

To appear in *Concentric* (<http://www.eng.ntnu.edu.tw/concentric-linguistics/>). 2007 ms.

Linking data to grammar in phonology: Two case studies*

James Myers
National Chung Cheng University

Abstract

Linguists have long debated the proper role of psychology in grammatical theorizing. Rather than taking the traditional formalist or functionalist positions in this debate, this paper instead advocates testing grammatical claims using the methods, but not necessarily the concepts, of psychology. This means developing explicit models of the links between hypothesized mental entities and observed data, often with the use of quantitative analysis. This more skeptical approach is applied in analyses of two phonological issues: epenthesis and vowel harmony in the Formosan language Pazih, and a comparison between the handshape inventories of Taiwan Sign Language and American Sign Language. In both case studies, the linking model approach undermines certain conclusions that might otherwise have been accepted uncritically, but at the same time it strengthens our confidence in others.

Keywords: phonology, psychology, quantitative linguistics, Pazih, Taiwan Sign Language

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, but by the time of his second book, Bloomfield had seen the introspective approach demolished in the rise of behaviorist psychology (Kess 1983). The two Bloomfields still define the range of attitudes

* Portions of this paper were presented at the 14th Annual Conference of the International Association of Chinese Linguistics & 10th International Symposium on Chinese Languages and Linguistics Joint Meeting in Taipei, May 2006. I thank the audience for their comments, as well as an anonymous reviewer for *Concentric*. Research was supported in part by National Science Council (Taiwan) grants NSC 94-2411-H-194-018 and NSC 95-2411-H-194-005.

that linguists take towards psychology and other fields of relevance to the study of language. 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 (functionalist phonologists would include physiology in their fundamental framework as well). On the other side are the New Bloomfieldians, who insist that language must be studied for itself, regardless of what the psychologists or physiologists may say; formal linguists would go in this category. Note that nobody seriously denies that linguistics is a branch of psychology; the contrast, rather, concerns the practical independence of the two disciplines. Thus the Old Bloomfieldians Derwing and Baker (1978:26) declare that "one should posit only those constructs [...] that are compatible, if possible, with those that have been shown to be required by psychological research," whereas the New Bloomfieldian Chomsky (1971:44) claims that "precisely because the proposed principles [of formal linguistics] are not essential or even natural properties of any imaginable language, they provide a revealing mirror of the mind (if correct)."

However, neither of these two extreme positions captures the relationship between linguistics and allied fields in the most productive way. In contrast to the attitude taken by the New Bloomfieldians, we shouldn't pretend that grammar floats around independently of the mind and body in which it is situated, especially while proclaiming simultaneously that linguistics is a cognitive science. But in contrast to the Old Bloomfieldians, we also shouldn't hitch the fate of linguistics to whatever psychological (or physiological) constructs one happens to find appealing (as Bloomfield himself learned from the downfall of introspectionism).

A more productive solution to the Bloomfield paradox, I suggest, is to unite the content of theoretical linguistics with the methodological structure of more mature sciences, especially psychology, arguably the closest neighbor to linguistics in the family of sciences. For example, the content of mainstream theoretical phonology consists of its facts, primarily data from dictionaries, and its hypotheses about mental grammar. The structure offered by psychology is what may be called linking models (see also Myers 2007). 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 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" (since, after all, it is the linguists who are the grammar experts, not the 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 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, linguistic data are usually treated as direct evidence for grammar, ignoring all the complex links in between. In this paper I turn the spotlight on these links.

I begin the discussion by demonstrating the nature and benefits of linking models in psychology. I then illustrate linguistic linking models by unpacking the logic implicit in the analyses of two very different case studies, both relating to the phonology of languages in Taiwan: epenthesis and vowel harmony in the Formosan language Pazih, and a comparison between the handshape inventories of Taiwan Sign Language and American Sign Language. The first case illustrates the problems posed by testing grammatical hypotheses with data from one language, and the second shows what happens when one tries to test typological hypotheses with cross-linguistic data.

2. Linking models

Linking models make the cognitive sciences possible. 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 an explicit and testable linking model, the behaviorists have every right to complain: A mind unlinked to behavior, or linked only in vague, commonsensical terms, is not something about which science is capable of saying anything.

An example of an explicit linking model commonly used in psychology is the additive model (S. Sternberg 1998), which links reaction time to experimental stimuli with hypothesized mental processes. Suppose we hypothesize that process A operates independently of process B. To test this, we vary the task in ways that are expected to affect, separately, how long processes A and B take to complete. Our hypothesis is supported if the effects of the two types of variation on reaction time (RT) are additive, that is, if $RT = A + B$ rather than $RT = A + B + AB$ or something more complex. This model has three equally important components: mental entities that cannot be observed directly (A and B), fully objective and well-defined measurements (RT), and the hypothesized structural relationship between them ($RT = A + B$). By contrast, traditional linguistic models tend to focus only on abstract mental entities (e.g. grammar), giving less emphasis to the nature of the observations (e.g. corpus data vs. native-speaker judgments) and their relationship with the mental entities (e.g. the fact that corpus data are records of language production whereas judgments reflect language perception and comprehension).

Adopting linking models immediately leads to two other defining characteristics of modern psychology, characteristics it shares with other mature sciences. The first 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 Unlike the behaviorists, who were forbidden to discuss the mind at all, cognitive psychologists are free to hypothesize anything they like, but that's only the beginning of the process. Precisely because the mind is not directly accessible, evidence for any claim made about it must be extremely compelling if skeptics are to be won over.

Yet with good evidence they can be won over, even about claims that are initially quite counter-intuitive. For example, consider the claim that a quick glance at a complex visual display generates a highly detailed but extremely short-lived image in the mind. This sounds inherently unlikely, and in any case, how could we prove that the image holds so much detail if it fades too quickly to be fully described? Sperling (1960) found a way: test experimental subjects 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.

The second characteristic of the psychologist's approach to science is the use of quantitative methods, including statistical analysis. This is because, in its essence, a linking model is a statement about the correlation between observed data and a theoretical claim.

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

Thus a schematic linking model can be sketched as in (1).

(1) Observed data = Grammar + Known extra-grammatical factors + Unknown

Importantly, if our measures are quantitative (as in the additive model, for example), we can interpret the model in (1) literally, as a calculable mathematical equation. In particular, that "Unknown" component can be tamed through statistics. Though by definition we can't know what exactly is in it (it is treated as "random"), we can measure the size of it quite precisely. Only if it is sufficiently small are we justified in claiming that the linking model really "explains" the observations. A statistical model is thus 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, $p < 0.05$, that is, chance probability less than 1 out of 20, is usually taken as statistically significant).

The reader should not conclude from the above discussion that linking models, skepticism, and statistics, designed as they are to put claims through the most stringent criteria before they are accepted, will inevitably take all the fun out of linguistics by debunking its most exciting 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 section.

3. 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 (2a). The problem is that there are a sizable number of exceptions in which the intervening vowels do not show vowel harmony, as in (2b).

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

Indeed, in the short but presumably complete list given by Li and Tsuchida (2001:20-21), over one fourth of the total (12 out of 45) are exceptions. This seems like a lot. Is it, indeed, enough for a skeptic to reject vowel harmony as being the result of mere chance?

To see the problem in linking-model terms, note that the data are corpus attestations: "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 chance alone. 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.

If our results are due to pure chance, the pattern can be described with the so-called binomial distribution, which is related to the familiar bell-shaped "normal" curve, but unlike the latter, can be applied even to very small samples like ours. We can easily compute probabilities with this distribution using the Excel function in (3), where a and b are the two event types (e.g. heads vs. tails for coin flips, or 1 vs. any other number for rolling a die) and

prob is the probability of each event (e.g. 1/2 for coin flips and 1/6 for rolls of a die).²

(3) =MIN(1, 2*BINOMDIST(MIN(*a*, *b*), *a+b*, 1-*prob*, TRUE)

According to this function, MIN(1, 2*BINOMDIST(MIN(33, 12), 45, 1-1/2, TRUE) gives us $p = 0.00246 < 0.05$, so we are justified in rejecting the null hypothesis: vowel harmony is a statistically significant pattern. This is good, but we can do even better. Note that this analysis assumes that the chance probability of a word harmonizing is 50%, but this only makes sense within a framework where harmonization is a discrete element of grammar that is either present or absent. A less theory-laden approach would assume that the fundamental units for calculating chance are the vowels themselves. Since Pazih has four phonemic vowels (/i/, /u/, /e/, /a/), the chance probability of choosing a harmonizing intervening vowel is 1/4. Similarly, 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 p value for the 33 patterned and 12 exceptional words can be calculated with MIN(1, 2*BINOMDIST(MIN(33, 12), 45, 1-1/4, TRUE), which gives the vanishingly small $p = 0.0000000003$. Thus the simultaneous harmonization of all four vowels is extremely unlikely to be an accident.

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 Ohala 1986 comes close to doing). Therefore it's worthwhile to see how much we can infer about Pazih mental grammar from the corpus 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 (4a), or a glide as in (4b). 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

² TRUE tells Excel to compute probability rather than merely the height of the distribution curve. Note that the given function actually doubles the probability (hence the "2*" in the function, which in turn necessitates the "MIN(1, ..." since doubling may raise p over 1). Doubling the p value is recommended since this makes the test more skeptical, setting aside any prior bias we may have about which event type (*a* vs. *b*) we would prefer to be more common.

structure in the language."

- (4) 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 do need featural content on the surface, this content must be predictable. One way to achieve this is via vowel harmony.

But now those twelve exceptions come back into our story. The featural content of the vowels in these twelve words is not predictable. Since speakers nevertheless know how to pronounce the exceptions, they must 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 are therefore forced to the conclusion (or so it seems) that the exceptional epenthetic vowels must be underlyingly floating. This analysis is illustrated in (5), where (5a) and (5b) show patterned and exceptional words, respectively (I set aside the question of what mechanisms lie behind the CVC reduplication).

- (5) a. u u
 | |
 h V r h V r
 "steam, vapor"
- b. u a u
 | | |
 h V r h V r
 "bald"

Look what we have discovered in the black box of the mind, a linguist might say at this point: floating vowels! 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 formal 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 is extremely unlikely that twelve independent mistranscriptions could have occurred by chance, and we have no reason for thinking that the fieldwork methodology was biased.

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 (6). If (6a) 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*?

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

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* may be morphologically related to *say* "question marker," which might put a protective morpheme boundary of some type 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* vs. *repeat-repetition*).

A third possibility would be that neither form in (6) 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 (6) involves syllables that are intermediate in illicit status: they do have codas, but "weak" (quasi-vowel) 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.

The next level beyond Li and Tsuchida's transcriptions and generalizations are their diachronic source. Here Li and Tsuchida 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 both the patterned forms and the exceptions as wholes? Note that if this hypothesis is correct, there is no need to posit floating vowels.

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, but such cases seem to be non-systematic, targeting random parts of a few random morphemes. By contrast, the allomorphs here would have to be freely generated for an entire lexical class.

This scenario also doesn't fit with what is known about the diachronic emergence of vowel epenthesis (see Blevins 2004:155-158). Vowel epenthesis typically 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 acoustically 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 of Anttila's

stochastic grammar). Eventually both harmonic and non-harmonic vowels were treated as lexically specified components of the morphemes, and simply memorized. It should be noted, however, that the examples in (6) pose the same problems for this diachronic analysis as for the floating-vowel 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 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 provided the motivation for epenthesis while segmental phonology provided the implementation. The floating vowel analysis distinguishes motivation from implementation in a similar way. While the prosodic/segmental split may not seem dramatic, not all phonological theories have recognized it (e.g. Chomsky and Halle 1968), and it is nice to get further independent evidence.

Second, the fact that the vowel harmony is total rather than partial shows that speakers are capable of treating segments as wholes, rather than as arbitrary collections of articulatory or perceptual features. Our statistical analyses are also consistent with there being a single harmony process rather than separate ones for each vowel phoneme. Again, this conclusion may have lost its drama through over-familiarity, but it represents a rather deep insight into the nature of the human language faculty. Generating identical copies of whole segments is not something that can phonetic processes can accomplish by themselves, 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 analogy or phonetics alone.

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), it also applies to the situation faced by scientists, who, like babies, are trying to test hypotheses underdetermined by the evidence (see 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 perhaps not sufficiently grammar-oriented for some formal phonologists to find interesting, I assumed *a priori* that Pazih was a "normal" language, and therefore, based on what was known about *other* languages, that Pazih probably went through certain historical stages.

These considerations lead us to the next case study, which explores the role of linking models in cross-linguistic research.

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

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) 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). A recent overview of sign language universals is Sandler & Lillo-Martin (2006), which includes discussion of universals in sign phonology. Sign phonology, 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. However, though sign universals are becoming better understood, little attention has been paid to typological differences across sign languages, especially in phonology. 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 has compared the handshapes of ASL and TSL. She supplements her typological data with an explicit theory of ease of articulation in terms of physiological facts, including those listed in (7).

- (7) 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.

These considerations allow Ann to quantify the ease of articulation for any particular hand configuration, independent of linguistic considerations. Her key prediction 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.

Ann's research points to an open question about the shape of phonological grammar in the mind. Namely, what do TSL and ASL signers actually 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?

Phonologists do not agree on the answer to this kind of question. On one side 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 side 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 (see McCarthy 2002:68-82 for a thorough discussion). In the latter view, the phonemic inventory is considered mere performance evidence for the underlying grammar; thus there may be accidental gaps and exceptional ("ungrammatical") elements in it, just as happens in the lexicon.

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 (8) I give a set of constraints and fixed universal rankings that attempt to encode her principles in OT terms.

- (8) a. *EXTEND(FINGER): Finger cannot extend independently of (both) adjacent finger(s)
 a'. Fixed universal ranking:
 { *EXT(MIDDLE), *EXT(RING) } » { *EXT(INDEX), *EXT(PINKY), *EXT(THUMB) }
 b. MRP: The middle, ring, and pinky fingers must act as a group.

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). The table in (9) gives a sense of the similarities and differences between the TSL and ASL handshape inventories. Each language has roughly 40-50 lexically contrastive handshapes, but I consider only a small subset 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). Each set of symbols in the first column of the table represents a forward-facing right hand: Up arrows are extended fingers, down arrows are closed fingers, and the tilde is the thumb, in any configuration. Four fingers with two configurations each make the 16 (= 4²) logically possible handshape patterns listed in the table. 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 on this issue below). Each of the handshape types is then evaluated according to the constraints in (8). Note that *EXT(MID) bans handshapes where the middle finger is extended and one or both of the adjacent fingers is closed, and similarly for *EXT(RING). Based on their constraint violations, the handshapes in (9) are ordered roughly from most to least marked.

(9) Markedness of handshape types in TSL and ASL inventories

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

The six handshape types shared across both languages all appear near the bottom of (9), 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.

What about the differences between the two languages? The data in (9) imply that there is only one handshape found in ASL but not TSL, namely (9i). The question mark is there because I am not sure how widespread this handshape really is in ASL. The lone ring finger in (9i) 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.³ However, the otherwise similar handshape (9f) 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, not closed (as in ASL "touch"). Similarly, ASL permits the finger configurations in (9b) (closed pinky) and (9k) (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 a special feature of ASL relative to TSL is that it generally favors thumb support for individually closed fingers. In particular, note that handshape (9b) 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 (9b), three additional handshapes are found in TSL but not ASL: (9e, 9g, 9h). All of them violate MRP and either *EXT(MIDDLE) or *EXT(RING). This suggests that TSL signers are more tolerant than ASL signers of certain markedness violations.

In summary, then, the constraints in (8) really seem to be doing some work. Not only do

³ This other handshape is not listed by Tennant and Brown (1998) and I have not found it in M. L. A. Sternberg (1998) either, though this dictionary is arranged by English glosses, making it very difficult to look for relevant examples. Handshape type (9i) is pictured in Corina and Sagey (1988), but unfortunately not in the context of a sign, so I'm not sure if it's truly contrastive.

they capture markedness "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 expressed using the same constraints that define the universals: TSL signers know that for them, some of these universals are turned off. This cross-linguistic difference can be expressed formally with ranked OT constraints as in (10).

(10) a. ASL

↓↑↑↑~	*EXT(RING)	FAITH
↓↑↑↑~	*	
☞ ↑↑↑↑~		*

b. TSL

↓↑↑↑~	FAITH	*EXT(RING)
☞ ↓↑↑↑~		*
↑↑↑↑~	*	

An independent reason for believing that an OT analysis like this is on the right track is a difference between TSL and ASL phonology that goes beyond 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 (11) as a sequence of handshapes. As intermediate steps, the sequence goes through two handshapes that violate *EXT(MIDDLE) or *EXT(RING) (marked *_M and *_R, respectively). Interestingly, the first is not in the TSL inventory and is only allowed in ASL with thumb support, while the latter is not found in the inventory of either language. Thus again we see that TSL is more tolerant of *EXT(FINGER) violations than is ASL.

(11) ↑↑↑↑~ ↑↑↑↓~ (*_M) ↑↑↓↓~ (*_R) ↑↓↓↓~ ↓↓↓↓~

Having made the case for an apparently reasonable OT analysis of the TSL and ASL inventory problem, let's apply some linking-model skepticism to it. Note first that we cannot justify the analysis by claiming that the constraints in (8) were translated directly from Ann's objective physiological metric, hence they must be correct. 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 analyses 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 (in this case, physiology).

To exemplify Bloomfield's lesson with the data at hand, 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 (12). ASL ranking (12b) requires thumb support if a lone pinky is to be closed as in (9b), which would otherwise violate *EXT(RING) by leaving the ring finger extended on its own. By contrast, TSL ranking (12c) permits lone pinky closing, but bans thumb support. But a constraint that blocks thumb support is unlikely to be motivated physiologically, given the extreme flexibility of the thumb (after all, we are the species with the famed opposable thumb). Nevertheless, the data seem to suggest that whether or not to use thumb support is part of what TSL and ASL signers know about their respective languages.

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

Because we cannot justify our analysis solely on the basis of grammar-external motivations, we must make sure that it is well-supported by the data we do have, within a linking model framework. In this regard, perhaps the most doubtful part of our analysis is the argument about handshapes that appear in sequential finger curving in TSL. In the fluent articulation of sequential finger curving, the closure of one finger is always accompanied by the partial closure of the next, and in any case the process is so rapid that any physiological awkwardness doesn't last long. Moreover, based on how we currently understand the structure of phonological systems, it's not obvious that handshapes that appear only as part of more complex movements should be linked to arguments about inventories, any more than allophones should be lumped together with phonemes in the analysis of spoken languages.

Even if we restrict our attention to the inventory data in (9), we must recognize that they do not represent grammar directly. Not only are there questions about the status of handshape (9i) in ASL and the best way to take thumb support into account, but an inventory, according to the theoretical assumption we are adopting, is merely a performance realization of grammar acting diachronically. Given these considerations, how justified are we in claiming that the data in (9) reveal a systematic rather than an accidental difference between the inventories of TSL and ASL?

If the TSL inventory showed all of the handshapes in (9a-k) while the ASL inventory were missing all of them, it would be easy to justify the claim that TSL grammar ranks the *EXT(FINGER) constraints above FAITH while ASL has the reversed ranking. Unfortunately, the actual situation is far less clear-cut. This is illustrated in (13), which classifies the fifteen handshapes (leaving out the unclear (9i)) as whether they are found only in TSL, only in ASL, in both, or in neither.

(13)

		ASL	
		✗	✓
TSL	✗	5	0
	✓	4	6

The eleven handshape pairs that are treated the same way by the two languages may help support the need for universal OT constraints (or at least Ann's physiological motivations for them), but they are irrelevant for testing the systematicity of possible cross-linguistic differences. This leaves only the four discordant handshape pairs. The problem now becomes a statistical one: We have flipped a coin four times (the four discordant handshape pairs), and all four times it came up heads (TSL). Applying the binomial distribution to discordant pairs in a table like (13) is technically called the exact McNemar test, but we can compute it the same way as before: $p = \text{MIN}(1, 2 * \text{BINOMDIST}(\text{MIN}(4, 0), 4, 1-1/2, \text{TRUE})) = 0.125 > 0.05$. We get an even worse result if (9i) turns out to be a legal handshape in ASL: $\text{MIN}(1, 2 * \text{BINOMDIST}(\text{MIN}(4, 1), 5, 1-1/2, \text{TRUE})) = 0.375$.

Do these non-significant results force us to reject the hypothesis that the TSL and ASL inventories differ systematically? Not at all. We may not be justified in rejecting the null hypothesis, but this doesn't mean that we must accept it. We simply need more cross-linguistic data. Perhaps if we examined all of the handshapes in the two languages we would be able to bring the trend seen in (13) up to statistical significance. Or perhaps a survey of handshape inventories from many sign languages would reveal that they fall into

two distinct classes, one with TSL-like languages and the other with ASL-like languages. A larger data set may even allow us to find statistical support for supplementary observations like the one about sequential finger curving.

There are three main lessons to be noted here about testing typological hypotheses. First, while extra-grammatical arguments, whether formalist or functionalist, can help motivate hypotheses, the hypotheses themselves only win or lose on the basis of actual linguistic data. Second, typological hypotheses are secondary hypotheses, built on top of hypotheses about individual grammars. These two lessons lead directly to the third: the linking models in typological research can be quite complicated and difficult to apply, requiring quite large data sets to test quite subtle hypotheses. This helps explain why controversy continues to rage among linguists over the proper way to understand cross-linguistic similarities and differences.

5. Conclusions

In both of the case studies presented in this paper, the core of the discussion focused on analyses typical of the mainstream theoretical literature. The Pazih analysis was essentially formalist, and the TSL/ASL analysis was partially formalist and partially functionalist (since the formal constraints were motivated by physiological considerations). In both cases, however, I applied a bit more skepticism than is typical in theoretical linguistics, whether in the formalist or the functionalist traditions. In particular, I explored how a skeptic's confidence in the analyses might be affected by an understanding of the linking models separating the hypothesized grammatical entities from the data as actually observed.

When placed within these linking models, confidence in some of the more dramatic theoretical claims was considerably weakened. For example, there are somewhat "boring" diachronic alternatives to the floating vowel hypothesis in Pazih, and the grammar-generated inventory hypothesis for TSL and ASL did not quite reach statistical significance. At the same time, however, the claims that did manage to survive the stringent tests of skepticism, such as the rule-governed (not analogical) nature of Pazih vowel harmony, can now be accepted with much greater confidence than they might have been otherwise.

Given the very different histories, goals, and philosophies of linguistics and psychology, a complete resolution of the Bloomfield paradox may still lie 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. Theoretical abstractions will still abound, even those currently rejected as implausible, but only if they're fully justified by the evidence. Adopting a more skeptical attitude about the links between data and hypotheses may make grammatical theorizing a bit harder and slower than linguists are accustomed to, but it will also make progress steadier and more reliable.

References

- 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.
- 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. Facsimile reprinted 1983 Amsterdam: John Benjamins.
- Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart & Winston
- Blust, R. (1999). Notes on Pazeh phonology and morphology. *Oceanic Linguistics*, 38 (2), 321-365.
- 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. (1971). *Problems of knowledge and freedom*. Random House.
- Chomsky, N. (1980). *Rules and representations*. Oxford: Basil Blackwell.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row. Reprinted 1990, Cambridge, MA: MIT Press.
- Corina, D., & Sagey, E. (1988). Predictability in ASL handshapes with implications for feature geometry. Unpublished ms., Salk Institute and UCSD.
- Derwing, B. L., and Baker, W. J. (1978). On the re-integration of linguistics and psychology. In R. N. Campbell & P. T. Smith (Eds.), *Recent advances in the psychology of language: Formal and experimental approaches* (pp. 193-218). Plenum Press.
- Duanmu, San. (2002). Two theories of onset clusters. *Chinese Phonology*, 11 (Special Issue: Glide, Syllable and Tone), 97-120.
- Kess, J. F. (1983). Introduction. In Bloomfield (1914/1983) (pp. xvii-xxxviii).
- Lee, Hsin-Hsien. (2003). *Analyzing handshape changes in Taiwan Sign Language*. National Chung Cheng University MA thesis.
- 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. (2002). *A thematic guide to Optimality Theory*. Cambridge University Press.
- Myers, J. (2007). Generative morphology as psycholinguistics. In G. Libben & G. Jarema (Eds.), *The mental lexicon: Core perspectives*. Elsevier.
- Ohala, J. J. (1986). Consumer's guide to evidence in phonology. *Phonology Yearbook*, 3,3-26.
- Reiss, C. (2003). Quantification in structural descriptions: Attested and unattested patterns. *The Linguistic Review*, 20, 305-338.
- Sandler, W., & Lillo-Martin, D. (2006). *Sign language and linguistic universals*. Cambridge, UK: Cambridge University 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.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, 74, 1-29.
- Sternberg, M. L. A. (1998) *American Sign Language (unabridged edition)*. Harper Collins.
- Sternberg, S. (1998). Discovering mental processing stages: The method of additive factors. In: *An Invitation to Cognitive Science, Vol. 4: Methods, Models, and Conceptual Issues* (D. Scarborough and S. Sternberg, eds.), pp. 703-863. MIT Press, Cambridge, MA.
- Tennant, R. A., & Brown, M. G. (1998). *The American Sign Language handshape dictionary*. Washington, DC: Gallaudet University Press.