Chinese as a Natural Experiment

James Myers

National Chung Cheng University

Address for correspondence:

James Myers

Graduate Institute of Linguistics

National Chung Cheng University

Min-Hsiung, Chiayi 62102

Taiwan

Phone: 886-5-242-8251

Fax: 886-5-272-1654

Email: Lngmyers@ccu.edu.tw

Chinese as a Natural Experiment

Abstract

The Chinese lexicon is characterized by its typologically unique one-to-one-to-one mapping of morphemes, syllables, and orthographic characters. This architecture poses practical difficulties for the psycholinguist wanting to study lexical processing in Chinese. More seriously, seen as a natural experiment, Chinese challenges assumptions that processing models traditionally make about the roles of phonemes, morphemes, lemmas, and words in lexical access. It is argued that cross-linguistic variation in lexical processing cannot be accommodated by simply modifying lexical processing models, but instead what is needed is a universal learning model. Suggestions are given for how such a model could be tested empirically by extending methods already used for testing language-specific lexical processing.

Keywords: Chinese, lexical processing, phonology, morphology, orthography, cross-language psycholinguistics

Why study lexical processing in Chinese? From a natural history perspective, of course, Chinese has to be studied (along with all other languages) if we want our description of the lexical world to be complete. A more interesting reason, however, arises in the perspective of natural philosophy: Chinese (along with all other languages) can be seen a natural experiment, or more properly, a natural experimental condition, representing a particular combination of independent variables, which can help in the testing of universal models of lexical processing.

The problem with natural experiments is that they are quite badly designed. Chinese differs from English, say, in many different ways, making it impossible to be entirely sure what causes what. In this article I illustrate this by focusing on the core architecture of the Chinese lexical system, namely its one-to-one-to-one mapping of morphemes, syllables, and orthographic characters. Western European languages, the primary inspiration for lexical processing models that strive for universality, do not have lexical architectures like this. Unsurprisingly, then, Western-inspired lexical models face difficulties, both practical and theoretical, when applied to Mandarin (the politically dominant language in the Sinitic language family, and thus the best studied).

The usual way to accommodate such cross-linguistic variation is to tweak the processing model, but tweaks alone cannot create a model capable of fitting languages with radically different lexical architectures. For this and other reasons, I end the article by arguing that what we really need is a universal model of lexical learning, and suggest how such a model could be tested empirically by extending methods already used for testing language-specific lexical processing.

The Architecture of the Chinese Lexicon

To a first degree of approximation, all Chinese morphemes are monosyllabic and written with one character. For example, the sentence 大象喝了蘋果汁 (written in the traditional characters still used in Taiwan and Hong Kong) has seven characters and seven syllables (romanized in Pinyin, with tone digits on stressed syllables, as *da4 xiang4 he1 le ping2 guo3 zhi1*), each a single morpheme (meaning ‘big’, ‘elephant’, ‘drink’, ‘completive aspect’, ‘apple’, ‘fruit’, ‘juice’). Segmented into three words as *da4xiang4 he1-le ping2guo3zhi1*, the sentence means ‘(The) elephant(s) drank apple juice’. Although there are no orthographic indications for word boundaries, there is linguistic and psycholinguistic evidence that Mandarin morphemes are grouped into larger stored units (Myers, 2006; Packard, 2000).

At a finer-grained (i.e., more accurate) level of description, the elegant one-to-one-to-one mapping has numerous exceptions. Of these, the most important are polysyllabic morphemes, written with strings of monosyllabic characters. Sometimes these characters are chosen solely for their sound (e.g., 沙拉 *sha1la1* ‘salad’, from English, literally ‘sand’ + ‘pull’), but sometimes they are coined specially for the polysyllabic morpheme (e.g., 葡萄 *pu2tao2* ‘grape’, an ancient borrowing, written with characters sharing the same radical). Taft and Zhu (1995) call this second type *binding characters*, though this is something of a misnomer, as we will see shortly.

Nevertheless, the vast majority of Mandarin morphemes are indeed monosyllabic.

Moreover, there is no cross-syllable tone spreading, vowel harmony, consonant assimilation, or resyllabification (though clitics may trigger cross-morpheme gemination; Duanmu, 2007; Lin, 2007). Thus 天安門 'Tiananmen' is pronounced *tian1an1men2*, with neatly separated syllables, not resyllabified as [tie.nan.men] as in English. Mandarin syllable structure is also quite simple. The closest it comes to consonant clusters are consonant-glide onsets, as in 天 *tian1* 'heaven', but even these are analyzable as single segments with secondary articulation (Duanmu, 2007). The simple structure generates a tiny syllable inventory. Even taking the four lexical tones into account, there are only around 5,000 logically possible syllables in the phonemic analysis underlying Pinyin, and of these, only around 1,300 are lexically attested, with virtually all of the gaps violating one or more phonotactic constraints (Duanmu, 2007; Lin, 2007). The small syllable inventory results in rampant homophony, with each syllable associated with an average of three characters (out of the 4,000 characters typically claimed to suffice for basic Chinese literacy; Huang & Hanley, 1994). Nevertheless, homophonous characters have remained visually distinct, even while syllable structure simplified diachronically (Baxter, 1992).

Turning to Mandarin morphology, its most striking feature is its blurring of what seem, from a Western perspective, to be fundamental distinctions. Thus despite the evidence for polymorphemic lexical units in Mandarin, the distinction between morphemes and words remains notoriously fuzzy (Myers, 2006; Packard, 2000). One oft-cited example is 擔心 *dan1xin1* 'worry (about)' (Huang, 1984; Zhang, 2007), literally 'carry heart'. This sometimes acts like an exocentric compound, since the internal "object" does not block the taking of a syntactic object, as in 他很擔心你 *ta1 hen3 dan1xin1 ni3* 'He is very worried about you' (literally 'he very *carry-heart* you'). Yet certain syntactic operations can split it up, as in 他擔了三年的心 *ta1 dan1 le san1 nian2 de xin1* 'He was worried for three years' (literally 'he *carry*-aspect three year modifier-marker *heart*'). The notion of affixation is also unclear in Mandarin, so much so that every potential affix is reanalyzable as a bound root, syntactic clitic, or meaningless formative (Myers, 2007).

Finally, consider the polysyllabic morphemes alluded to earlier. These forms may be more consistent with one-to-one morpheme-syllable mapping than first appears, since speakers seem to treat them much the way they treat compounds. Thus the ancient and apparently native (DeFrancis, 1984) disyllabic morpheme 蝴蝶 *hu2die2* 'butterfly' (note the radical shared by the two binding characters) has since been reanalyzed as a right-headed compound, permitting coinages like 彩蝶 *cai3die2* 'colorful butterfly'. A striking example of the ambiguity of the compound/morpheme distinction is the word *jia1ju4* 'furniture', which is written variously as the transparent compound 家具 (literally 'home equipment'), the monomorphemic 傢俱 (note the shared radical), and the mixed versions 傢具 and 家俱. Quantitative evidence of the non-binding nature of binding characters comes from an analysis of the materials used by Hsu and Myers (2009). The mean mutual information (MI) value (a measure of collocation; Church & Hanks, 1990) for a sample of thirty nouns composed of binding characters was significantly higher (M = 15.34, SD = 2.23) than the mean MI for a sample of thirty compound nouns matched in word frequency (M = 7.17, SD = 3.26) ($t(58) = 11.33$, $p < .0001$). Though this is just what we would expect of characters that prefer to bind, note that the standard deviation for binding word MI was far from zero, and in fact there were three apparently genuine compounds with MI values higher than the lowest binding word MI.

These collocation patterns arise in part because Mandarin permits words to be truncated to one character/syllable, then recombined with others (which may themselves have been derived by truncation), regardless of the word's internal structure. For example, the apparently monomorphemic 葡萄 *pu2tao2* 'grape', which we saw earlier, forms an orthographic component of the monomorphemic 葡萄牙 *pu2tao2ya2* 'Portugal' (the third character means 'tooth', but is chosen here solely for its sound). This trisyllabic morpheme can then be truncated to 葡 *pu2* in compounds, as in 葡語 *pu2yu3* 'Portuguese language', where 語 *yu3* itself is truncated from the compound 語言 *yu3yan2* 'language' (composed of two near synonyms). Truncation, like nickname formation (Mester, 1990), is subject to prosodic constraints (Duanmu, 2010); Mandarin words are preferentially disyllabic, making 葡語 *pu2yu3* better formed than 葡萄牙語言 *pu2tao2ya2yu3yan2*.

Packard (2000) observes that when non-binding characters inherit whole-word semantics via truncation, they end up gaining additional meanings. Myers (2006) cites the example of 機場 *ji1chang3*, which does not mean 'machine area' but 'airport', the first character being truncated from the transparent compound 飛機 *fei1ji1* 'airplane' (literally 'fly machine'). Truncation is one reason why most Chinese characters have multiple meanings (Myers, Taft, & Chou, 2007). As in other languages, it is difficult to distinguish homonymy (distinct morphemes sharing one character) from polysemy (one morpheme with multiple senses). The new meaning of 機 *ji1*, for example, may merely be a new sense of the old morpheme.

These orthographic, phonological, and morphological features prove to have serious consequences for how experiments on Mandarin lexical processing are designed and interpreted, a point we turn to next.

Psycholinguistic Consequences

The practical and theoretical implications of Mandarin lexical architecture are surprisingly pervasive. As a practical matter, designing an experiment on lexical processing entails dealing with intrinsic quantitative properties of lexical items, such as lexical frequency, neighborhood density (Luce, Pisoni, & Goldinger, 1990), and morphological family size (Schreuder & Baayen, 1997). Whether they are matched or manipulated in the material design or factored out in the statistics, they first have to be measured. The nature of the Mandarin lexicon is such that it is quite a challenge to do this in a principled way.

This is most obviously the case for word frequency, given how vague the notion of wordhood is in Chinese. For example, the inherent limitations of using internet searches to estimate word frequency (Eu, 2008) become even more serious in Chinese; an exact Chinese character string search could include hits for strings split by an orthographically invisible word boundary. This may be part of the reason why the currently best frequency estimates for disyllabic Mandarin words come from the word-segmented movie subtitle corpus of Cai and Brysbaert (2010); their estimates predict lexical decision times in Myers, Huang, and Wang (2006) and other studies better than any other available Chinese corpus. Of course, their word segmentations had to be imposed on their raw data; other segmentations may prove to work better yet.

The vagueness of the Chinese word infects all lexical measures, but each has additional problems of its own. Neighborhood density, for example, is usually defined as the number of lexical items differing from a target in exactly one phoneme (Luce et al., 1990; cf. Bailey & Hahn, 2001). Even if we set aside the problem of how to decompose a Mandarin syllable into phonemes (recall that onset consonant-glide clusters can be analyzed as single segments), the experimentalist looking for a robust neighborhood effect in Mandarin monosyllables is faced with the relative uniformity of neighborhood densities across phonotactically legal targets, a result of the high proportion of possible legal syllables that are actually attested (Myers & Tsay, 2005).

Defining morphological family size faces a different sort of problem. If we made no distinction between “true” morphemes and “mere” characters, we would instead be studying orthographic families (e.g., Huang, Lee, Tsai, Lee, Hung, & Tzeng, 2006). Yet even if we restrict our attention to transparent compounds, character polysemy/homonymy makes it difficult to define morphological families. Fortunately, progress towards dealing with this problem is being assisted by Chinese WordNet (e.g., Huang & Lee, 2010), a lexical semantics database inspired by the original English version (Fellbaum, 1999).

The theoretical implications of Mandarin lexical architecture are even more serious. Consider just three generalizations derived from research on Western European languages and built into prominent lexical processing models. First, spoken word recognition and word production are both claimed to involve the activation of phonemic units (e.g., McClelland & Elman, 1986, for speech input; Levelt, Roelofs, & Meyer, 1999, for speech output). Second, morphemes are claimed to play an active role in word production (Bien, Levelt, & Baayen, 2005; Roelofs, 1996), even when semantically opaque (Roelofs & Baayen, 2002). Third, again in word production, the semantic and syntactic information encoded in lemmas are claimed to be prepared prior to activation of the form-encoding lexemes (e.g., Levelt et al., 1999). Two of these generalizations appear to be false in Mandarin, while a peculiarity of Mandarin makes it hard at first to see the truth of the third.

Take the units of spoken word processing. Consistent with their syllable-based phonological, morphological, and orthographic systems, Mandarin speakers appear not to depend on phoneme-sized units in either word recognition or word production. Tseng, Huang, and Jeng (1996) found that Mandarin listeners were faster to detect onset consonants in lexically attested syllables (e.g., *bai2* 白 'white') than in lexically unattested syllables (e.g., the accidental gap *bou2*), whether the lexical syllable appeared in a disyllabic word (e.g., *bai2yun2* 白雲 'white cloud') or nonlexical string (e.g., 白力 *bai2li4* 'white force'). Similarly, in word production, O'Seaghdha, Chen, and Chen (2010) showed that English speakers initiate speech faster when trained targets started with the same consonant (e.g., *day* trained with *dew*, *dough*, *dye* vs. with *pea*, *rye*, *sow*), just like Dutch speakers (Meyer, 1991). Yet for Mandarin speakers, no such effect was found, whether targets were monosyllabic or disyllabic.

Thus whether or not a phoneme-processing component is necessary in a universal model of lexical processing, it seems inefficient to hard-wire it in, since Mandarin speakers may never use it. As an example of the sort of model we must reject, consider Newport and Aslin (2004) and Newport, Hauser, Spaepen and Aslin (2004). These studies tested the learning of patterns in CVCVCV structures, in both cotton-top tamarins and humans, the latter all monolingual English speakers. Both species were able to learn patterns involving vowels and adjacent syllables, but only humans could learn patterns with consonants, and only monkeys could do so for patterns with nonadjacent syllables. The authors concluded that humans are innately biased to perceive syllables via their component phonemes. It is certainly conceivable that Mandarin speakers have not achieved their full human potential, but it is also possible that the difference between the monkeys and English speakers may have had something to do with their different pre-experiment experiences.

Turning now to morphology, we might expect that, given the morphemic nature of Chinese characters, Mandarin speakers will show the same robust influence of morphemes in complex word production as has been found in other languages. This expectation is wrong. Chen and Chen (2006, 2007) found no morpheme frequency effect and no additional benefit when the first syllable of training items were not only phonologically identical, but also represented the same morpheme, sharing the same semantics and character (e.g., 家事 *jia1shi4* ‘housework’ in a training set of transparent compounds starting with the same morpheme 家 *jia1* ‘home, family’, compared with a training set of compounds starting with the homophonous but otherwise unrelated morphemes 佳 *jia1* ‘excellent’, 加 *jia1* ‘add’, and 嘉 *jia1* ‘glorious’). Janssen, Bi, and Caramazza (2008) also failed to find morpheme frequency effects for Mandarin compounds in a picture naming task.

Chen and Chen explain the lack of morpheme activation in Mandarin word production as deriving from the lack of resyllabification. Production tasks designed to probe the lexeme stage of word production do not actually detect morphemes as Saussurean signs (i.e., arbitrary form-meaning pairings), but rather only the phonological side-effects of morphological structure, in particular the blocking of resyllabification; Levelt et al. (1999,

p. 26) cite the English compound *popart*, syllabified *pop-art*, not *po-part* as a monomorphemic word would be. Mandarin never resyllabifies, regardless of morphological structure, so lexeme preparation does not require activation of morphemes.

A wrinkle in this picture is the fact that Janssen et al. (2008), using the same task they used for Mandarin, found no morpheme frequency effect in English either. Janssen et al. suggest that the disagreement with the Dutch results of Roelofs (1996) and Bien et al. (2005) may be due to a language or task difference, but English resyllabifies in a very similar way as Dutch, and Chen and Chen (2006), who failed to find morpheme frequency effects in Mandarin, used the same task as Roelofs (1996). Such cross-language/study discrepancies are ubiquitous in psycholinguistics, and at the end of the article I make some suggestions about how to side-step them entirely.

Mandarin idiosyncrasies do not always threaten lexical processing models; sometimes taking them into account allows us to see that the model does not need to be changed at all. For example, Perry and Zhuang (2005) report data claimed to challenge the widely assumed feed-forward model of word production, whereby lemma preparation strictly precedes lexeme preparation (Levelt et al., 1999). Mandarin speakers were shown pictures of objects whose names freely varied between monosyllabic and disyllabic forms, such as ‘elephant’, which can be named as either 大象 *da4xiang4* or 象 *xiang4* (dropping 大 *da4* ‘big’). Speakers were found to be less likely to choose monosyllabic forms if the pictures were shown in lists that also included objects with unambiguously polysyllabic names. The authors concluded that lemma choice can be primed by lexeme form, a reversal of the standard assumed order. However, given that the alternate word forms used in this study always shared a morpheme and whole-word semantics, a simpler interpretation is that the alternate forms actually had the same lemma, and the variation in word length occurred in the lexeme stage, when speakers did or did not apply truncation, under prosodic influence.

The primary lesson of this and the previous section is quite simple: languages are different. Yet, as we discuss next, this bit of banality has profound effects on how lexical models should be built and tested if they are to have serious pretensions of universality.

Considerations in the Building and Testing of Universal Models

The usual approach to cross-linguistic variation is to take a preexisting processing model and add, delete, or rename just enough bits to handle the exotic new data. This approach is taken by O'Seaghdha et al. (2010) to harmonize their findings of the different roles for phonemes versus syllables in English versus Mandarin. Namely, they posit a universal principle of proximate units in word production; in English these are phonemes, while in Mandarin they are syllables.

From a natural philosophy perspective, however, we cannot be satisfied with such model-tweaking. Given the cross-language variation, the Mandarin speaker's apparent neglect of phonemes cannot be innate, so it must be learned from experience, and similarly for all of the other typological curiosities we have been discussing. Immediately we are led to the familiar conclusion that the only truly explanatory model of language is a model of language learning (a position advocated by researchers as different as Chomsky, 1965, and the non-Chomskyan psycholinguists Bates, Devescovi, & Wulfeck, 2001).

Moreover, while a universal learning model must exist, there is no reason to expect that a universal model of mature processing is even possible. Nativist logic applies to babies, not adults, and indeed, (apparent) cross-linguistic incommensurability is easy to find (see Evans & Levinson, 2009, for a recent review). O'Seaghdha et al. (2010) inadvertently illustrate this in two box-and-arrow networks representing the hypothesized word production systems for English and Mandarin (their Figure 1, p. 286). These networks are topologically so different that only if their proximate unit principle is reinterpreted as a component of a learning model, one that generates networks rather than being a component of any mature network itself, can we even see them as consistent with a unified theory.

Two of the coauthors of O'Seaghdha et al. (2010) did take a small step towards a learning model in Chen, Dell, and Chen (2007). They presented translation equivalents of children's stories in both Mandarin and English to a simple recurrent network, training it to predict the next phoneme from the current one (following Dell, Juliano, & Govindjee, 1993). The network's error rate for predicting within-syllable phonemes was much lower in Mandarin than in English, presumably because of the stricter phonotactics of Mandarin. However, this is a only small step towards a learning model because, among other things, it does not model the causal links between phoneme transition probabilities and the choice of proximate units.

More fundamentally, models only generate predictions; to test them, we need evidence. The problem is that when testing universal models, cross-language evidence is a double-edged sword. On the one hand, different languages represent different levels of a factor fixed language-internally (Cutler, 1985); this is the logic of the natural experiment. On the other hand, cross-language comparisons involve confounds of their own. For instance, besides phonotactic probability, English and Mandarin differ in many other ways potentially relevant to their choice of proximate unit, including orthography (English letters are phoneme-sized, Chinese characters are syllable-sized), segmental phonology (English has resyllabification and phonemic alternations, Mandarin does not), and syllable inventory (large and sparse in English, small and dense in Mandarin).

Some confounds can be dealt with by studying child language. For example, Leong (2006) reviews evidence that preliterate Mandarin-speaking children already do worse on phoneme-manipulating tasks than preliterate English-speaking children. However, child studies cannot eliminate all confounds; since Mandarin and English differ in many ways, the learning patterns associated with them differ in many ways as well (see Tardif, 2006, for an attempt to disentangle the disparate variables affecting noun bias vs. verb bias in child English vs. child Chinese).

In the remainder of this article, then, I want to explore a more ambitious approach. The key insight is that the relationship between a universal learning model and a specific lexicon is analogous to that between a language-specific model and a lexical item. In both cases behavior is modeled as a function of lexical input, except that for a learning model, the input is an entire lexicon and the behavior an entire suite of responses. This analogy suggests that universal learning models could be built and tested by extending techniques originally developed for language-specific models.

These techniques are of two general types: factorial experiments and data-mining. A factorial experiment on a learning model involves presenting artificial lexicons that systematically vary in architecture. For example, to test the hypothesis that strict syllable-internal phonotactics discourages phonemic segmentation in word production, we could train English speakers on a set of novel words with high cross-phoneme transitional probabilities, and then test whether consonant priming is reduced in this language, as compared with a control artificial lexicon with lower transitional probabilities.

Artificial lexicon experiments, a species of artificial language experiment (Pothos, 2007), have attracted considerable attention in recent years, though only a subset of studies explicitly focus on the learning of lexical architectures per se. We have already discussed one such study earlier, namely Newport and Aslin (2004). However, recall also that their English-speaking participants extracted phonemes from the artificial words (something Mandarin speakers presumably would not have done), suggesting that their English-honed skills were carried over into the new lexicon. Language transfer can take surprising forms; Ehrich and Meuter (2009) even observed transfer of reading habits, with Mandarin speakers better than English speakers at learning novel words written in an invented logographic (i.e., Chinese-like) orthography. Fortunately, language transfer is not inevitable; Magnuson, Tanenhaus, Aslin, & Dahan (2003) found that the processing of a small artificial lexicon was not influenced by neighbors in the participants' native English lexicon.

Another concern is whether brief laboratory experiments truly engage the same processes involved in natural first language acquisition. Artificial lexicons can begin to influence processing with surprisingly little training, but the effects can disappear again just as quickly (Onishi, Chambers, & Fisher, 2002). There are also hints that learning changes qualitatively over time; Warker and Dell (2006) found that novel complex phonotactic constraints only began to affect speech errors on the second day of training. Even with the ideal training regimen, an artificial lexical experiment can still only aspire to simulate second language learning. Fortunately, second-language lexicons seem not to be learned or processed substantially differently from native lexicons (French & Jacquet, 2004).

An alternative to factorial experiments is data-mining. Psycholinguists use this technique in so-called megastudies, where speaker responses to a large number of lexical items are analyzed post-hoc for statistical regularities (Balota, Yap, Hutchison, & Cortese, in press; Liu, Shu, & Li, 2007). Megastudies obviate the impossible task of squeezing fixed lexical variables into a true factorial design (Baayen, 2010). Data-mining has precisely the same benefit for testing hypotheses about the universal learner, where what is fixed is the set of attested human languages. For example, to test the hypothesized relationship between syllable-internal phonotactics and phoneme activation in production, we could select a hundred or so languages, compute various lexical statistics, find ten or so speakers of each language, run separate picture naming (mega-)experiments on all of the languages, and analyze the naming latencies with the lexical statistics as fixed independent variables and languages (ideally also items and speakers) treated as random. If the hypothesis is correct, we predict that higher by-language mean phoneme transition probability will be associated with reduced priming across words sharing the same initial consonant.

Such a study need not be overly difficult. A surprising degree of linguistic diversity already thrives on the Internet, simplifying the computation of lexical statistics. Linguistic diversity is also the norm in any large, multicultural city, which can supply both participants and consultants to help with material selection. Although in psycholinguistics multi-language studies have generally involved small-scale comparisons (among the most ambitious are Bates et al., 2003; Slobin, 1985-1997), in typological linguistics it is routine to treat language as a random variable (Cysouw, 2005; Haspelmath, Dryer, Gil, & Comrie, 2005). Language megasampling does not require much overhead either; the survey of almost six hundred languages reported in Mielke (2008) was conducted at a single library for a doctoral thesis. It also seems to be an approach whose time has come. Jaeger and Norcliffe (2009) bemoan their discovery that fewer than thirty of the estimated five to ten thousand extant human languages have been studied in sentence processing experiments (the number for lexical processing cannot be much higher); following Hawkins (2007), they call for psycholinguists to work more closely with typologists.

Language megasample studies pose challenges of their own, however. Sibley, Kello, & Seidenberg (2009) show that the high error variance of megastudies can hide effects detectable in smaller controlled studies. Balota et al. (in press) argue that larger samples and standardized scores fix the problem, but cross-language samples can never be as large or homogeneous as within-language samples. Typological research also requires finding an unbiased descriptive terminology (Haspelmath, 2010; Jaeger & Norcliffe, 2009); notions like wordhood are even harder to define cross-linguistically than language-internally. Fortunately, with data-mining, various definitions can be applied post-hoc to see which gives the best overall fit of the data.

Although artificial lexicon and language megasample studies have their limitations, the limitations are complementary: artificial yet well-designed lexicons versus natural yet arbitrary lexicons. Both are also more logistically challenging than traditional studies, but perhaps it is time for linguistics to think in terms of big science, the way physics and biology have learned to do. Ultimately, bold methodological steps seem crucial if we are to respond decisively to the cross-linguistic challenges discussed in this article, so that lexical processing research can become a fully respectable branch of natural philosophy.

References

Baayen, R. H. (2010). A real experiment is a factorial experiment? *The Mental Lexicon, 5* (1), 149-157.

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory & Language, 44*, 569-591.

Balota, D. A., Yap, M. J., Hutchison, K.A., & Cortese, M. J. (in press). Megastudies: Large scale analysis of lexical processes. In James S. Adelman (Ed). *Visual word recognition, Vol. 1: Models and methods, orthography and phonology*. London: Psychology Press.

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Lu, C-C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A., & Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review, 10* (2), 344-380.

Bates, E., Devescovi, A., & Wulfeck B. (2001). Psycholinguistics: A cross-language perspective. *Annual Review of Psychology, 52*, 369-96.

Baxter, W. H. (1992). *A handbook of Old Chinese phonology*. Berlin: Mouton de Gruyter.

Bien, H., Levelt, W. J. M., & Baayen, H. (2005). Frequency effects in compound production. *Proceedings of the National Academy of Sciences, 102*, 17876-17881.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE, 5* (6) e10729: 1-8.

Chen, J.-Y., & Chen, T.-M. (2006). Morphological encoding in the production of compound words in Mandarin Chinese. *Journal of Memory and Language, 54*, 491-514.

Chen, J.-Y., & Chen, T.-M. (2007). Form encoding in Chinese word production does not involve morphemes. *Language and Cognitive Processes, 22*(7), 1001-1020.

Chen, T.-M., Dell, G., & Chen, J.-Y. (2007). A cross-linguistic study of phonological units: Syllables emerge from the statistics of Mandarin Chinese, but not from the statistics of English. *Chinese Journal of Psychology, 49*(2), 137-144.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics, 16*, 22-29.

Cutler, A. (1985). Cross-language psycholinguistics. *Linguistics, 23*, 659-667.

Cysouw, M. (2005). Quantitative methods in typology. In R. Kohler, G. Altmann, & R. G. Piotrowski (Eds.) *Quantitative Linguistik: Ein internationales Handbuch* [Quantitative linguistics: An international handbook] (pp. 554-578). Berlin: Walter de Gruyter.

DeFrancis, J. (1984). *The Chinese language: Fact and fantasy*. Honolulu: University of Hawaii Press.

Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, *17*, 149-195.

Duanmu, S. (2007). *The phonology of Standard Chinese* (second edition). Oxford: Oxford University Press.

Duanmu, S. (2010, January). Phonology and the Chinese lexicon. Talk presented at the CUNY Conference on the Word in Phonology, New York City.

Ehrich, J. F., & Meuter, R. F. I. (2009). Acquiring an artificial logographic orthography: The beneficial effects of a logographic L1 background and bilinguality. *Journal of*

Cross-Cultural Psychology, 40, 711- 745.

Eu, Jinseung. 2008. Testing search engine frequencies: patterns of inconsistency. *Corpus Linguistics and Linguistic Theory, 4*, 177-208.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32* (5), 429-492.

Fellbaum, C. (1999). (Ed.) *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

French, R. M., & Jacquet, M. (2004). Understanding bilingual memory: Models and data. *TRENDS in Cognitive Sciences*, *8* (2), 87-93.

Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. *Language*, *86* (3), 663-687.

Haspelmath, M., Dryer, M. S., Gil, D., & Comrie, B. (2005). *The world atlas of language structures*. Oxford: Oxford University Press.

Hawkins, J. A. (2007). Processing typology and why psychologists need to know about it. *New Ideas in Psychology, 25*, 87-107.

Hsu, C.-W., & Myers, J. (2009, September). Units of phonological encoding in Mandarin speech production. Talk presented at Taiwan Psychology Association 48th Annual Conference, Taipei, Taiwan.

Huang, C.-Y., & Lee, C.-Y. (2010). Hemispheric processing of Chinese polysemy in the disyllabic verb/ noun compounds: An event-related potential study. *Proceedings of the NAACL HLT 2010 First Workshop on Computational Neurolinguistics* (pp. 45-51). Los Angeles, California.

Huang, H. S., & Hanley, J. (1994). Phonological awareness and visual skills in learning to read Chinese and English. *Cognition, 54*, 73-98.

Huang, H.-W., Lee, C.-Y., Tsai, J.-L., Lee, C.-L., Hung, D. L., & Tzeng, O. J.-L. (2006). Orthographic neighborhood effects in reading Chinese two-character words. *NeuroReport*, *17* (10), 1061-5.

Huang, J. C. T. (1984). Phrase structure, lexical integrity, and Chinese compounds. *Journal of the Chinese Language Teachers Association, 19* (2), 53-78.

Jaeger, T. F., & Norcliffe, E. J. (2009). The cross-linguistic study of sentence production. *Language and Linguistics Compass, 3/4,* 866-887.

Janssen, N., Bi, Y., & Caramazza, A. (2008). A tale of two frequencies: Determining the speed of lexical access for Mandarin Chinese and English compounds. *Language and Cognitive Processes, 23*, 1191-1223.

Leong, C. K. (2006). Making explicit children's implicit epilanguage in learning to read Chinese. In Ping Li, Li Hai Tan, Elizabeth Bates, & Ovid J. L. Tzeng (Eds.) *The handbook of East Asian psycholinguistics, vol. 1: Chinese* (pp. 70-80). Cambridge, UK: Cambridge University Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in spoken production. *Behavioral and Brain Sciences, 22* (1), 1-75.

Lin, Y.-H. (2007). *The sounds of Chinese*. Cambridge: Cambridge University Press.

Liu, Y., Shu, H., & Li, P. (2007). Word naming and psycholinguistic norms: Chinese. *Behavior Research Methods*, *39* (2), 192-198.

Luce, P. A., Pisoni, D. B., & Goldinger, S.D. (1990). Similarity neighborhoods of spoken words. In G. T. M Altman (Ed.), *Cognitive models of speech processing:*

Psycholinguistic and computational perspectives (pp. 105-121), Cambridge, MA: MIT Press.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The Time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General, 132* (2), 202-227.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18* (1), 1-86.

Mester, R. A. (1990). Patterns of truncation. *Linguistic Inquiry, 21* (3), 478-485

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language, 30*, 69-89.

Mielke, J. (2008) *The emergence of distinctive features*. Oxford: Oxford University Press.

Myers, J. (2006). 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. (2007). Generative morphology as psycholinguistics. In G. Jarema & G. Libben (Eds.), *The mental lexicon: Core perspectives* (pp. 105-128). Amsterdam: Elsevier.

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

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

Myers, J., Taft, M., & Chou, P. (2007). Character recognition without sound or meaning. *Journal of Chinese Linguistics, 35* (1), 1-57.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127-162.

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology 49*, 85-117.

O'Seaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition, 115*, 282-302.

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory exposure. *Cognition, 83*, B13-B23

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

Perry, C., & Zhuang, J. (2005). Prosody and lemma selection. *Memory & Cognition*, *33* (5), 862-870.

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin, 133* (2), 227-244.

Roelofs, A. (1996). Morpheme frequency in speech production: Testing WEAVER. In G. E. Booij & J. van Marle (Eds.), *Yearbook of morphology* (pp. 135-154). Dordrecht, the Netherlands: Kluwer Academic Press.

Roelofs, A., & Baayen, H. (2002). Morphology by itself in planning the production of spoken words. *Psychonomic Bulletin & Review, 9*, 132-138.

Schreuder, R., & Baayen, H. (1997). How complex simplex words can be. *Journal of Memory and Language, 37*, 118-139.

Sibley, D.A., Kello, C.T., & Seidenberg, M.S. (2009). Error, error everywhere: A look at megastudies of word reading. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1036-1041). Amsterdam, Netherlands: Cognitive Science Society.

Slobin, D. I. (Ed.). (1985-1997). *The crosslinguistic study of language acquisition. Vol. 1-5*. Mahwah, NJ: Lawrence Erlbaum Associates.

Taft, M., & Zhu, X. (1995). The representation of bound morphemes in the lexicon: A Chinese study. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 293-316). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Tardif, T. (2006). The importance of verbs in Chinese. In Ping Li, Li Hai Tan, Elizabeth Bates, & Ovid J. L. Tzeng (eds.) *The handbook of East Asian psycholinguistics, vol. 1: Chinese* (pp. 124-135). Cambridge: Cambridge University Press.

Tseng, C.-H., Huang, K.-Y., & Jeng, J.-Y. (1996). The role of the syllable in perceiving spoken Chinese. *Proceedings of the National Science Council, Part C: Humanities and Social Sciences, 6* (1), 71-86.

Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32* (2), 387-398.

Zhang, N. (2007). Root merger in Chinese compounds. *Studia Linguistica, 61* (2), 170-184.

Author note

Thanks to Sau-chin Chen, San Duanmu, Chiungwen Hsu, Yuguang Ko, Chia-Ching Ruan, Jane Tsay, Chi-Shing Tse, Niina Zhang, the editors, and two anonymous reviewers for many helpful suggestions, and apologies if I ignored or misunderstood any of them. Research support came from National Science Council grant 97-2410-H-194-067-MY3.