (first posted online in 1996) concluded that this assumption of (7) is probably correct.³

Moreton (2004) and Tsay and Myers (1996) then raised another skeptical alternative, namely that the pattern in (7) doesn't involve general rules, but merely allotone selection: speakers simply memorize arbitrary tone pairs ([55]~[33], [33]~[21], etc). This would mean that the tone circle is a meaningless coincidence. For Moreton (2004) this conclusion is particularly crucial, since he proves that Optimality Theory is mathematically incapable of handling circles like (7) if they are real.

To the best of my knowledge, however, nobody has ever completed the argument by showing that the pattern in (7) is indeed likely to be a coincidence. We can only do this within the constraints of a linking model. Without any constraints at all, the pattern in (7) is extremely unlikely to arise by chance alone (e.g. why isn't it [13] ↔ [42]?). The appropriate linking model would be one where the independent variables conform to the assumptions of the allotone selection hypothesis, listed in (8).

- (8) a. Southern Min has the five lexical long tones [55], [33], [24], [51], [21].
 b. Tone sandhi consists of pairs of these lexical tones.
 c. Tone sandhi is thus structure-preserving: no new tones are created.

When we now come to consider what the linking model's target outcome should be, it is important to note that it is *not* (7) itself. What makes (7) notable is the circular shape, not the specific value and order of the tones. Thus any five-tone system with a loop anywhere in it would attract our attention. Indeed, in the universe of loops formed of these five tones, the pattern in (7) is not the most noteworthy imaginable: more amazing would be a system consisting solely of a five-tone loop, without any "lollipop stick."

Suppose we take the position that in order to count as theoretically interesting, the loop must be at least as big as the one that is actually observed: either four tones, as in the actual tone circle in (7), or five. The probability of getting such a "big" loop by chance is the number of such looped systems divided by the total number of tone systems defined as in (8). To calculate this, we start by encoding a "big" loop in a 5-tone system as an ordered series ($t_1, t_2, t_3, t_4, t_5, t_6$) where either $t_6 = t_1$ (5-tone loop) or $t_6 = t_2$ (4-tone loop). In the first case, there are $5! (= 1 \times 2 \times 3 \times 4 \times 5)$ possible ordered series (t_1, t_2, t_3, t_4, t_5), but since it doesn't matter which tone we actually start the loop with (e.g. $(t_1, t_2, t_3, t_4, t_5, t_1) = (t_2, t_3, t_4, t_5, t_1, t_2)$), the number of 5-tone-loop systems is actually $5!/5 = 4! = 24$. The logic works the same way in the second case, except now we must treat the "lollipop stick" as special (e.g. $(t_1, t_2, t_3, t_4, t_5, t_2) \neq (t_2, t_3, t_4, t_5, t_1, t_3)$), so the number of 4-tone-loop systems is $5! = 120$. Put together, then, the total number of "big" loops is $144 (= 24 + 120)$.

As for the number of logically possible tone systems, we start with the observation that for each of the five tones, there are five logically possible outputs (including vacuous rules

³ Note, however, that proving phonetic identity involves *accepting* the null hypothesis, rather than *rejecting* it; this poses the same logical problems as trying to prove a negative. Indeed, some still-unpublished phonetic experiments I conducted with Jane Tsay did show miniscule but significant differences in [33] as output of [55] vs. [24]. Saving (7) would then require showing that these differences are extra-grammatical (e.g. side-effects of our reading-aloud task), but the necessary experimental work to confirm this has yet to be done.

where input and output are identical). A tone system will then be a set of five tone rules, each starting with a different input and each ending with one of five outputs. If each rule is independent of the others (permitting neutralization, since the real system permits it), the total number of possible tone systems, assuming an ordered set of lexical tones, is (5 possible rules for t_1) $\times \dots \times$ (5 possible rules for t_5) = $5^5 = 3125$.

This means that the probability of getting a "big" loop by chance in a system conforming to (8) is $p = 144/3125 = 0.04508$, just marginally significant ($p < .05$). If we make the additional assumption that vacuous rules are not permitted (this is at least somewhat plausible on functional grounds, given the potential usefulness of tone sandhi to adult listeners as a marker of syntactic constituent boundaries; see Tsay, Myers, and Chen 2000), then the number of logically possible tone systems drops to $4^5 = 1024$, so $p = 0.140625 (= 144/1024)$, which is not statistically significant ($p > .05$). If we follow Moreton (2004) and assume that a loop of any size is theoretically problematic, and thus interesting, we must consider the set of all looped systems, including systems containing multiple small loops. However big this set is (I haven't bothered to calculate it), it must be a proper superset of the set of 5-tone-loop systems, increasing the size of the numerator and thus the p value.

Note that this exercise is psychology, not merely mathematics. The allotone selection hypothesis is a hypothesis about mental grammar (or the lack of it). The hypothesis makes a specific prediction: the Southern Min Tone Circle is a coincidence. We put the hypothesis into one end of the linking model (as in (8)) and see what comes out the other end. What comes out is a p value that is at best only marginally significant. Thus based on the data we have chosen to test, there is no reason to reject the allotone selection hypothesis. At the same time, the high p value gives us good reason to cast a highly skeptical eye on the earlier analyses that worked so hard to explain an apparently illusory pattern.

The reader should not conclude from this case study that linking models and statistics take all the fun out of linguistics by debunking its most interesting claims. On the contrary, it is precisely when theoretical claims survive a highly skeptical examination that they make the most dramatic impression. I illustrate this point in the next case study.

4. Epenthesis and vowel harmony in Pazih

The Formosan language Pazih (or Paze) has a set of morphemes consisting of reduplicated CVC syllables with an intervening medial vowel (all examples come from Li and Tsuchida 2001; see also Blust 1999). This vowel is usually identical to that of the reduplicated syllables, as in (9a). The problem is that there are a sizable number of exceptions in which the intervening vowels do not show vowel harmony, as in (9b).

- | | | | | | | |
|-----|----|-----------|----------------|----|-----------|--------|
| (9) | a. | bak-a-bak | "native cloth" | b. | bar-e-bar | "flag" |
| | | hur-u-hur | "steam, vapor" | | hur-a-hur | "bald" |

The existence of these exceptions presents a challenge even more basic than that posed by the Southern Min Tone Circle. Here it is not a matter of establishing whether a set of rules form a coherent system, but whether there really is any rule at all. As it happens, in the short but presumably complete list given by Li and Tsuchida (2001:20-21), almost one fourth of the total (12 out of 45) are exceptions. This seems like a lot, but is it enough to undermine any claim of a vowel harmony pattern?

To see the problem in linking-model terms, note that the data are corpus attestations, so the dependent measure is categorical: "patterned" (obeying vowel harmony) vs. "exceptional." The question is whether the probability of a /CVC-CVC/ word surfacing as patterned is significantly higher than would be expected by pure chance. According to the

simplest null hypothesis, each word has an equal probability of being patterned or exceptional. This means that we can calculate the p value in terms of the probability that our target event (being patterned) will happen at least as many times as observed out of a total number of independent trials (words). That is, we treat words as coin flips, where heads, say, represents the property of being patterned, and calculate the probability of getting 33 or more heads in 45 flips. More generally, given a certain number of patterned forms, what is the maximum number of additional exceptional forms we can tolerate before $p > .05$? Table 2 below gives the answer for various sizes of patterned sets. According to the table, with 33 patterned words we would need more than 18 exceptions before we would no longer be justified in rejecting the null hypothesis. But we only have 12 exceptions. This means that we may indeed reject the null hypothesis: vowel harmony is statistically significant.

Table 2. INSTRUCTIONS: Find the number of patterned items in the grey cells. The number in the adjacent white cell then shows the maximum number of additional exceptions that can be tolerated before the pattern becomes nonsignificant ($p > .05$).⁴

Patterned	Max exceptional						
1	*	11	2	21	9	31	16
2	*	12	3	22	9	32	17
3	*	13	4	23	10	33	18
4	*	14	4	24	11	34	18
5	*	15	5	25	12	35	19
6	0	16	5	26	12	36	20
7	0	17	6	27	13	37	21
8	1	18	7	28	14	38	21
9	1	19	7	29	15	39	22
10	2	20	8	30	15	40	23

* Too few patterned items to be significant, even if there are no exceptions

Unaware that the twelve exceptions are not fatal, Li and Tsuchida (2001) are eager to reduce their number by showing that some of them are not really exceptions. Thus they observe that among these twelve words, the vowel "/a/" appears to be the most common" (p. 21). This would make 7 patterned words (with /a/) and 5 exceptional words (with some other vowel). Unfortunately, from Table 2 we can see that 7 patterned words form too small a set to tolerate any exceptions at all. There is thus no reason to reject the null hypothesis that the vowels in the non-vowel-harmony words truly are random exceptions.

The above analyses assumed that the odds of getting a patterned (vowel harmony) surface form by chance is 50/50, but a skeptical observer should rightly ask whether we are justified in treating cases like *bak-a-bak* and *hur-u-hur* the same way, since they involve different vowels. This leads to a different null hypothesis, one where the intervening vowel could be any of the four phonemic vowels of Pazih (/i/, /u/, /e/, /a/). The chance probability that vowel harmony appears in any given word is thus 1/4, the chance probability that vowel harmony appears in any two words is 0.0625 (= (1/4)²), and so on: now the analogy is not the repeated flipping of a coin, but the repeated rolling of a four-sided die. The resulting p value for the 33 patterned and 12 exceptional words is a vanishingly small $p = 0.0000000003$: the

⁴ This table was generated via the binomial version of the sign test (also called the exact McNemar test). The p values were calculated using the Excel function = MIN(1, 2 * BINOMDIST(MIN(#patterns, #exceptions), total observations, 0.5, TRUE)), where 0.5 represents the assumed chance probability of an exceptional outcome.

simultaneous harmonization of all four vowels is very unlikely to be an accident.⁵

However, establishing statistical significance is only a necessary but not sufficient condition for vowel harmony being a productive part of the mental grammar of Pazih speakers: they might very well memorize these forms as wholes (after all, they are morphemes, not generated by morphological rules). Demonstrating productivity would require experiments with novel forms (e.g. asking which made-up word sounds better, *dux-u-dux* or *dux-a-dux*). Unfortunately, Pazih is nearly extinct; even its most fluent speakers are losing command of it as they age. This is hardly atypical for the sort of data sources phonologists rely on. Linguists don't have the luxury of simply dismissing all non-experimental data as virtually useless (as Ohala 1986 comes close to advocating). Therefore I think it's worthwhile to see how much we can infer about Pazih mental grammar from the corpus data that we actually have, impoverished as it is.

What does the Pazih pattern imply? First, regardless of whether or not it is synchronically productive, its systematicity must have come from somewhere. The theory of this systematicity must be a psychological theory, even if the true story is merely diachronic and the mental grammar of adults is not directly relevant. Second, our theory must be able to handle both the vowel harmony pattern and its exceptions. Third, and most intriguingly, the theory must also be able to handle the evidence suggesting that the intervening vowel is epenthetic. That is, implicit in the above discussion is the fact that /CVC-CVC/ morphemes always have the surface form [CVC-V-CVC]. There are no exceptions at all (45 heads vs. 0 tails): no *[CVC-CVC], no *[CVC-CVC-V], and so on. The epenthetic vowel thus also demands a psychological explanation.

Any self-respecting phonologist should have already thought of an explanation: the epenthetic vowel is there to split up the consonant cluster, or rather, to turn the preceding coda into an onset. The first benefit of this hypothesis is that it correctly predicts that Pazih generally avoids morpheme-internal consonant clusters, only permitting them if the first segment is a nasal as in (10a), or a glide as in (10b). As Li and Tsuchida (2001:20) note, "[i]t is clear that the empty vowel is added to a reduplicated form as a result of limiting the syllable structure in the language."

- (10) a. sampuy "early harvest"
 bintul "star"
 riŋxaw "rice gruel"
 b. tawtaw "peanut"
 saysay "anything"

The second benefit of the epenthetic hypothesis is that it helps explain why there is vowel harmony at all. By hypothesis, the epenthetic vowels are inserted solely to fulfill a surface prosodic constraint, so they need not be present underlyingly. If they are not present underlyingly, they have no underlying featural content. Since they need featural content on the surface, this content must be predictable (generated by regular rule). One way to achieve this is via vowel harmony.

But now those twelve exceptions come back into our story. As we have established, the featural content of the vowels in these twelve words is not predictable. Since speakers nevertheless know how to pronounce the exceptions, generative phonological theory obliges us to store the features in the input (underlying representation). This is so even though the vowels, as epenthetic, have no underlying prosodic structure for the features to link to. We

⁵ This was calculated by replacing 0.5 in the Excel function with 0.75, the chance probability of getting a non-harmonizing vowel.

are therefore forced to the conclusion (or so it seems) that the features of the exceptional epenthetic vowels must be underlyingly floating.

Look what we have discovered in the black box of the mind, a linguist might say at this point: floating features! Given the equally bizarre things that cognitive psychologists have established (like highly detailed, rapidly fading visual images), we should not reject this conclusion for its bizarreness alone, as critics of generative linguistics are often wont to do. Nevertheless, skepticism is still warranted, since skepticism is *always* warranted.

Let's carefully retrace the steps that led us to this conclusion, since it is these steps that make up the relevant linking model. On one end is the raw data, which consist of a list of transcriptions in Li and Tsuchida (2001). A skeptic might wonder whether the problematic exceptions were simply mistranscribed, erasing the problem entirely. However, it seems extremely unlikely that twelve independent mistranscriptions could have occurred by chance, and we do have reason for thinking that the fieldwork methodology was biased in this way.

A more serious concern at this initial step in our journey from data to theory relates to the key generalization made by Li and Tsuchida (and Blust 1999) about the restriction on codas, which plays an essential role in the epenthesis analysis. Evidence that something is wrong with this generalization comes from the near minimal pair in (11). If (11a) shows that glide-initial word-internal clusters do not violate any prosodic constraint, and if epenthesis is required solely for prosodic reasons, why is there epenthesis in *hay-a-hay*?

- (11) a. saysay "anything"
 b. hay-a-hay "stalk of miscanthus"

At least three very different possibilities suggest themselves. Perhaps *saysay* is the anomaly; as the only example of /y/ + obstruent cluster cited by Li and Tsuchida (Blust 1999 cites none at all), we are free to speculate that epenthesis is unnecessary here for some word-specific reason (*saysay* seems to be morphologically related to *say* "question marker," which would put a protective morpheme boundary between the /y/ and /s/). This first possibility would allow us to preserve the hypothesis that epenthesis is always prosodically induced.

Alternatively, the anomaly may instead be *hay-a-hay*, forcing us to admit that epenthesis may occur even without a phonological need. This would make epenthesis a quasi-morphological process, which in turn would mean that the non-harmonizing vowels in those twelve exceptions could be analyzed as lexically specified formatives, similar to those that arbitrarily appear in English *-ion* nominalizations (e.g. *delete-deletion*, *repeat-repetition*, *elicit-elicitation*).

A third possibility would be that neither form in (11) is anomalous. Lexical items can themselves be seen as performance data, indirectly reflecting the operation of a hidden grammar. Suppose this grammar is a stochastic grammar of the sort posited by Anttila (2002) for lexical variation in Finnish, outputting A some of the time and not-A the rest of the time, particularly in contexts where phonological constraints are not strong. This would cause instances of both A (epenthesis) and not-A (no epenthesis) to emerge in the lexicon. Consistent with Anttila's approach, the conflict in (11) involves syllables that are intermediate in illicit status: they do have codas, but "weak" codas. More problematic would be variation in epenthesis with obstruent codas or open syllables, but this is not found.

Unfortunately, all three alternative analyses seem equally plausible, and there appears to be no empirical way to choose among them based on the present data set.

One level beyond Li and Tsuchida's transcriptions and generalizations are the native-speaking consultants themselves. The possibility that all of the exceptional words were the result of momentary speech errors seems extremely unlikely on statistical grounds, but there is another possibility: the speakers not only had no knowledge of the vowel harmony pattern,

but they devoted more cognitive resources to the reduplicated /CVC/ portion of the morphemes. Thus occasionally they randomly substitute non-harmonic vowels. However, this possibility seems to be ruled out by the minimal pair *hur-u-hur* "steam, vapor" vs. *hur-a-hur* "bald" in (9), specifically highlighted by Li and Tsuchida, and in any case, we can reasonably assume these experienced fieldworkers carefully tested for lexical variability.

Whether or not the vowel harmony pattern is a productive part of adult grammar, a yet deeper level is its diachronic source. Here Li and Tsuchida have a specific proposal. Rather than analyzing the exceptions as involving floating vowel features, they suggest that the exceptions "may have come into being and fossilized at an early stage before the rule of adding an empty vowel applied" (p. 21). As support for the fossilization hypothesis, they emphasize their monomorphemic, non-derived status; morphemes must be stored in memory anyway, so why not store the exceptions as wholes?

At first sight, this seems like an appropriately skeptical approach to the data: diachronic fossilization of patterns is a well-attested phenomenon. The problem is that the exceptions are not pure exceptions. They may have arbitrary vowel features, but the presence of the epenthetic vowel is fully systematic and apparently prosodically motivated, as Li and Tsuchida note themselves. Li and Tsuchida's putative earlier stage, therefore, would have been populated by speakers who were obliged to epenthesize vowels but who nevertheless had total freedom about which vowel they chose for this purpose. Free variation in allomorphy is not impossible (cf. English [ɛ]*conomic* vs. [i:]*conomic*), but such cases seem to be non-systematic, targeting random parts of a few random morphemes. Here allomorphs would be freely generated for an entire lexical class (all /CVC-CVC/ forms).

This scenario also doesn't fit with what is known about the diachronic emergence of vowel epenthesis (see Blevins 2004:155-158). Vowel epenthesis often begins as the phonologization of consonant release, which explains why the resulting vowel quality is predictable (phonetic processes tend not to add much lexical information), why epenthesis is most common after obstruents (obstruents are the most likely to have release), and why word-final codas don't induce epenthesis (interconsonantal release is more audible, hence learnable). If Pazih followed this more typical diachronic path, at an early stage the epenthetic vowel would have been prosodically present but featurally ambiguous. Later on, speakers filled in the vowel features according to a stochastic process similar to that modeled statistically above: speakers aimed at vowel harmony, but occasionally missed the target for whatever reason (note how this can be seen as a diachronic linking-model variant on Anttila's stochastic grammar). Eventually both harmonic and non-harmonic vowels were treated as lexically specified components of the morphemes. It should be admitted, however, that the examples in (11) pose the same problems for this diachronic analysis as for the floating-feature analysis, since both assume that epenthesis is an across-the-board phenomenon.

Of course, for most of the speculations thrown about in this section, we will never have the necessary evidence to show which, if any, is on the right track. Thus it may seem that even after carefully considering a multi-level linking model, the most useful thing we can say is discouragingly negative: on the basis of this data set alone, there is no compelling reason to prefer the floating vowel feature analysis over any of the alternatives.

Nevertheless, at least two interesting claims remain constant across all of the analyses, suggesting that this exercise has indeed yielded some genuine insights into how phonology works in the mind. First, there must be a split between prosodic phonology and segmental phonology. This split follows from the fact that the prosodic pattern (epenthesis) is exceptionless while the segmental pattern (vowel harmony) is not, implying that they have inherently distinct properties. Both diachronic analyses presume this split: according to my proposal, the prosodic change occurred before the segmental change, while for Li and Tsuchida, prosodic phonology provides the motivation for epenthesis while segmental

phonology provides the implementation. The floating-feature analysis distinguishes motivation from implementation in a similar way. While the prosodic/segmental split may not seem dramatic, it hasn't always been recognized, by either phonologists or psycholinguists (as was shown in Table 1), and it is nice to get further independent evidence for it.

Second, the fact that the vowel harmony is total rather than partial shows that speakers are capable of treating segments as wholes, rather than merely as arbitrary collections of articulatory or perceptual features. Moreover, since our statistical analysis suggests that there is a single harmony process rather than separate ones for each vowel phoneme, speakers seem to treat harmony in terms of symbol manipulation, where a general V slot can have any specific vowel plugged into it. Again, these points have lost their drama through over-familiarity, but they represent rather deep insights into the nature of the human language faculty. Complete identity is unlikely to arise by phonetic processes alone, and as Reiss (2003) shows, formalizing an identity function when only features are phonetically "real" poses serious challenges. Among the approaches known to be incapable of handling the identity function are computational models of exemplar-driven analogy (see Marcus 2001 for extensive discussion). Whatever gave rise to Pazih vowel harmony, it wasn't pure analogy.

As promised, then, skeptical application of the linking model framework can bring clarity to issues of importance to theoretical linguistics, and in fact, sometimes gives us more confidence in certain abstract claims than in their "boring" alternatives. What we can conclusively conclude about the Pazih pattern may seem far less dramatic than what generative linguists are accustomed to, but if slowing the pace of research somewhat means that work actually progresses rather than going in circles, the sacrifice seems well worth it.

One final point needs to be made about this case study before moving on to the next. I claimed above that the Pazih corpus is highly impoverished, and therefore we'll never really know what's going on in Pazih grammar. However, this argument neglects a key dogma of generative linguistics: the poverty of the stimulus argument. While this argument is most familiar from the innateness debate (starting with Chomsky 1965, though the term was not invented until later), it also applies to the situation faced by the scientist, since all theories are underdetermined by evidence, as Chomsky, in his consistently rationalist way, has pointed out many times (e.g. Chomsky 1980). Thus it is not possible to look at one language in isolation, without any preconceived notions, and "let the data speak for themselves." Even when arguing for my preferred diachronic analysis of the Pazih data, an analysis too empiricist for some generative phonologists to find interesting, I assumed *a priori* that Pazih was a "normal" language, and therefore, based on what was known about *other* languages, Pazih probably went through certain historical stages. My logic was no different in kind from what a hard-core generativist would use when taking a constraint discovered in English and applying it to Chinese. Similarly, psycholinguists, despite the dismay many of them express for Chomskyan nativism, are in the business of discovering principles of the human language processor in general, even if they conduct no cross-linguistic research at all.

This leads us to the next case study, which explores the role of linking models in cross-linguistic research.

5. Handshape inventories in Taiwan Sign Language and American Sign Language

Linguists base their claims of naturalness (unmarkedness, innateness, etc) on three types of evidence: learnability (or the lack thereof), ease of processing, and cross-linguistic typology. The first two, if studied properly, require a methodological approach familiar to psychologists (well-designed experiments with multiple subjects, statistical analysis, perhaps computer modeling), which is why linguists tend to rely on the third. Typological research is merely a form of corpus analysis, but one where the corpus is a collection of languages rather

than a collection of forms within one language. Since phonologists traditionally work with corpus data already, it has been much easier for them to amass large cross-linguistic databases than it has been for syntacticians; *Aspects of the theory of syntax* (Chomsky 1965), despite its all-purpose title, cites only English data, whereas *The sound pattern of English* (Chomsky and Halle 1968) actually has a lot to say about a wide variety of languages.

When it comes to sign languages, however, for a long time one language has stood for all: American Sign Language (ASL). Only in the past decade or so has serious research begun on the many other historically unrelated sign languages of the world, including Taiwan Sign Language (TSL; e.g. Smith and Ting 1979, 1984). Even today, most of the cross-linguistic sign literature is concerned with syntax (e.g. Yang and Fischer 2002) or morphology (e.g. Aronoff et al. 2005). Still relatively rare is cross-linguistic comparisons of sign phonology (which, like spoken phonology, relates to semantically neutral form patterns, in this case the shape, location, orientation, and movement of hands, fingers, and nonmanual features like facial expressions). The relatively unexplored nature of sign language typology, not to mention the unfamiliarity of sign language research in general to most linguists, is a particularly attractive feature given my purposes in this paper: we can explore the logic of typological research without preconceptions about what should or shouldn't be considered "natural."

An oasis in the desert of typological research on sign phonology is Ann (1996; see also Ann 1993, 2005, 2006), who compares the handshapes of ASL and TSL. She supplements her typological data with the second source of evidence about naturalness listed above, namely an explicit theory of ease of articulation in terms of the three physiological facts listed in (12).

- (12) a. The thumb, index, and pinky each has its own independent extensor muscle while the middle and ring do not, so it's easier to extend the thumb, index, and pinky independently than to do so with the middle or ring.
- b. The middle, ring, and pinky together share a muscle and thus tend to configure together.
- c. Due to the combinations of muscles involved, finger curving is harder than extension, which is harder than bending, which is harder than closing.

These considerations allow Ann to quantify the ease of articulation for any particular hand configuration, independent of any linguistic consideration. Her key claim is that signers' frequency of use of handshapes should be correlated with the handshapes' ease scores. This proves to be true, for both TSL and ASL, which therefore end up showing the same quantitative distribution of handshapes in their lexicons.

As important as Ann's study is, however, it doesn't fit in the scope of this paper for two reasons. First, its focus is on the lexicon. To linguists interested in mental grammar, the lexicon is at best considered a necessary evil, since, in the memorable phrase of Di Sciullo and Williams (1987:3), it "is like a prison" of "the lawless"; it is not a research end in itself. The surface form of a lexical item is merely performance data, providing evidence for the hidden grammar that gave rise to it, but nothing more. Second, the linking model underlying Ann's analysis focuses on performance as well, postulating that the probability that a sign will be added or remain in the lexicon is determined, in part, by ease of articulation, presumably because harder handshapes are either avoided or modified over time into easier handshapes. Grammar doesn't seem to come into play at all.

Yet a generativist twist on Ann's study can be formed around the following question: What do TSL and ASL signers know about the handshape inventories of their respective languages? Since their handshape inventories are not identical (as we will see), they must be learned; they cannot be entirely innate or shaped deterministically by ease of articulation, or

else they would be identical. Since they are learned, knowledge becomes relevant, and linguistic knowledge involves grammar. What is it about the grammars of TSL and ASL that makes their handshape inventories different and what role, if any, is played by markedness?

Generative phonologists themselves do not agree on the answer to this kind of question. On the one hand is a long tradition of treating phonemic inventories as essentially arbitrary collections like lexicons. Chomsky and Halle (1968) seem to take a position like this throughout most of their book, only acknowledging evidence for "natural" vs. "unnatural" inventories in their final chapter where they give a half-hearted theory of markedness. A recent advocate of this arbitrary-inventory position is Duanmu (2002). On the other hand are frameworks like radical underspecification theory (Archangeli and Pulleyblank 1994) and Optimality Theory (OT), which treat phonemic inventories as generated by grammar in the same way that grammar is presumed to generate the lexicon. A generativist (i.e. rationalist) argument in favor of the second (and currently dominant) view is that it requires less learning: rather than the possibly large set of phonemes themselves, the child only has to learn a small set of grammatical constraints on features and their combinations. OT makes it even simpler, since the constraints are presumed to be innate; now all that must be learned is their ranking. In this view, the phonemic inventory is still considered merely performance evidence for the underlying grammar, not a direct reflection of it; thus there may be accidental gaps in it, just as happens in the lexicon (see McCarthy 2002:68-82 for a thorough discussion).

In the case of TSL and ASL handshape inventories, it seems reasonable to suppose that the grammar's markedness constraints, if they really exist, would relate somehow to Ann's ease scores. In (13) I give a set of constraints and fixed universal rankings that attempt to encode her principles in OT terms.

- (13) a. *OPEN-FINGER family: The specified finger cannot be extended independently (i.e. without adjacent finger(s) also being extended).
- a'. Fixed universal ranking:
 {*OPEN-MIDDLE, *OPEN-RING} » {*OPEN-INDEX, *OPEN-PINKY, *OPEN-THUMB}
- b. MRP: The middle, ring, and pinky fingers must act as a group.
- c. *CURVED » *EXTENDED » *BENT » *CLOSED

Now let's see how well they handle the facts (the TSL data come from Chang et al. 2005, and the ASL data come from Tennant and Brown 1998). In Table 3, I have tried (and perhaps failed) to express in as clear a way as possible the similarities and differences in the TSL and ASL handshape inventories; each language has roughly 40-50 lexically contrastive handshapes, but I consider only a small subset. The subset is defined in terms of four fingers (index, middle, ring, pinky), each of which is in one of two finger configurations: open (fully extended or at most bent at the base) vs. closed (fully closed or at least as tightly curved as possible). Four fingers with two configurations each make 16 (= 4²) logically possible handshape patterns. Since Ann's constraints relate to fingers configured under their own power, holding fingers closed with the thumb is "cheating," so such handshapes are not included (more details relating to this point are given below). Each of the handshape types is then evaluated according to the constraints in (13ab); the universal rankings in (13a') and (13c) are set aside since they didn't prove to be helpful here. Note that the ordering of the handshape types in Table 3 runs roughly from most to least marked.

The six handshape types shared across both languages all appear near the bottom of Table 3, showing that they are relatively unmarked. The two languages also agree on rejecting five handshape types, and most of these are in the top, more marked, half of the table. So far these observations simply replicate those made by Ann (1996), but now in the context of handshape inventories instead of whole lexicons.

Table 3. Markedness of handshape types in TSL and ASL inventories. Each set of symbols represents a forward-facing right hand (the tilde is the thumb, in any configuration). Up arrows are extended fingers; down arrows are closed fingers.

		*OPEN-MID	*OPEN-RING	MRP	TSL	ASL
a.	↓↑↑↓~	*	*	*	✗	✗
b.	↓↑↑↑~		*	*	✓	✗
c.	↑↑↓↓~		*	*	✗	✗
d.	↓↑↓↑~		*	*	✗	✗
e.	↓↑↓↓~		*	*	✓	✗
f.	↑↑↓↑~		*	*	✗	✗
g.	↑↓↑↓~	*		*	✓	✗
h.	↓↓↑↓~	*		*	✓	✗
i.	↑↓↑↑~	*		*	✗	✓?
j.	↓↓↑↑~	*		*	✓	✓
k.	↑↑↑↓~	*			✗	✗
l.	↑↓↓↑~			*	✓	✓
m.	↑↓↓↓~			*	✓	✓
n.	↑↑↑↑~				✓	✓
o.	↓↓↓↑~				✓	✓
p.	↓↓↓↓~				✓	✓

What about the differences between the two languages? Table 3 implies that there is only one handshape type found in ASL but not TSL, namely (i). The question mark is there because I am not sure how widespread this handshape really is in ASL. The lone ring finger in (i) is usually held down by (or at least makes finger tip contact with) the thumb (as in ASL "seven"), but apparently some ASL signers have another lexically distinctive handshape involving a (nearly) closed ring finger without thumb support.⁶ However, the parallel handshape (f) really does seem to be missing in ASL; when the lone middle finger is closed or nearly so, it is held down by (or at least makes finger tip contact with) the thumb (as in ASL "eight"), but when it appears without support, it is only slightly curved (as in ASL "touch"). Similarly, ASL permits the finger configurations in (b) (closed pinky) and (k) (closed index) only if the lone finger is held down by the thumb. By contrast, TSL never seems to require thumb support in order to achieve finger closure. Despite some uncertainty about the data, then, it appears that the only special feature of ASL relative to TSL is that it generally favors thumb support for individually closed fingers. In particular, note that handshape type (b) is perfectly fine in TSL (as in TSL "eight"), while the closest parallel in ASL involves thumb support to hold the pinky closed (as in ASL "six").

In addition to (b), three additional handshape types are found in TSL but not ASL: (e, g, h). All of them violate MRP and either *OPEN-MIDDLE or *OPEN-RING. This suggests that TSL signers are more tolerant than ASL signers of certain markedness violations.

The constraints in (13) really seem to be doing some work. Not only do they capture "universals" (at least with regard to TSL and ASL), but also they seem to provide the right framework for understanding the differences between the two languages. Namely, at least in the case of what TSL signers know that ASL signers don't, this knowledge seems to be

⁶ This handshape is not listed by Tennant and Brown (1998) and I have not found it in Sternberg (1998) either, though this dictionary is arranged by English glosses, making it very difficult to look for relevant examples. Handshape type (i) is pictured in Corina and Sagey (1988), but unfortunately not in the context of a sign, so I'm not sure how closed the finger actually is in actual use.

expressed using the same constraints that define the universals: TSL signers know that for them, some of these universals are turned off.

Another possible reason for believing that an OT analysis like this is on the right track is a difference between TSL and ASL outside of their handshape inventories. As observed by Lee (2003), TSL and ASL share most local movements (e.g. finger wiggling), but only TSL has nonrepeated, sequential finger curving. This is found in TSL "how many" (a morphemic component of many signs), represented schematically in (14) as a sequence of handshapes labeled as in Table 3 (the thumb is closed against the palm throughout). As intermediate steps, the sequence goes through handshapes (k) (not in the TSL inventory and only allowed in ASL with thumb support) and (c) (not found in the inventory of either language). Thus again we see that TSL is more tolerant of difficult handshapes than is ASL. Of course, in fluent signing the closure of one finger is accompanied by the partial closure of the next, and in any case the process is so rapid that any physiological violation doesn't last long.

(14) (n) ↑↑↑↑~ (k) ↑↑↑↓~ (c) ↑↑↓↓~ (m) ↑↓↓↓~ (p) ↓↓↓↓~

After having made a case for a very reasonable (perhaps even convincing) OT analysis of the TSL and ASL inventory problem, let's apply some psychologist-style skepticism to it (in addition to the caveats made above about Table 3 and (14), which are of the sort that linguists already know they have to make). Most fundamentally, how justified are we in seeing any patterns in Table 3 at all? What is the chance probability that two languages will divvy up these sixteen (or thirty-two, if we include thumb support as a parameter) handshape types in ways at least as interesting as what we've observed here?

This question is much more difficult than the similar question we asked about the Southern Min Tone Circle. With the Tone Circle we could take loop size as an objective measure of theoretical interest, but we don't have that luxury now. But aren't the constraints in (13) directly translated from Ann's objective articulatory metric? Not quite. Just as data don't really speak for themselves without a theoretical context to make sense of them, the use of extra-grammatical information in linguistic analysis must also take into consideration the linguistic data themselves. To do otherwise is to ignore Bloomfield's hard-won lesson that linguists are really only qualified to analyze language, not the myriad forces that may or may not influence language. The Old Bloomfieldians are right about psychology (and physiology) occupying a more "basic position" than linguistics, but as the New Bloomfieldians point out, psychologists (and physiologists) need to learn from us linguists just as much as we do from them. It has often been the case that higher-level sciences have reached conclusions that lower-level sciences caught up with only later; Chomsky (2000) cites the example of hydrogen bonds, whose properties were well understood by chemists long before physicists developed a theory (quantum mechanics) that could explain them.

To exemplify Bloomfield's lesson with the data at hand, first consider the role of thumb support. ASL signers use it to sidestep the articulatory difficulty of lone finger closing, but TSL signers do not. This difference could be handled in OT by positing the constraint and rankings in (15). ASL ranking (15b) requires thumb support if a lone pinky is to be closed as in (b); TSL ranking (15c) permits lone pinky closing without thumb support. But a constraint that blocks thumb support is surely not motivated physiologically, given the extreme flexibility of the thumb. Thus if the linguistic data demand it (which is admittedly far from clear), its sole empirical support will come from the linguistic data themselves.

- (15) a. *THUMBSUPPORT: The thumb cannot be used to hold down fingers.
 b. ASL: *OPEN-RING » FAITH » *THUMBSUPPORT
 c. TSL: *THUMBSUPPORT » FAITH » *OPEN-RING

The proto-analysis in (15) hints at the same point in another way. In order to capture the observation that TSL permits unsupported pinky closure while ASL does not, we had to rank some markedness constraint below faithfulness. The only constraint made available by Ann's metric is *OPEN-RING. Note that ranking this constraint below faithfulness in TSL can also explain why handshape (e) is permitted in TSL but not ASL. The four other logically possible handshape types violating *OPEN-RING (a, c, d, f) are missing in TSL, but this could be dismissed as an accidental inventory gap. So far so good.

But now consider the other constraint against individual finger extension: *OPEN-MIDDLE. This also seems to be more violable in TSL than ASL, given that handshape types (g, h) are found only in the former. But three other handshape types violating this constraint have a more problematic status: handshape (k) isn't found in either language, handshape (j) is found in both languages, and handshape (i), as noted above, is found in some varieties of ASL, but not TSL. The first case might merely represent an accidental gap, but the non-surface-apparent appearance of the other two others in ASL cannot be handled without modifying the constraints (assuming that we refuse to give up the assumption that inventories are grammatically generated). Whatever these modifications should be like, they will no longer be guided by Ann's ease of articulation metric: we are again on our own, just as the New Bloomfieldians say we should be.

The unfortunate result, however, is that without external guidance, the "linguistically significant generalization" becomes an ever-shifting target. It is impossible to calculate how likely the pattern in Table 3 is to have arisen by chance because there is no way to determine ahead of time what will count as a "hit" and what will count as a "miss." This is a fundamental weakness of all observational (non-experimental) research where the hypotheses are generated and tested on the same data set. On top of this, we have the convenient (if reasonable) excuse of the "accidental gap." Consider the six handshape types violating *OPEN-RING. If we classify the two found in TSL (b, e) as "patterned" and the rest (a, c, d, f) as "exceptional," we can consult Table 2 to see that $p > .05$: we have insufficient evidence to support the reality of the pattern in the first place, let alone the accidental nature of the gaps. Moreover, by classifying the four missing handshape types in TSL as "exceptional" we undermine the claim we made at the outset that the handshapes in the top half of Table 3 are universally more marked.

Does this mean that we must dismiss all of the insights arising in the above analysis as mere illusions? Is it, in fact, impossible to carry out the sort of analysis I was trying to do in a scientifically valid way? I don't think so, but it should be apparent that a lot more evidence is needed, both within and across languages, before typological claims can be validated. This is true even when we have some external guidance for our hypotheses, as we do here; most of the time, linguists lack even this.

In a sense this case study isn't really fair, given the intentions of this paper. Typological research may indeed pose serious challenges, but at least the linguists are brave enough to try it; as noted earlier, psycholinguists generally are not. In the final case study I return to territory more comfortable from a psycholinguistic perspective: the use of non-traditional data and high-powered mathematical analysis.

6. Searching for grammar in Mandarin acceptability judgments⁷

Judgments have not been the primary data source used in phonology, which has instead relied on corpus data, usually dictionaries. This is so despite general acknowledgment of the inconclusive nature of corpus attestations. As pointed out by Kenstowicz and Kisseberth

⁷ Portions of this section are rewritten from Myers and Tsay (2005).

(1979) and many others, attested phonological forms may simply be memorized as wholes, already patterned from diachronic sound change, thereby making synchronic phonological grammar superfluous (Blevins 2004 argues that this may in fact be very close to correct). The phonologist's reliance on corpora is thus in sharp contrast to the methodology in generative syntax, where acceptability judgments are the primary data source (a kind of experimental data, as noted earlier). Of course, even if all phonologists suddenly began to rely on judgments like syntacticians, they would still have to deal with the competence-performance distinction. Judgments don't provide a direct route to grammar any more than corpora do.

Nevertheless, it is somewhat easier to think about judgments in terms of linking models than corpus attestations. Now the dependent variable is the directly observed judgment, not a probability that can only be estimated by collecting a lot of data, and among the independent variables are relatively well-understood performance factors operating within the head of a single speaker, rather than far more mysterious, perhaps unknowable, diachronic forces.

The most basic of the factors affecting phonological judgments is lexical status (attested vs. not). As we've seen, the lexical status of a form is only indirectly relevant to its grammaticality, since gaps may be accidental. Though real words are much more likely to be judged as acceptable than nonwords, from a generative perspective this is a mere performance effect. Similarly, more frequent words are likely to be judged as more acceptable. Again this must be recognized as a performance bias rather than a grammatical fact, since lexical exceptions aren't necessarily distributed in accordance with frequency.

Two further factors are less familiar, but prove to be crucial. The first is phonotactic probability, the typicality of the subsequences in a target item relative to words in the lexicon (e.g. English *stop* starts with typical sequence /st/, *sphere* with atypical sequence /sf/). The second is neighborhood density, which relates to the number of words similar to a target item (e.g. the form *lat* has more lexical neighbors in English, like *cat* and *lap*, than *zev*). These two factors are highly confounded: forms with high values for one also tend to have high values for the other. Nevertheless, there is strong evidence that they have distinct processing effects: phonotactic probability acts as a filter prior to lexical access, whereas neighborhood density only affects processing after competing forms in the lexicon have been accessed. Vitevitch and Luce (1999), Bailey and Hahn (2001), and Pylkkänen et al. (2002) provide three very different types of psycholinguistic evidence for the reality of the distinction between phonotactic probability and neighborhood density.

Because neighborhood density involves comparison among lexical items, it is essentially a measure of that old generativist bugaboo: analogy. By contrast, since phonotactic probability operates independently of the lexicon, it represents something closer to what a linguist would consider true grammatical knowledge, except that it is gradient and that it is defined in terms of the lexicon, rather than the other way around. For example, most generativists would consider /sf/ fully grammatical in English, with its rarity merely an accident of history.⁸

The story so far can be summarized with the linking model in (16): phonological acceptability judgments are a function of various known performance factors ("phonotactics" stands for phonotactic probability, "neighbors" for neighborhood density), hypothesized grammatical knowledge, and unknown sources of noise. The factors are each multiplied by coefficients, as described in section 3, but for simplicity I leave them out here.

$$(16) \text{ Judgment} = \text{Lexicality} + \text{Frequency} + \text{Phonotactics} + \text{Neighbors} + \text{Grammar} + \text{Noise}$$

⁸ Generativists don't form a monolithic group, of course. Hammond (2004) explicitly claims that phonotactic knowledge is gradient and lexically driven, and in fact includes both phonotactic probability and neighborhood density as components in the grammar.

The model in (16) assumes that the factors have independent, additive effects: raising or lowering any one factor will always raise or lower the judgment score, regardless of the values of the other factors. This picture is not necessarily correct, however. In particular, if the psycholinguist's notion of phonotactic probability really overlaps with the linguist's notion of grammar, these two factors should interact, rather than working independently. By contrast, if grammar works nothing like analogy, there should be no interaction between grammar and neighborhood density. Mathematically, the interaction of factors A and B is simply expressed as their product $A \times B$, so we can easily add interactions to an equation like (16). Like any other factor, then, we can test whether the coefficient of an interaction is sufficiently far from zero to reject the null hypothesis that the factors really do not interact.

The performance factors in (16) are well recognized in the psycholinguistic literature, but what about grammar itself? Can we justify its presence in the linking model? We addressed this question as part of a larger study on native-speaker acceptability judgments of Mandarin syllables (Myers and Tsay 2005). Our test items were syllables derived from all possible combinations of eight onset phonemes /p, p^h, m, f, t, t^h, n, l/, four vowel phonemes /a, i, u, ə/, three endings /n, ŋ/ and #, and four tones. Of these 384 syllables, 235 appear in real words in Mandarin (lexical syllables) and 149 do not (nonlexical syllables). Speakers were asked to judge, without time pressure, how similar to Mandarin each item was on a scale from 1 to 6, where 6 represented *zui xiang Guoyu* ("most like Mandarin"); for more details on our experimental procedures and data analyses, see Myers and Tsay (2005).

We coded each of the 384 syllables as grammatical or ungrammatical, adopting the standard generative assumption that grammaticality is a categorical property, even if judgments vary gradiently. Our classification was based on widely accepted principles of phonology in general, and Mandarin phonology in particular. These principles are listed below in (17). Together they defined 56 of our syllables as ungrammatical.

- (17) a. Lexical syllables are grammatical.
 b. Tone plays no role in determining grammatical status.
 c. Labial onsets may not combine with /ə/ in an open syllable (e.g., */pə/).
 d. Labial onsets may not combine with /əŋ/ (e.g., */fəŋ/).
 e. Labial onsets may not be followed by /u/ in closed syllables (e.g., */mun/).
 f. The sequence */fi/ is ungrammatical.

A few comments about (17) are in order. First, despite the distinction made above between lexicality and grammaticality, (17a) is justified by the fact that each of lexical syllable represents multiple morphemes; it is unlikely that all of these morphemes are lexical exceptions. Second, (17d) is a local feature of the Taiwan Mandarin spoken by our judges. Finally, the definition of grammar is based solely on patterns observed in the Mandarin lexicon, with no consideration given to "naturalness." In particular, we included principle (17f) even though it's not obvious that it is cross-linguistically well motivated.

To begin our analyses, we removed grammar entirely from the model in (16). The remaining factors turned out to do a very good job in accounting for judgments all by themselves: 72% of the variability was explained, much higher than what Bailey and Hahn (2001) had found for a similar grammar-free model of English syllable judgments. This means that only a rather small minority (28%) of the judgment pattern had to be ascribed to unknown noise factors. Then we added grammar as defined above, producing a new model that captured 73% of the judgment pattern. This increase was statistically significant, so grammar as we defined it did make a real contribution, but clearly it is only a very tiny contribution.

Next we turned to the issue of interactions. Recall that if our definition of "grammar"

really represents something like the lexicon-independent knowledge that linguists would demand it to be, we might expect an interaction between grammar and phonotactic probability but not between grammar and neighborhood density (analogy). What we actually found was more complex and interesting. When we added the grammar \times phonotactics interaction to the model in (16), the interaction was indeed significant, though the overall model still only accounted for 73% of the judgment pattern, no different from when we ignored the interaction. The reason for this lack of increase seems to relate to the fact that the interaction coefficient was negative, indicating that grammar and phonotactic probability complemented each other. This seems to suggest that what we called "grammar" and "phonotactic probability" are different facets of the same thing, just as linguists might expect.

When we returned to the model in (16) and added the grammar \times neighbors interaction, the coverage of the model increased significantly to 75%, implying that this interaction truly described new information about the judgment process. A final model variant, with both interactions, accounted for 76% of the judgment pattern, slightly but significantly more. The interaction between grammar and analogy contradicts what a linguist might naively expect, but I think that's only because the expectation is indeed naive. In fact, interactions between competence and performance have been observed for decades. Classic syntactic examples are shown (18): (18a) (from Chomsky 1965) is grammatical but unacceptable, while (18b) (from Montalbetti 1984) is ungrammatical but acceptable.

- (18) a. The man who the boy who the students recognized pointed out is a friend of mine.
 b. More people have been to Berlin than I have.

Our discovery of grammar-analogy interactions may thus be old news in a sense, but findings like this pose a serious challenge to generative reasoning. It's hard enough for linguists to take seriously the truism that linguistic data may be tainted by extra-grammatical factors, but if the effects of competence and performance are blended together, disentangling them becomes a major undertaking, perhaps even impossible.

The challenge of interpreting phonological judgment data is serious even if the interaction issue is set aside. Phonologists worried about the problems with corpus data addressed in the previous three sections should not think that judgment data somehow avoids them. The phonotactic probability and neighborhood density that have such strong influences on phonological acceptability judgments are themselves derived from the lexicon. Hence to a large extent a native-speaking judge is merely parroting back corpus statistics, rather than providing a totally new window into grammar.

7. Concluding remarks

As the reader might have guessed by now, one of my goals in this paper was to shock the average linguist out of an unjustified sense of complacency, but I don't want to be overly discouraging. Paraphrasing Bruce Hayes (via Sorace and Keller 2005:1498), it's true that "linguistics isn't hard enough," but that doesn't mean that a harder sort of linguistics is necessarily less fun. Progress will be a bit slower, but more steady. Abstractions will still abound, though only if they're really justified.

One of my current interests is on the engineering side of this: How can I help make "hard linguistics" as easy as possible? This paper is part of that; Table 2, for instance, can be taken as a practical tool for estimating how reliable a pattern is in a small corpus. More tools for basic phonological corpus analysis are still in the works, but already I've made public an early version of a program, called MiniJudgeJS, that helps to design, run, and analyze small-scale judgment experiments (Myers 2006). Though this program is designed specifically for

syntax, it can also be used to test some types of pragmatic, semantic, morphological, and phonological hypotheses.

Given the very different histories, goals, and philosophies of linguistics and psychology, a complete resolution of the Bloomfield paradox is probably still in the distant future. Nevertheless, it seems to be inevitable that linguistics will be gradually integrated more and more into the family of sciences, starting with psychology, its closest, if long estranged, neighbor.

References

- Abdel Rahman, R., van Turenout, M., & J. M. Levelt, W. (2003). Phonological encoding is not contingent on semantic feature retrieval: An electrophysiological study on object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29 (5), 850-860.
- Ann, J. (1993). *A linguistic investigation of physiology and handshape*. University of Arizona doctoral dissertation.
- Ann, J. (1996). On the relation between the difficulty and the frequency of occurrence of handshapes in two sign languages. *Lingua*, 98, 19-41.
- Ann, J. (2005). A functional explanation of Taiwan Sign Language handshape frequency. *Language and Linguistics*, 6 (2), 217-245.
- Ann, J. (2006). *Frequency of occurrence and ease of articulation of sign language handshapes: The Taiwanese example*. Washington, DC: Gallaudet University Press.
- Anttila, A. (2002). Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory*, 20, 1-42.
- Archangeli, D. & Pulleyblank, D. (1994). *Grounded phonology*. Cambridge, MA: MIT Press.
- Aronoff, M., Meir, I., & Sandler, W. (2005). The paradox of sign language morphology. *Language*, 81 (2), 301-344.
- Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory & Language*, 44, 569-591.
- Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge, UK: Cambridge University Press.
- Bloomfield, L. (1914). *An introduction to the study of language*. New York: Holt.
- Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart & Winston.
- Blust, R. (1999). Notes on Pazeh phonology and morphology. *Oceanic Linguistics*, 38 (2), 321-365.
- Bodman, N. C. (1955) *Spoken Amoy Hokkien*. Revised and published 1987, Ithaca, NY: Spoken Language Services.
- Botha, R. P. (1982). On 'the Galilean style' of linguistic inquiry. *Lingua*, 58, 1-50.
- Chambers, J. K. (1973). Canadian raising. *Canadian Journal of Linguistics*, 18, 113-135.
- Chang, J.-h., Su, S.-f., & Tai, J. H.-Y. (2005). Classifier predicates reanalyzed, with special reference to Taiwan Sign Language. *Language and Linguistics*, 6 (2), 247-278.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1980). *Rules and representations*. Oxford: Basil Blackwell.
- Chomsky, N. (2002). *On nature and language*. Cambridge, UK: Cambridge University Press.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row. Reprinted 1990, Cambridge, MA: MIT Press.
- Chomsky, Noam. (2000). Linguistics and brain science. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.) *Image, language, brain* (pp. 13-28). Cambridge, MA: MIT Press.
- Corina, D., & Sagey, E. (1988). Predictability in ASL handshapes with implications for feature geometry. Unpublished ms., Salk Institute and UCSD.
- Dell, Gary S. (1989). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. In W. Marslen-Wilson (Ed.) *Lexical representation and process* (pp. 136-165). MIT Press.
- Di Sciullo, A. M., & Williams, E. (1987). *On defining the word*. Cambridge, MA: MIT Press.
- Duanmu, San. (2002). Two theories of onset clusters. *Chinese Phonology*, 11 (*Special Issue: Glide, Syllable and Tone*), 97-120.

- Edelman, S., & Christiansen, M. H. (2003). How seriously should we take Minimalist syntax? *Trends in Cognitive Science*, 7 (2), 60-61.
- Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language* 47:27-52.
- Garrett, M. F. (1980). Levels of processing in sentence production. In B. L. Butterworth (Ed.) *Language production, vol 1: Speech and talk* (pp. 177-220). London: Academic Press.
- Garrett, M. F. 1988. Processes in language production. In F. J. Newmeyer (Ed.) *Linguistics: The Cambridge survey, vol III: Language: Psychological and biological aspects* (pp. 69-96). Cambridge, UK: Cambridge University Press.
- Gussenhoven, C., & Jacobs, H. (2005). *Understanding phonology, second edition*. London: Arnold.
- Halle, M. (1962). Phonology in generative grammar. *Word*, 18, 54-72.
- Halle, M., & Mohanan, K. P. (1985). Segmental phonology of Modern English. *Linguistic Inquiry*, 16, 57-116.
- Hammond, M. (2004). Gradience, phonotactics, and the lexicon in English phonology. *International Journal of English Studies*, 4, 1-24.
- Inkelas, S., Orgun, C. O., & Zoll, C. (1997). The implications of lexical exceptions for the nature of grammar. In I. Roca (Ed.) *Derivations and constraints in phonology* (pp. 393-418). Oxford: Clarendon Press.
- Joos, M. (1942). A phonological dilemma in Canadian English. *Language*, 18, 141-144.
- Kaye, J. (1990). What ever happened to dialect B? In J. Mascaró & M. Nespór (Eds.) *Grammar in progress: GLOW essays for Henk van Riemsdijk* (pp. 259-263). Dordrecht: Foris.
- Kenstowicz, M., & Kisseberth, C. (1979). *Generative phonology: Description and theory*. New York: Academic Press.
- Kiparsky, P. (1982). Lexical morphology and phonology. In *Linguistics in the Morning Calm: Selected Papers from SICOL-1981* (pp. 3-91). Seoul: Hanshin Publishing Company.
- Kiparsky, P. (2000). Opacity and cyclicity. *The Linguistic Review*, 17, 351-366.
- Kuhn, T. S. (1970). *The structure of scientific revolutions*. (Second edition, enlarged). University of Chicago Press.
- Lasnik, H. (2002). The minimalist program in syntax. *Trends in Cognitive Sciences*, 6, 432-437.
- Leben, W. (1978). The representation of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 177-219). New York: Academic Press.
- Lee, Hsin-Hsien. (2003). *Analyzing handshape changes in Taiwan Sign Language*. National Chung Cheng University MA thesis.
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. MIT Press.
- Levelt, W. J. M., and Indefrey, P. (2000). The speaking mind/brain: Where do spoken words come from? In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.) *Image, language, brain* (pp. 77-94). Cambridge, MA: MIT Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1-75.
- Li, P. J.-K., & Tsuchida, S. (2001). *Pazih dictionary*. Taipei: Institute of Linguistics.
- Marcus, G. F. (2001). *The algebraic mind: Integrating connectionism and cognitive science*. Cambridge, MA: MIT Press.
- McCarthy, J. (1999). Sympathy and phonological opacity. *Phonology* 16:331-399.
- McCarthy, J. (2002). *A thematic guide to Optimality Theory*. Cambridge University Press.
- McCarthy, J. (2006, April). Derivations: Optimal and otherwise. Paper presented at GLOW 2006, Barcelona, Spain.
- Mendoza-Denton, N., Hay, J., & Jannedy, S. (2003). Probabilistic sociolinguistics: Beyond variable rules. In R. Bod, J. Hay, & S. Jannedy (Eds.) *Probabilistic linguistics* (pp. 97-138). Cambridge, MA: MIT Press.
- Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In R. D. Luce et al. (Eds.) *Handbook of mathematical psychology, Vol. 2*. Wiley.
- Montalbetti, M. M. (1984). *After binding: On the interpretation of pronouns*. MIT Ph.D. dissertation.
- Moreton, E. (2004). Non-computable functions in Optimality Theory. In J. J. McCarthy (Ed.), *Optimality Theory in phonology: A reader* (pp. 141-163). Oxford, UK: Blackwell.
- Myers, J. (2006). MiniJudgeJS (Version 0.9.1) [Computer software]. Retrieved from <http://www.ccunix.ccu.edu.tw/~lngproc/MiniJudgeJS.htm>

- Myers, J., & Tsay, J. (2001). Testing a production model of Taiwanese tone sandhi. *Proceedings of the Symposium on Selected National Science Council Projects in General Linguistics from 1998-2000*, 257-279. National Taiwan University, Taipei.
- Myers, J., & Tsay, J. (2005, May). The processing of phonological acceptability judgments. *Proceedings of Symposium on 90-92 NSC Projects*, 26-45. Taipei, Taiwan.
- Ohala, J. J. (1986). Consumer's guide to evidence in phonology. *Phonology Yearbook*, 3, 3-26.
- Phillips, C., & Lasnik, H. (2003). Linguistics and empirical evidence: Reply to Edelman and Christiansen. *Trends in Cognitive Science*, 7 (2), 61-62.
- Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition*, 95, 201-236.
- Prince, A., and Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar*. Published 2004, Blackwell.
- Pylkkänen, L., Stringfellow, A., & Marantz, A. (2002). Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain & Language*, 81, 666-678.
- Reiss, C. (2003). Quantification in structural descriptions: Attested and unattested patterns. *The Linguistic Review*, 20, 305-338.
- Rice, K. (1989). *A grammar of Slave*. Berlin: Mouton de Gruyter.
- Schütze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.
- Smith, W. H., & Ting L.-f. (1979). *Shou neng sheng qiao* [Your hands can become a bridge], Vol. 1. Taipei: Deaf Sign Language Research Association of the Republic of China.
- Smith, W. H., & Ting L.-f. (1984). *Shou neng sheng qiao* [Your hands can become a bridge], Vol. 2. Taipei: Deaf Sign Language Research Association of the Republic of China.
- Sorace, A., & Keller, F. (2005). Gradience in linguistic data. *Lingua*, 115, 1497-1524.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, 74, 1-29.
- Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, 21, 291-311.
- Sternberg, M. L. A. (1998) *American Sign Language (unabridged edition)*. Harper Collins.
- Tennant, R. A., & Brown, M. G. (1998). *The American Sign Language handshape dictionary*. Washington, DC: Gallaudet University Press.
- Tsay, J. & Myers, J. (1996). Taiwanese tone sandhi as allomorph selection. *Proceedings of the Berkeley Linguistic Society*, 22, 394-405.
- Tsay, J. (1994). *Phonological pitch*. Doctoral dissertation, University of Arizona, Tucson.
- Tsay, J., Myers, J., & Chen X-J. (2000). Tone sandhi as evidence for segmentation in Taiwanese. *Child Language Research Forum*, 30, 211-218.
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory & Language*, 40, 374-408.
- Wang, W. S-Y. (1967). Phonological features of tone. *International Journal of American Linguistics*, 33, 93-105.
- Yang, J. H., & Fischer, S. D. (2002). Expressing negation in Chinese Sign Language. *Sign Language & Linguistics*, 5 (2), 167-202.
- Yip, M. (1980). *The tonal phonology of Chinese*. Doctoral dissertation, MIT.
- Zoll, C. (2003). Optimal tone mapping. *Linguistic Inquiry*, 34 (2), 225-268.