[Preprint of: Myers, J. (2007). Generative morphology as psycholinguistics. In G. Jarema & G. Libben (Eds.), *The mental lexicon: Core perspectives* (pp. 105-128). Amsterdam: Elsevier.]

Generative morphology as psycholinguistics

James Myers

Graduate Institute of Linguistics, National Chung Cheng University, Min-Hsiung, Chia-Yi 62102, Taiwan

Keywords: generative linguistics, compounds, Chinese, competence, performance, Distributed Morphology

1. Introduction

The title of this chapter is, of course, deliberately provocative. How can generative linguistics be psycholinguistics? Both sciences test mentalistic hypotheses about language, but their differences go beyond the cliché that linguists study competence while psycholinguists study performance. Instead they represent two different cultures, with linguistics more deductive and rationalist, psychology more inductive and empiricist. Psychologists, who are concerned with "cause and effect" (as Miller, 1990, p. 321, observes), run carefully designed experiments, conduct quantitative analyses, and write papers with explicit method sections. Generative linguists, who (Miller says) prefer "simplifications as explanations," don't.

And what is "generative morphology"? Isn't morphology lexical knowledge, and isn't grammar what permits one to go beyond rote lexical memory? Chomsky (1957) put regular interactions between word and sentence structure into the syntax, not the lexicon, and Chomsky and Halle (1968) put regular interactions between word and sound structure into the phonology; even today, most generative linguists are syntacticians and most of the rest are phonologists. As for those few who specialize in morphology itself, Marantz (1997, p. 202) complains that "when morphologists talk, linguists nap."

Yet my title is no oxymoron. Generative morphology, like the rest of generative linguistics, is indeed psycholinguistics, albeit methodologically sloppy psycholinguistics. Moreover, the lack of canonical pronouncements on what generative morphology is supposed to be is actually fortuitous, since it forces us to think hard about the notion of "generative grammar," and this is essential if we are to see grammar as psychological and psycholinguistics as grammar-oriented.

I start in section 2 by introducing the central notion of competence-performance linking models and explaining their relevance to morphology. Special focus is given to the theory of Distributed Morphology (Halle & Marantz, 1993; Harley & Noyer, 2003), not because I have any special fondness or animosity towards it, but because it is arguably the "most generative" of morphological theories. In section 3 I apply competence-performance linking models to a specific morphological question: Is there any difference between affixation and compounding in Chinese, and more to the point, how could we tell? I then analyze three sources of new evidence bearing on the question, each requiring its own linking model: two traditionally "linguistic"

(native-speaker judgments and corpus analyses) and one traditionally "psycholinguistic" (reaction times in lexical decision tasks).

2. Competence and performance in morphology

Far from defining the border between linguistics and psychology as often assumed, the competence-performance distinction of Chomsky (1965) actually provides the conceptual basis for fitting the study of mental grammar comfortably into the broader study of language in the mind. In this section I show how, focusing on morphology.

2.1. Competence-performance linking models

Generative linguistics is given its most succinct definition on page 4 of Chomsky (1965). There Chomsky states that "linguistic theory is mentalistic, since it is concerned with discovering a mental reality underlying actual behavior," that is, "a description of the ideal speaker-hearer's intrinsic competence." Competence is defined as "the speaker-hearer's knowledge of his language," that is, mental grammar, as opposed to performance, which is "the actual use of language in concrete situations." A linguist's grammar is called "generative" if it is "perfectly explicit — in other words, if it does not rely on the intelligence of the understanding reader but rather provides an explicit analysis of his contribution."

This passage has had a rather controversial history, but if we translate the linguistese into psychologistese, the concepts actually transfer quite well. The most obvious cognate is mentalism; nobody wants to return to the bad old behaviorist days, when the mind was dismissed as unscientific. Less obviously, the "ideal speaker-hearer" (particularly notorious among sociolinguists) also plays a starring role in psycholinguistics: S/he hovers in the Platonic statistical space over the heads of the actual experimental participants, the population mu to their sample x-bars. Similarly, psychologists believe in "universal psychology" just as much as generative linguists believe in "universal grammar," or how else could a study on Chinese lexical access have any bearing on theories of lexical access more generally?

Moreover, the goal of "discovering a mental reality underlying actual behavior," which is at the core of the competence-performance distinction, is what cognitive psychologists aim for every day. Psychologists know not to confuse overt behavior with the underlying mental operations themselves. In the same way, syntacticians know that acceptability judgments, their favorite data source, are partly contaminated by parsing effects (Phillips & Lasnik, 2003), and phonologists know that dictionaries, their favorite data source, are partly contaminated by historical accidents (Kenstowicz & Kisseberth, 1979). As Penke & Rosenbach (2004, p. 492) emphasize, "there is no such thing as competence data," only performance data, and performance data include not just judgments and dictionaries, but also traditionally "psycholinguistic" evidence like "rapidity, correctness, and uniformity of recall and recognition" (Chomsky, 1965, p. 10). It may not always be obvious where to draw the line between competence and performance, but similar demarcation problems plague psychology as well, as with prelexical vs. postlexical stages in lexical access (see 2.3).

But what really makes generative linguistics psychology is the notion of generativity itself. Chomsky (1980, p. 48) suggests that (generative) linguistics can be taken as "the study of the computation in language" in Marr's sense of "computation" (e.g., Marr, 1982): an explicit description of the engineering problem to be solved by the cognitive system. Computation is key

because the modern cognitive revolution (as it emerged in the second half of the twentieth century) is defined by it. Without computation, the behaviorists had every right to assert that the black-box nature of the mind prevents scientific study. With it, a tool becomes available for linking the observable world to the hidden contents of the black box. Computation is the "mind-stuff" out of which we can construct inference chains to probe as deeply into the mind as we care to go.

I'll call a cognitive inference chain of this sort a linking model. A particularly elegant type is the additive model of Sternberg (1998), with which one can make justifiable inferences about the order, duration, and modularity of mental processes simply by measuring reaction times in different behavioral tasks. (Of course, there's no reason to share Sternberg's assumption that linking models must be linear; see Westbury & Hollis, this volume.) An independently justified linking model gives us just as much confidence about the reality of hidden mental entities as we have about any other natural phenomena that can only be observed indirectly.

Linking models need not be fully explicit in order to be useful. The usual strategy in cognitive psychology is to describe whatever is systematic in behavioral evidence, and leave the rest for later. This strategy is necessitated by the complexity of behavior, and it is made rigorous by statistics, another essential tool of the modern cognitive sciences (e.g., in Sternberg's additive model, modularity is ruled out by a significant interaction). Despite its empiricist connotations, a statistical linking model still permits the rationalist view of grammar as static knowledge, because inferential statistics boils down to correlation, and correlation, as every statistics student knows, isn't the same as causation. Thus there is no obligation for a linking model to treat grammar as having "real time" influences on behavior. That is, it doesn't have to be an algorithm, the next level in Marr's hierarchy, as long as some aspects of behavior are predicted by it.

Therefore, generative linguistics is psycholinguistics. At least it is in principle. Due to the rationalist nature of their culture, linguists generally fail to take seriously the empirical challenge of testing mental claims with behavioral data. On the one hand, linguists often treat performance as transparent reflections of competence; even Penke and Rosenbach (2004, p. 492) include "grammaticality judgments" (rather than acceptability judgments, i.e., performance data) on their list of "direct evidence" (as if metalinguistic musing is what human grammar is for). On the other hand, when linguists do invoke performance, they use it primarily as a dumping ground for counterevidence to competence claims, though this move can be valid if the performance-based explanation is independently motivated (see 2.2).

The failure to respect competence-performance linking models has widespread consequences in generative linguistics. Schütze (1996) surveys the serious difficulties with interpreting informally collected syntactic judgments, and Cowart (1997) shows how proper experimental protocols resolve them. Ohala (1986) diagnoses and advises on similar problems in phonology. This chapter attempts to perform a similar methodological service for the study of morphological competence.

2.2. Morphological competence

What makes a morphological theory generative? If we take "generative" as a synonym for "computational" in Marr's sense, and then add the rationalist preference for simplicity, the question splits into two: What is the design problem that morphology exists to solve, and what is the simplest way to describe it?

The obvious answer to the first question is compositionality: The central morphological fact is that words show sound-meaning regularities. Compositionality demands analysis at the

computational level whether decomposition is obligatory at the algorithmic level (see Marantz, this volume) or merely simulated (see Baayen, this volume). If we then consider the second question, the "most generative" answer must be a theory that is maximally compositional.

Of course, several morphologists in the generative tradition have given empirical arguments against extreme compositionality. Aronoff (1976), for instance, points out that speakers know things about words that they cannot have derived from morphological structure alone; thus in English, *transmission* can refer to a specific bit of automotive hardware, not merely the act of transmitting. Extending the argument, Becker (1993) notes that not only are the meanings of *seaman* and *airman* not derived transparently from those of *sea*, *air*, and *man*, but they also share a complex of other notions, such as "travel": Word-level semantic idiosyncrasies can be productive (a point reiterated by Aronoff, this volume).

Yet such arguments do not really falsify extreme compositionality at the computational level. Grammar needn't be given the power to memorize arbitrary meanings for morphologically complex words, since this ability is already sufficiently well explained by the extra-grammatical fact that brains are voracious memorizers: If *transmission* is treated as a unit at any stage of processing (as what Packard, 2000, calls a psycholinguistic word), it is a potential target for semantic barnacles. Not even syntactic phrases are immune, as idioms demonstrate. Similarly, we need not ascribe the productivity of the *seaman* schema to grammar if we admit that humans also possess the extra-grammatical ability to form analogies (as discussed further in 2.3).

For reasons like these, Di Sciullo & Williams (1987) dismiss word-level phenomena as mere performance effects; they call memorized forms "listemes" to give a name to what they don't care about. This may sound radical to a psycholinguist, but it is justifiable: A linguistic form is not automatically "grammatical" merely by virtue of existing, since it may be an exception, and an unattested form isn't automatically "ungrammatical," since it may be an accidental gap.

Yet the theory of Di Sciullo & Williams (1987), like those of Aronoff (1976), Anderson (1992), and most other "generative" morphologists, is still not positioned at the extreme of the compositionality scale, since it assumes that word-internal and word-external composition are fundamentally distinct. Much closer to the extreme is Distributed Morphology (DM; Halle & Marantz, 1993; Harley & Noyer, 2003), so named because it distributes the jobs traditionally given to the "word level" among different components of the mind. Word-level semantic idiosyncrasies (like the special meaning of *transmission*) are extracted from grammar and placed in what DM calls the Encyclopedia. Underlying phonological forms and morphosyntactic conditioning environments are bundled into what DM calls Vocabulary Items. These are then combined with universal morphosyntactic features by essentially syntactic operations, and have their forms readjusted and otherwise transformed by essentially phonological operations. In terms introduced by Harley & Noyer (2000), "f-morphemes" are the functor-like morphosyntactic features, like PLURAL, that license "l-morphemes" like DOG (the parallels, if any, between the l- vs. f-morpheme dichotomy and the traditional root-affix dichotomy are the focus of section 3).

Readers unfamiliar with DM but who know a little generative history may wonder how it deals with Chomsky (1970). This is the paper that introduced the lexicalist hypothesis, which is what led virtually every generative morphologist for over twenty years to assume a strict division of labor between morphology and syntax. Moreover, linguistic folklore credits Chomsky (1970) with killing the theory of Generative Semantics, which, like DM, brought syntax down inside words and morphemes. Marantz (1997) diffuses the apparent challenge to DM very simply: He

shows that Chomsky (1970) has exactly the *opposite* implications from those generally ascribed to it, and in fact its fatal blow was aimed at the lexicalist hypothesis itself. Chomsky was not criticizing syntactic compositionality below the word level, since that's precisely what he means by the "lexicalist hypothesis." Rather, he was only criticizing the use of Generative-Semantics-style transformations to convert forms like *destroy the city* into forms like *destruction of the city*. Though not transformationally related, such forms, Chomsky argues, do share syntactically parallel structures. Crucially (as Marantz emphasizes), in order for Chomsky's argument to go through, the root DESTROY shared by *destroy* and *destruction* must be visible to the syntax, thereby falsifying the "lexicalist hypothesis" as it is usually (mis)understood.

Chomsky supports this claim with, among other examples, the contrast between *destroy/destruction* and *grow/growth*. In DM terms, the l-morpheme DESTROY, whether inside a noun or a verb, expects to go with a patient, and that's why both (1a) and (1b) below are acceptable. By contrast, the l-morpheme GROW describes an internally caused change; its default verbal use is intransitive, as in (2a). A transitive usage like (2b) only works with a causative meaning, represented in the syntax by the silent f-morpheme CAUSE, which is only permitted in verb phrases. The result is that the noun *growth* cannot have a causative implication: (3a) is acceptable, but not (3b) (the star indicates unacceptability, not necessarily ungrammaticality). Since the syntactic difference between the pairs *destroy/destruction* and *grow/growth* is predictable from the semantic difference between the l-morphemes DESTROY and GROW, these roots must be visible to syntax.

(1) a. the enemy destroyed the city
    b. the enemy's destruction of the city

(2) a. the tomatoes grew
    b. John grew the tomatoes

(3) a. the growth of tomatoes
    b. *John's growth of tomatoes

Importantly in the context of this chapter, Harley & Noyer (2000, pp. 364-366) later invoked the competence-performance distinction to deal with a gap in the DESTROY/GROW argument. The problem involves l-morphemes like SEPARATE, which describe an internally caused change, as in (4a), yet unlike GROW do permit causative noun phrases, as in (4b).

(4) a. Jim and Tammy Faye separated
    b. the teacher's separation of the children

Harley and Noyer argue that such observations do not undermine Chomsky's argument because knowledge of the difference between GROW and SEPARATE belongs in the grammar-external Encyclopedia. They do not consider this performance-based explanation as special pleading because they provide independent evidence for it: The acceptability of forms like (4b) depends on how causer-like the subject is perceived as being, which in turn depends on real-world knowledge. Thus some speakers balk at examples like (5).

(5) the Cold War's separation of E. and W. Germany

As Harley and Noyer admit, however, not all speakers reject examples like (5). They therefore predict that cross-speaker judgment variation must derive from cross-speaker variation in thresholds for the perception of causer-hood. Unfortunately this prediction remains rather speculative at this point, since they don't explain how it could be tested against the alternative hypothesis that Chomsky's (1970) argument is essentially mistaken.

Given its extreme compositionality, DM is quite "generative" indeed, and its use of performance as a dumping ground for apparent counterevidence (word-level idiosyncrasies, GROW vs. SEPARATE) is also quite typical of argumentation in the generative tradition. Before we can consider the validity of this latter move, however, we need to take a closer look at the nature of morphological performance.

2.3. Morphological performance

Syntax tends to stick with judgment data, and phonology relies primarily on corpus (dictionary) data, but morphology, the middle child, uses both types, in roughly equal proportions. There are good reasons for this.

Morphology is syntax-like in dealing with rather productive forms. The more productive a system is, the less likely one is to find theoretically relevant examples in a random sample; this observation is at the core of Chomsky's (1957) argument against corpus-based syntactic analysis. Productivity in (lexical) phonology is limited by the small number of phonological units, so it is relatively easy for phonologists to distinguish systematic from accidental gaps, even in small corpora, but morphologists and syntacticians aren't so lucky. This is why syntacticians rely on judgments, which can be elicited by the experimenter in unlimited supply. As we saw above, judgments also play an important role in DM argumentation.

At the same time, however, morphology is like (lexical) phonology in dealing with word knowledge; syntactic productivity must be greater than morphological productivity because sentences are made of words (even in polysynthetic languages, where morphological productivity reaches syntactic levels). Thus morphology is actually in an intermediate position between phonology and syntax with regard to the usefulness of corpora, which is why dictionaries have been a major data source in generative morphology since the beginning. Some generative morphologists have also begun to use corpora of fluent discourse, most often as a tool for finding examples that sound right but are hard to think up spontaneously; Marantz (2005) applies this method to argumentation within a DM framework.

Judgments and corpora provide complementary performance windows into competence: Judgments reflect both grammar and comprehension processes, while corpora reflect both grammar and production processes. Studying both thus makes it easier to factor out competence from performance. Moreover, the proper interpretation of judgments about listeme-like forms depends crucially on understanding corpus patterns. The reason is analogy. Most linguists assume they can eliminate lexical effects on judgments simply by testing novel forms that could not be memorized, but this doesn't block speakers from analogizing, that is, generalizing superficially (extra-grammatically) from memorized exemplars. Even with morphosyntactic judgments, Chomsky (1970, pp. 27-9) shows how speakers can accept ungrammatical forms "by analogy" (his term) with similar-sounding grammatical forms. Since both grammar and analogy involve generalizing across memorized tokens (the former via biologically-constrained

acquisition processes, the latter haphazardly on the fly), judgments of novel forms are, to some extent, merely a sort of intuitive corpus analysis.

For (lexical) phonological judgments, the influence of analogy has been precisely quantified. Bailey & Hahn (2001) had English speakers make wordlikeness judgments on nonword syllables, and calculated the proportions of judgment variation accounted for by phonotactic probability (e.g., the likelihood of /v/ being word-final) and by neighborhood density (e.g., the number of real words differing from the target in at most one phoneme). Both variables had significant effects on judgments (Myers & Tsay, 2005, replicated this in Mandarin Chinese). Since phonotactic effects in spoken word recognition are apparently more "prelexical," representing listener expectations, and neighborhood effects more "postlexical," involving comparison with lexical neighbors (Vitevitch & Luce, 1999), the latter are arguably more like analogy than the former.

The good news from this research is that (phonological) judgment patterns are not solely analogy. The bad news is that what's left after analogy is extracted (including phonotactic probability) cannot really be identified with "grammar." Phonotactic probability and neighborhood density are highly correlated, and each may be defined in many different ways, making it difficult to figure out where the line should be drawn. Moreover, reaction-time measures reveal that they interact with each other (Luce & Large, 2001), so according to the additive model (Sternberg, 1998) they are not associated with distinct processing modules. Similarly, the "phonotactic" MEG component identified by Pylkkänen, Stringfellow, & Marantz (2002) is also sensitive to lexical frequency (Embick, Hackl, Schaeffer, Kelepir, & Marantz, 2001). There thus may well be no fully "prelexical" component in lexical access at all, suggesting that grammar, which is supposed to be independent of the lexicon, plays no real-time role in the processing of lexical phonology (though this does not necessarily rule out an essential role for it at the computational level).

Note how this discussion of the place of traditional linguistic data sources in competence-performance models has led naturally to data sources that are traditionally "psycholinguistic." This is a simple consequence of the principle that answering new types of questions requires collecting new types of data. Unfortunately, this principle has had very little influence on linguistic practice, where data always involve acceptability judgments or corpora of productions, never any other type of linguistic behavior. Myers (2006b) illustrates this by comparing the histories of generative research on lexical phonology and psycholinguistic research on spoken word production between 1970 and 2000. The psycholinguists began by studying only corpus data (natural speech errors), then added experimentally-induced speech errors, then other types of word-production experiments, then brain-imaging studies, along the way sharpening theories by implementing them on computers. During the same period, however, the vast majority of generative linguists made no changes in their data sources at all. This self-imposed limitation is deeply strange: Why should the science of competence be defined in terms of its preferred performance types?

Fortunately, some linguists do occasionally take the extra effort to apply new data sources to grammatical questions. A case in point is Stockall & Marantz (2006), who test deep compositionality (which happens to be a central claim of DM) at Marr's most fundamental level, that of neural implementation, by looking for MEG evidence of root priming in irregularly inflected English verbs. If the test had utterly failed (it didn't), this wouldn't have falsified DM since there are many ways to implement a computational goal, and a successful test doesn't automatically "prove" DM, not only because DM makes many other claims, but more

importantly because linking performance data of any kind to competence claims requires numerous auxiliary assumptions that must eventually be tested themselves. Nevertheless, studies like this do add to our understanding of grammar because they falsify alternative hypotheses that traditional data sources cannot. In particular, as we have seen, it is difficult to rule out an analogical analysis of word-internal patterns solely from corpus data (which are memorized) or judgments (which are partly influenced by analogy). Yet if we are safe in assuming that analogical effects are relatively slow (though perhaps not literally "postlexical"), then a root priming effect in an early MEG component cannot be dismissed as mere analogy. (Another application of "psycholinguistic" data to DM claims is Barner & Bale, 2002.)

In short, there are many sources of performance data bearing on competence claims, the ones traditionally used in linguistics can be useful, the same holds for non-traditional ones, and most importantly, every data source is associated with a different competence-performance linking model that must be understood, at least in rough outline, before data can be interpreted properly.

3. Case study: Affixation vs. compounding in Chinese

In keeping with the back-to-basics approach of this chapter, I've chosen to illustrate morphological competence-performance linking models with a fundamental yet understudied issue: Does grammar really make a distinction between affixation and compounding? Rather than testing it empirically, this distinction is simply assumed by all of the morphological theories I am aware of: Affixes are assumed to be functor-like, whereas roots, the concatenation of which defines compounding, are assumed to be argument-like. In DM this contrast, instantiated in terms of f-morpheme and l-morphemes, is absolutely central to its syntactic approach.

Inconveniently, however, languages don't always draw a very sharp boundary between affixes and roots, and Chinese, as I will show, seems to draw no real boundary at all. Does this mean that something is fundamentally wrong with theories of morphological competence, DM in particular? Or is it another one of those cases where the competence-performance distinction can (validly) come to the rescue?

Obviously I can't really resolve such enormous issues in the space available, and anyway, the point is the journey, not the destination. What I do instead is sketch out the basic facts in Chinese, list some options available in the DM framework, and then see what of relevance can be extracted from data sources of various types.

3.1. The basic facts

Chinese is the prototypical compounding language, and some readers may be surprised to learn that it has anything affix-like at all, but it does (Li & Thompson, 1981; Packard, 2000). One of the clearest examples of a (Mandarin) Chinese affix is *zi* (here and elsewhere transcriptions are in Hanyu Pinyin), found in nouns like *háizi* (child), *chēzi* (car), and *shuāzi* (brush). This morpheme is affix-like because it can never be used in isolation (though the character used to write it is also used for the root morpheme *zǐ* "child"), it is phonologically reduced (toneless, indicated in Pinyin by the lack of a tone diacritic), it always appears in a fixed location (immediately after a root morpheme), it indicates the syntactic category of noun (even if the base morpheme is a verb, like *shuā* "to brush"), and it is semantically "bleached" (in fact it seems to mean nothing).

Chinese has been argued to have several other affixes as well. A few could be called inflectional, like the human plural suffix *men* (*háizimen* "children") and verbal aspect markers like *zhe* (*xuézhe* "studying"). Others seem derivational, like the human nominal suffix *zhě* (*xuézhě* "scholar"). The rest are native roots that are also used to fill in for the affixes in loan translations; an example we meet again below is *jiā* ("expert"), used to translate *-ist* (*xīnlǐxuéjiā* "psychologist," literally "heart reason study expert").

Yet something is not quite right with this picture. First, unlike English, the great majority of morphologically complex words in Chinese are composed entirely out of roots. Perhaps that's why the affix list has to be padded out with loan translations. Should "affixes" like *jiā* be on the list at all? Morphological status in the donor languages is synchronically irrelevant. Why not consider them to be roots that have acquired somewhat bleached semantics?

Second, boundedness cannot be used as a reliable diagnostic for affixhood in Chinese, which has a large number of bound roots. Bound-root compounds in Western languages tend to be opaque curiosities (e.g., *helico* + *pter*), but Chinese has many fully transparent examples like *xiàozhǎng* "school principal," where both *xiào* (school) and *zhǎng* (chief) are bound (as free words, "school" and "chief" are respectively *xuéxiào* "study school" and *shǒuzhǎng* "head chief").

Third, it's not clear if all of the "affixes" are really word bound. Chinese orthography doesn't help clarify the issue, since it fails to mark word boundaries at all. The aspect markers seem more like clitics (though clitic groups can be treated as listemes by Chinese readers; Myers, Huang, & Wang, 2006). Even the "derivational" human marker *zhě* can be attached to things that look a lot like phrases (e.g., *wèi liáng tǐwēn zhě* "people who have not measured their body temperature," ubiquitous on public signs in Taiwan during the SARS scare).

Finally and most importantly, Chinese "affixes" don't really behave like functors. Even noun-final *zi*, the most affix-like morpheme, is not really a nominalizer, since it's most often used with roots that are already nouns (*chēzi* "car"). Pairs like *shuā* (to brush) - *shuāzi* (a brush) are misleading because conversion in Chinese requires no (overt) affixation; *shuā* itself can be a noun (Tai, 1997). A related observation is that derivational "affixation" in Chinese can't feed output into other "affixation" processes (Packard, 2000), quite unlike languages like English (cf. *nationalization*).

Though Chinese blurs the affix-root distinction to an extreme, all languages have their controversial cases. In English we have *able* doing double duty as suffix (*lovable*) and root word (*able to love*), *ful/full* (why would a true affix permit *spoonsful*?), *tele* (prefix or bound root?), and so on. The question is whether this sort of fuzziness in the lexicon should be a fatal problem for theories like DM that assume a sharp distinction in the grammar.

3.2. What would DM do?

Even separate from the affix-root question, root compounding has attracted far less attention in the generative literature than affixation (Aronoff, 1976, doesn't discuss compounding at all), and DM is no exception. Xue (2001) uses DM to argue that whether a string of Chinese roots is a word or a phrase is contextually conditioned, but his focus isn't on compounding per se. Harley (2004) restricts her attention to compounds like X + verb + *er/ing* (e.g., *scriptwriter*, *script-writing*), which, conveniently for DM's syntactic approach, contain overt affixes. About affixless compounds DM seems to have had nothing to say so far, not even well-behaved modifier-noun compounds like *teapot* or *bookstore* (I offer a modest proposal of my own in 3.3.3).

Nevertheless, DM does make it clear that the notions of f-morpheme and traditional affix are not identical. In particular, boundedness is morphosyntactically irrelevant; free function words like *the* and *will* are f-morphemes. Some f-morphemes even have competing free and bound Vocabulary Items, like COMPARE in English being realized as *more* or *-er*. Heidi Harley (personal communication, June 5, 2006) believes that *able* may perhaps be amenable to a similar analysis, taking advantage of syntactic devices to flip the argument roles in *able to love* vs. *lovable* (= "able to be loved").

Other properties of f-morphemes do seem to have relatively invariant correlates in performance, however. The simplest of these is semantic bleaching; f-morphemes are assumed to have extremely general, universal semantics. This fits well with one of the arguments in Packard (2000, pp. 72-3) for analyzing *zhě* (which he translates "one who does/is X") as an affix, but *yuán* ("person whose job/position is X") as a root. Since the semantics of the latter entails that of the former, *zhě* has the more basic meaning; it is more "bleached." Semantic bleaching can have many causes other than grammar, a fact that weakens an argument made by Marantz in unpublished work (summarized in Harley & Noyer, 2000, pp. 369-371) for treating suppletive pairs (*go-went*) as competing realizations of semantically bleached f-morphemes (GO); both suppletion and semantic bleaching are correlated with high frequency, so neither has to cause the other. Nevertheless, on average, f-morphemes must still be more bleached than l-morphemes.

To quantify semantic bleaching in Chinese morphemes (for experiments described in 3.3.3), Wang & Myers (2004) ran two semantic relatedness judgment pretests (1= very unrelated, 6 = very related), one with morpheme-word pairs, as in (6), and the other with word-word pairs, as in (7). As the mean ratings below indicate, Packard's semantic contrast between *zhě* and *yuán* was borne out.

(6) a. *zhě* — *ruòzhě* — 2.0
weak-*zhě*
"weakling"

b. *yuán* — *fǎngyuán* — 3.6
inquire-*yuán*
"reporter"

(7) a. *ruòzhě* *yìzhě* — 1.4
weak-*zhě* translate-*zhě*
"weakling" "translator"

b. *guǎnyuándiànyuán* — 4.0
house-*yuán* store-*yuán*
"librarian" "salesclerk"
(*guǎn* = *túshūguǎn* "library")

As Packard (2000) points out, *zhě* has a second property consistent with affixhood: It is functor-like (converting "X" into "one who does/is X"). Like bleaching, functorhood is partly in the eye of the beholder, but as functors, f-morphemes have two very specific properties. First, there are selectional restrictions on what they may license; this is why *growth* can't have a causative implication, because CAUSE is restricted to verbal environments. Second, unlike

l-morphemes, the Vocabulary Items of f-morphemes are in competition with each other for the same syntactic slots (Halle & Marantz, 1993). Since each f-morpheme defines its own syntactic frame (similar to Construction Grammar; Goldberg, 1995), different f-morphemes themselves compete as well.

In the following exploration, then, I take surface evidence of semantic bleaching, (morphosyntactic) selectional restrictions, and competition as indicative of underlying f-hood.

3.3. Three kinds of evidence

This section examines three distinct data sources for evidence of f-hood. Two would be considered traditionally "linguistic," since they involve, respectively, native-speaker judgments and corpus data. The third data source is traditionally "psycholinguistic," since it involves reaction times in lexical decision experiments.

3.3.1. Judgments

Searching for a judgment-based diagnostic for f-hood, my first thought was to find a Chinese parallel to the English judgment contrast in (8b,c) claimed by Di Sciullo & Williams (1987, p. 33). These examples were intended to show that affixes like *-er* are transparent to the assignment of semantic roles (*bread* is the patient of *bake*), whereas roots like *man* are not.

(8) a. to bake bread
b. a baker of bread
c. *a bake-man of bread

Unfortunately, this turns out to be yet another example of what happens when linguists neglect careful methodology. The claim involves a relationship between two elements, *-er*/*man* and *bread*, and therefore the judgment experiment should involve two factors, not just *-er* vs. *man*, but also *bread* vs. no *bread*. Adding the [±*bread*] factor results in the two additional forms in (9), where (9b) is clearly worse than (9a).

(9) a. a baker
b. *a bake-man

There is thus a fatal confound in (8), making the pattern there irrelevant. As Spencer (1991, p. 333) points out, (9b) is no fluke; *man* generally doesn't attach to verbs, especially not transitive ones. It's not surprising, then, that my search for a Chinese parallel to this pattern didn't turn up anything useful either.

Then I stumbled upon an interesting empirical claim made by He (2004) regarding the interaction, in *zhě* forms, of the affixation of the plural marker *men* and the order of verb and object. Affixation of *men* is permitted with object-verb strings, as in (10a), but not with verb-object strings, as in (10b). Crucially, both forms are acceptable without *men*, as in (10c-d); thus unlike Di Sciullo & Williams, He provides a complete experimental design.

(10) a. OV-*zhě-men* yáoyán zhìzào zhě men
rumor make *zhě* PLURAL
"rumor-mongers"

b. VO-*zhě-men* *zhìzào yáoyán zhě men

c. OV-*zhě* yáoyán zhìzào zhě
"rumor-monger"

d. VO-*zhě* zhìzào yáoyán zhě
"rumor-monger"

The object-verb *zhě* forms in (10a,c) impress any English-speaking linguist as being strikingly similar to the *scriptwriter*-type compounds analyzed by Harley (2004). Perhaps in (10a,c) *zhě* licenses the adjacent verb, whereas in (10b,d) it cliticizes to an entire phrase, and this structural difference then affects whether or not *men* affixation is permitted (this is He's own analysis, in broad outline).

Note that this interpretation assumes that *zhě* has selectional restrictions: It behaves differently with a verbal vs. phrasal sister in the syntactic tree. Thus if *zhě* is a "true affix" but *yuán* is not, as Packard (2000) argues, we would not expect *yuán* to share this selectional restriction. He's paper is not focused exclusively on *zhě*, and in fact, many of his examples involve root morphemes, including *yuán*. If the judgments are subtle enough, however, He may have simply missed some difference between *zhě* and *yuán*. Indeed, when I informally checked He's paradigm examples with Chinese-speaking colleagues and students, I merely found that some didn't like *men* at all (unlike the plural in English, *men* is used rather sparingly), and others didn't like the verb-object order at all. These two biases are irrelevant to He's generalization, which concerns the interaction between the two factors, but they are so powerful they overwhelm any informal search for such an interaction.

Hence I decided to conduct a more careful judgment experiment. To do so, I took advantage of MiniJudge (Myers, 2006c), a free online software tool for designing, running, and analyzing linguistic judgment experiments in accordance with the general principles advocated in Cowart (1997). MiniJudge makes things easier for the novice experimenter by automating the generation of materials and using powerful statistics (generalized linear mixed effect modeling, a sort of repeated-measures logistic regression; Agresti, Booth, Hobert, & Caffo, 2000) run in the free statistics package R (R Development Core Team, 2005) to extract the maximum amount of information from the minimal amount of data: a small number of speakers making binary good/bad judgments on a small number of items. One of the innovations of MiniJudge is the treatment of the (random) order of items as a continuous variable in the statistics to remove lingering order confounds, and if the experimenter so chooses, interactions with order can be factored out as well; among other things, this helps reduce the influence of cross-item analogizing.

The materials were designed primarily with He's two binary factors in mind ([±VO] and [±*men*]), but the head morphemes also varied across the three types shown in (11). Surveys with 48 randomly ordered items modeled on those in (10) were distributed by email to students and faculty in my linguistics department (in southern Taiwan) who didn't know the purpose of the experiment. I received 18 completed surveys.

(11) a. semantically bleached, bound: *zhě*, *jiā* (-ist)
b. semantically rich, bound: *yuán*, *shī* (teacher)
c. semantically rich, free: *rén* (person)

The default MiniJudge analysis revealed significant main effects of [VO] (verb-object strings lowered acceptability) and [*men*] (*men* suffixation lowered acceptability), reconfirming the irrelevant biases I had already noted in the informal judgments. As in the informal test, the [VO] × [*men*] interaction predicted by He was not detected. However, there was also a main effect of order, indicating a (performance-based) change in judgment strategy over the course of the experiment. Hence I used the option in MiniJudge to test for order interactions, and this revealed a significant interaction between order and [VO] × [*men*]. More importantly, factoring out this three-way interaction allowed the previously hidden [VO] × [*men*] interaction to emerge as a significant effect itself. He's generalization was thus confirmed: For [+VO] forms, [-*men*] forms were judged better than [+*men*] forms almost four-to-one (10d vs. 10b), about twice as high as the ratio for [-VO] forms (10c vs. 10a).

The patterns associated with the three-way contrast among morpheme heads (tested directly in R) were more complex. The simplest prediction is that He's generalization should only hold for "true affixes" like *zhě* and *jiā*. Thus if we look only at bound morphemes, there should be an interaction between semantic bleaching (e.g., *zhě* vs. *yuán*) and the [VO] × [*men*] interaction. Unfortunately no such interaction was found, no matter how I played with the data.

Another interesting pattern did emerge, however: Speakers preferred the head morpheme to be as affixlike as possible (*zhě* and *jiā* were better than *yuán* and *shī*, which were in turn better than *rén*). Unfortunately this cannot be taken as evidence for f-morphemes, since bound roots were also preferred over free roots; bleaching is a diagnostic for f-hood, but boundedness is not.

Though this judgment experiment confirmed He's generalization, it failed to confirm any special status for "true affixes." As a null result, it can hardly be considered conclusive. Still hoping that my extension of He's generalization might pan out, I made another attempt from a different angle, as described next.

3.3.2. Corpus analysis

Another argument Packard (2000) gives for treating *zhě* as an affix and *yuán* as a root is the much greater productivity of the former, as demonstrated by a dictionary analysis. The notion of productivity is inextricably linked with generativity, not only because of the similarity of the concepts "produce" and "generate" (about which Chomsky, 1965, p. 9, expresses some regret), but also because productivity indicates that speakers are going beyond rote memory (though they may do this by analogy). In corpus linguistics, however, it's usually applied to comparisons among morphemes of the same type, typically affixes (e.g., Baayen & Renouf, 1996). When comparing affixes and roots, other factors complicate the interpretation of corpus data. In particular, the semantically richer a morpheme is, the more specific its information content, and thus the more restricted its usefulness in discourse; productivity may diagnose semantic bleaching rather than affixhood directly.

Nevertheless, as long as we maintain respect for the extra-grammatical forces influencing corpus frequencies, it seems reasonable to assume that grammatical status should have some measurable affect on the rate at which linguistic forms are coined and used. I decided to apply this logic in a second test of the generalization of He (2004) and my extension of it.

My data source was the ten million word Academia Sinica Balanced Corpus of Modern Chinese (Chen, Huang, Chang, & Hsu, 1996), which is parsed by words (no mean feat given the lack of word boundaries in Chinese orthography) and tagged for syntactic category. I searched for examples of each of the twenty classes implied by the design of the judgment experiment ([±VO], [+*men*], *zhě*/*jiā*/*yuán*/*shī*/*rén*), ensuring that I only chose transitive verbs and common nouns that were both disyllabic (He, 2004, observes that different principles apply with monosyllabic roots). I then tabulated type frequencies as a measure of coinage rate.

The type frequencies shown in (12) yet again revealed clear biases against *men* (5 types with *men* vs. 927 without) and against verb-object strings (214 verb-object types vs. 718 object-verb types). Unfortunately, there wasn't even the slightest hint of an interaction; He's generalization was not confirmed. There's thus no point looking for further interactions with affixhood.

(12) a. [+VO, +*men*] 0
b. [-VO, +*men*] 5
c. [+VO, -*men*] 214
d. [-VO, -*men*] 713

The preference found in the judgments for more affixlike morphemes was only partially replicated, as suggested by the ranking of type frequencies by head morpheme shown in (13).

(13) a. *zhě* 766
b. *rén* 101
c. *jiā* 49
d. *shī* 5
e. *yuán* 1

The surprisingly high type frequency of *rén* forms may simply be due to its own high token frequency, which is far higher than that of the other five morphemes combined, including *zhě*, which has the second-lowest token frequency after *yuán*. Setting *rén* aside, the corpus results actually provide better evidence than the judgments for the special status of semantic bleaching (as opposed to boundedness), since the "true affixes" *zhě* and *jiā* are preferred in coinages far more than the "roots" *shī* and *yuán*. With such a small sample of types, however, this could merely be a statistical accident.

Thus the little that came out of the corpus analyses ended up conflicting, partly, with patterns detected in the judgments. Since each data source has its pros and cons, we can't take one as inherently more reliable than the other. In any case, my extension of He's generalization did not prove very useful as an f-hood diagnostic.

3.3.3. Context and morpheme frequency effects on lexical decision

The final data source I consider here is a set of reaction-time experiments reported in Wang & Myers (2004). Our original motivations were traditionally "psycholinguistic," in that we were testing hypotheses about real-time lexical access, not grammatical knowledge. Nevertheless, as we will see, even traditionally "psycholinguistic" experiments must be interpreted in the context of competence-performance models, and doing so may even reveal new insights into competence itself.

Our experiments were based on Andrews (1986), who used morpheme frequency effects as a diagnostic for morphological decomposition in visual lexical decision tasks in English. Such studies generally find that decisions about complex words are faster if the words contain higher-frequency morphemes, even with whole-word frequency matched. Since frequency effects indicate lexical access, such results suggest decomposition of complex words into morphemes. However, Andrews suspected that decomposition is not automatic, but depends on the contextual influence of inherently more parsable forms on the access of inherently less parsable forms. Consistent with her suspicions, morpheme frequency effects were found with compounds presented alone (inherently easy to parse), but not with suffixed words presented alone (inherently hard to parse), and when she mixed both word types together, both showed morpheme frequency effects (context effects).

Our interest in these experiments had nothing to do with decomposition, automatic or otherwise; in Chinese there's nothing to debate, given the morpheme-based nature of its orthography (Myers, 2006a). We were instead attracted by the possibility that the Andrews paradigm could provide a diagnostic for distinguishing between affixation and compounding in Chinese. We thus ran one experiment with "suffixed" words only (i.e., bimorphemic words ending with semantically bleached morphemes according to our pretests), one with compound words only, and one with mixed morphological types. The results replicated the English results almost exactly, suggesting that at least according to this diagnostic, Chinese has "affixes" just as English does.

Yet in the context of this chapter it's important to consider what sort of competence-performance linking model is presumed by the Andrews diagnostic. Clearly the effect of mixing word types must be a performance effect; grammar can't change itself depending on the processing context. What about the finding that it's inherently easier to parse roots out of compounds than out of suffixed words? Could we posit that the licensing relationship between the affix (f-morpheme) and root (l-morpheme) forces the parser to treat the root-affix string as a whole listeme? Perhaps not; Taft & Forster (1975) and follow-up work found that roots are as easy to parse out of prefixed words as out of compounds. Aren't prefixes just as much f-morphemes as suffixes are?

My current suspicion is that the Wang & Myers (2004) results don't bear on the f-hood issue at all, but actually reveal nothing more than a secondary performance difference. The asymmetry between prefixed words and compounds on the one hand, and suffixed words on the other, can be at least partly explained as arising from the directionality of time itself, which causes a left-to-right bias in lexical access generally, even in monomorphemic words (Marslen-Wilson, 1987).

A more productive research strategy might be to focus on affix competition in primed lexical decision experiments. The basic insight derives from the twin observations that root priming is robust when both prime and target are compounds (*teacup* - *teapot*; Zhou & Marslen-Wilson, 2000) but not when prime and target are suffixed words (*confession* - *confessor*). The latter observation comes from an English study by Marslen-Wilson, Tyler, Waksler, & Older (1994), and they explain it as resulting from competition of different suffixes for the same root. Recall from 3.2 that in DM, f-morphemes define competing syntactic frames; thus competing is just what we would expect suffixes to do. The most obvious wrinkle with using competition as an f-hood diagnostic is that yet again, prefixes behave differently; Marslen-Wilson et al. found that prefixed words do prime each other (*unfasten* - *refasten*).

This time, however, ironing out such wrinkles may be worth the effort, since the competition notion also fits quite neatly with recent research on relation priming in modifier-noun compounds (see Gagné & Spalding, 2004, for a review). For example, *student vote* (head BY modifier) is primed by *student accusation* (head BY modifier) relative to *student car* (head FOR modifier). What's particularly interesting about this research is that it points towards a possible DM analysis of root compounding: Perhaps the psychologically active compound-internal relation is expressed by a covert f-morpheme, similar to the overt affix *-er* in *scriptwriter*. Thus the compounds in (14a) would have the abstract structures in (14b).

(14) a. student vote
student accusation
student car

b. student-BY vote
student-BY accusation
student-FOR car

In favor of this hypothesis, note that the relations used in compounds are semantically bleached and universal: Li & Thompson (1981) list the same relations for Chinese compounds as those found in English, and Ji & Gagné (2004) have replicated the relation priming effect in Chinese compounds. Moreover, priming of *student vote* by *student accusation* relative to *student car* means that the BY and FOR relations compete (indeed, Gagné calls her theory Competition Among Relations in Nominals), just like the competition between overt suffixes found by Marslen-Wilson et al. This is why I treat BY and FOR as suffixes in (14b), on the assumption that the prefix/suffix priming asymmetry has a competence-based rather than performance-based explanation.

This assumption is further supported by an asymmetry first discovered by Gagné (2001), whereby relation competition is found if compound and target compounds have the same modifier, as in (15a), but not if they have the same head, as in (15b) (Gagné, 2002, found that this competition also occurs with semantically related modifiers, e.g., *scholar* and *student*).

(15) a. student-BY vote student-FOR car
b. student-BY vote reform-FOR vote

Gagné, Figueredo, & Spalding (2006) have recently confirmed the positional restriction on relation priming by showing that it does not occur when compounds merely share roots and relations; thus there is no priming of (16a) by (16b) relative to (16c).

(16) a. reading-FOR lamp
b. lamp-FOR shade
c. lamp-PRODUCED_BY light

This pattern of results makes sense if the covert relation-marking f-morpheme in modifier-head compounds is indeed suffixed to the modifier, so that the f-morphemes in (15a) compete for l-morphemes, just as *-ion* and *-or* compete in *confession-confessor*. If this interpretation of relation priming holds up, it would represent a striking example of how the

study of performance, even when conducted solely for its own sake, can provide useful insights into the nature of competence. Indeed, it is not obvious how silent f-morphemes like BY or FOR, or their precise position inside a compound, could ever be established except via traditionally "psycholinguistic" evidence like priming.

What does all this mean for Chinese? Though Ji & Gagné (2004) found relation priming in Chinese compounds, they failed to find any modifier-head asymmetry. This failure is consistent with those noted in 3.1, 3.3.1, and 3.3.2, suggesting that Chinese word formation may not, in fact, require the use of functor-like f-morphemes (relation priming in Chinese compounds would instead involve some sort of word-level semantic priming). If further exploration continues to uncover nothing but null results like these, Chinese may indeed pose a genuine challenge to theories like DM, if only a typological one.

4. What next?

Generative morphology, like generative linguistics as a whole, is psycholinguistics: The two disciplines do not work on qualitatively different issues or in qualitatively different ways. Humans being what they are, maintaining the division between the cultures of linguistics and psychology assumes an illusory importance that does more harm than good. If only linguists would learn from psychologists about how to collect and analyze data in ways that respect the noisy channel between behavior and mind. If only psychologists would learn from linguists that interpreting performance data always depends on testable competence assumptions.

Interdisciplinary interaction can only be improved through mutual respect and education. This truism applies to morphology as much as to any other aspect of language, but here the immediate future looks brighter. The undeniable importance of rote memory in word knowledge has long obliged generative morphologists to keep abreast of psycholinguistic findings, and (de)composition has long played a central role in the study of lexical access. This very book is testimony to the relative ease with which psychologists and linguists exchange views on the nature of words. Perhaps it isn't overly idealistic to hope that a mutually respectful multiculturalism will someday become the norm in the language sciences as a whole.

Acknowledgements

Portions of the work described in section 3 were supported by National Science Council (Taiwan) grant NSC 94-2411-H-194-018, and MiniJudge is co-copyrighted by National Chung Cheng University. I have benefited from comments from Edith Aldridge, Christina L. Gagné, Heidi Harley, and Niina Zhang, as well as the editors; Ko Yuguang helped with the judgment experiment, and Hong Jiafei helped with the corpus analysis. Any mistakes are all my fault.

- **References**

Agresti, A., Booth, J. G., Hobert, J. P., & Caffo, B. (2000). Random-effects modeling of categorical response data. *Sociological Methodology, 30*, 27-80.

Anderson, S. R. (1992). *A-morphous morphology.* Cambridge: Cambridge University Press.

Andrews, S. (1986). Morphological influence on lexical access: Lexical or nonlexical effects? *Journal of Memory and Language, 25*, 726-740.

Aronoff, M. (1976). *Word formation in generative grammar.* Cambridge, MA: MIT Press.

Aronoff, M. (this volume). Language: Between words and grammar. This volume.

Baayen, R. H. (this volume). Storage and computation in the mental lexicon. This volume.
Baayen, R. H., & Renouf, A. (1996). Chronicaling the *Times*: Productive lexical innovations in an English newspaper. *Language, 72*, 69-96.
Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory & Language, 44*, 569-591.
Barner, D., & Bale, A. (2002). No nouns, no verbs: Psycholinguistic arguments in favor of lexical underspecification. *Lingua, 112*, 771-791.
Becker, T. (1993). Back-formation, cross-formation, and "bracketing paradoxes" in paradigmatic morphology. In G. Booij & J. van Marle (Eds.) *Yearbook of morphology 1993* (pp. 1-25). Dordrecht: Kluwer.
Chen, K.-J., Huang, C.-R., Chang, L.-P., & Hsu, H.-L. (1996). SINICA CORPUS: design methodology for balanced corpora. *Language, Information and Computation, 11*, 167-176.
Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
Chomsky, N. (1970). Remarks on nominalization. Reprinted 1972 in *Studies on semantics in Generative Grammar* (pp. 11-61). The Hague: Mouton. [Pagination in the text refers to the 1972 reprint.]
Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences* 3:1-15, 42-75.
Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row. Reprinted 1990, Cambridge, MA: MIT Press.
Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments*. London: Sage Publications.
Di Sciullo, A. M., & Williams, E. (1987). *On the definition of word*. Cambridge, MA: MIT Press.
Embick, D., Hackl, M., Schaeffer, J., Kelepir, M., & Marantz, A. (2001). A magnetoencephalographic component whose latency reflects lexical frequency. *Cognitive Brain Research, 10* (3), 345-348.
Gagné, C. L. (2001). Relation and lexical priming during the interpretation of noun-noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27* (1), 236-254.
Gagné, C. L. (2002). Lexical and relational influences on the processing of novel compounds. *Brain and Language, 81*, 723–735.
Gagné, C. L., Figueredo, L., & Spalding, T. L. (2006, October). Does snow man prime plastic snow? The effect of position in accessing relational information during conceptual combination. Poster to be presented at the Fifth International Conference on the Mental Lexicon. Montreal, QC.
Gagné, C. L., & Spalding, T. L. (2004). Effect of relation availability on the interpretation and access of familiar noun-noun compounds. *Brain and Language, 90*, 478-486.
Goldberg, A. E. (1995). *A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In Kenneth Hale and Samuel Jay Keyser (Eds.) *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger* (pp. 111-176). MIT Press.
Harley, H. (2004). Merge, conflation, and head movement: The First Sister Principle revisited. In K. Moulton & M. Wolf (Eds.) *Proceedings of Northeast Linguistics Society 34*. Amherst: University of Massachusetts, GLSA.

Harley, H., & Noyer, R. (2000). Formal versus encyclopedic properties of vocabulary: Evidence from nominalisations. In B. Peeters (Ed.) *The lexicon-encyclopedia interface* (pp. 349-374). Oxford: Elsevier.

Harley, H., & Noyer, R. (2003). Distributed morphology. In L. Cheng & R. Sybesma (Eds.) *The second GLOT International state-of-the-article book: The latest in linguistics* (pp. 463-496). Studies in Generative Grammar 61. Berlin: Mouton de Gruyter.

He, Yuanjian (2004). The words-and-rules theory: Evidence from Chinese morphology. *Taiwan Journal of Linguistics, 2* (2), 1-26.

Ji, H., & Gagné, C. L. (2004). Lexical and relational influences on the processing of Chinese modifier-noun compounds. Presented at the Fourth International Conference on the Mental Lexicon, Windsor, Canada, June-July.

Kenstowicz, M., & Kisseberth, C. (1979). *Generative phonology: Description and theory.* New York: Academic Press.

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language & Cognitive Processes, 16* (5/6), 565-581.

Marantz, A. (this volume). The mental lexicon in the decompositional brain. This volume.

Marantz, A. (1997). No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics, Vol. 4.2*, pp. 201-225.

Marantz, A. (2005). *Rederived generalizations*. Handout for talk presented at Tsinghua University, Hsinchu, Taiwan, December.

Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition, 25*, 71-102.

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review, 101* (1), 3-33.

Miller, G. A. (1990). Linguists, psychologists and the cognitive sciences. *Language* 66:317-322.

Myers, J. (2006a). Processing Chinese compounds: A survey of the literature. In G. Libben and G. Jarema (Eds.). *The representation and processing of compound words* (pp. 169-196). Oxford: Oxford University Press.

Myers, J. (2006b). Linguistics as cognitive psychology. Pre-Conference proceedings of the 14th Annual Conference of the International Association of Chinese Linguistics & 10th International Symposium on Chinese Languages and Linguistics Joint Meeting (pp. 150-174). Taipei, Taiwan, May.

Myers, J. (2006c). MiniJudgeJS (Version 0.9.9) [Computer software]. URL: http://www.ccunix.ccu.edu.tw/~lngproc/MiniJudgeJS.htm

Myers, J., & Tsay, J. (2005). The processing of phonological acceptability judgments. *Proceedings of Symposium on 90-92 NSC Projects* (pp. 26-45). Taipei, Taiwan, May.

Myers, J., Huang, Y.-C., & Wang, W. (2006). Frequency effects in the processing of Chinese inflection. *Journal of Memory and Language, 54*, 300-323.

Ohala, J. J. (1986). Consumer's guide to evidence in phonology. *Phonology Yearbook, 3*, 3-26.

Packard, J. L. (2000). *The morphology of Chinese: A linguistic and cognitive approach.* Cambridge, UK: Cambridge University Press.

Penke, M., & Rosenbach, A. (2004). What counts as evidence in linguistics? An introduction. *Studies in Language, 28* (3), 480-526.

Phillips, C., & Lasnik, H. (2003). Linguistics and empirical evidence: Reply to Edelman and Christiansen. *Trends in Cognitive Science, 7* (2), 61-62.

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.

R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Schütze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.

Spencer, A. (1991). *Morphological theory: An introduction to word structure in generative grammar*. Oxford: Basil Blackwell.

Sternberg, S. (1998). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Eds.) *An invitation to cognitive science, vol. 4: Methods, models, and conceptual issues* (pp. 703-863). MIT Press.

Stockall, L., & Marantz, A. (2006). A singe route, full decomposition model of morphological complexity. *The Mental Lexicon, 1* (1), 85-123.

Taft, M., & Forster, K. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14*, 638-647.

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., & Myers, J. (2004). The processing of affixation and compounding in Chinese. Presented at Fourth International Conference on the Mental Lexicon, Windsor, Canada, June-July.

Westbury, C. F., & Hollis, G. (this volume). Putting Humpty together again: Synthetic approaches to nonlinear variable effects underlying lexical access. This volume.

Xue, N. (2001). *Defining and automatically identifying words in Chinese*. Unpublished doctoral thesis, University of Delaware, Newark, DE.

Zhou, X., & Marslen-Wilson, W. (2000). Lexical representation of compound words: Cross-linguistic evidence. *Psychologia, 43* (1), 47-66.