Cross-morphemic predictability
and the lexical access of compounds in Mandarin Chinese

James Myers and Shuping Gong

To appear in *Folia Linguistica*, 2001

Abstract

Chinese poses a challenge for models of compound processing, since the basic notions of morpheme, word and phrase are not consistently distinguished by native speakers. It is thus proposed that the mental lexicon consists of linked and overlapping listemes (in the sense of Di Sciullo and Williams 1987) which can potentially be of any size (morpheme, word, phrase). One implication of this approach for compound processing is that cross-morphemic predictability should play an important role: the more predictable one morpheme is from the other, the easier the compound should be to access. To study this implication, cross-morphemic predictability is quantified using the measure of Mutual Information from information theory, which divides the frequency of the constituent of interest (i.e. a compound) by the frequencies of the components (i.e. morphemes). This leads to two specific predictions: compound frequency should have a positive effect on lexical access, but morpheme frequency should have a negative effect. In Experiment 1, it is demonstrated that a very simple connectionist network, built according to the overlapping listeme approach, conforms to these two predictions. This suggests that separate effects of word and morpheme frequency need not require separate processing levels for words and morphemes. Experiment 2 then compares the network's behavior with Chinese native speakers in a lexical decision task involving spoken Mandarin Chinese compounds. As predicted, there was a positive effect of word frequency on response speed and accuracy, and a negative effect of morpheme frequency. Suggestions are made for reconciling these results with the more familiar positive or neutral morpheme frequency effects found in other studies.

Acknowledgements

The authors would like to thank the audience at the Compound Processing Workshop at the Morphology 2000 Conference, held in Vienna on February 28, 2000, an anonymous reviewer, and especially Dominiek Sandra for insightful comments and suggestions. Experiment 2 was originally performed as part of Gong (1999), which also reports a modeling experiment similar to Experiment 1. Support for S. Gong's studies came from National Science Council (Taiwan) grant 87-2418-H-194-012. We are solely responsible for any errors or misanalyses that may remain.

Address of the first author:

James Myers
Graduate Institute of Linguistics
National Chung Cheng University
Min-Hsiung, Chia-Yi 621
Taiwan
lngmyers@ccunix.ccu.edu.tw

1. Introduction

The central question in the psycholinguistics of morphology concerns what role, if any, the component morphemes play in the processing of words. This question is complex enough in the languages which have provided the bulk of the data, but in Chinese, the peculiar properties of this language turn it into a multidimensional challenge of quite fascinating proportions, with implications, we think, for the understanding of word processing cross-linguistically.

First, since Chinese doesn't have much in the way of inflection, and productive affixation processes are rather limited, focus must be placed on compounds, a structure that has been less intensively investigated in the psycholinguistic literature (though one hopes this special issue will go some way to redress this imbalance). Compounding is itself already understudied in linguistics as compared to other sorts of morphology (Fabb 1998), and so psycholinguists have less guidance about what theories to test or challenge.

Second, in spite of their ubiquity in the Chinese lexicon, Chinese compounds don't fit well into the standard definition (i.e. compounds are concatenations of free morphemes). Instead, as Starosta, Kuiper, Ng and Wu (1997) have argued, many Chinese words traditionally labeled compounds are perhaps not. Thus they suggest that words that are completely semantically opaque (e.g. *dongxi* "thing", but literally "east-west") should be treated as monomorphemic; the characters misleadingly imply a morphological structure that is actually mere historical residue. At the same time, they point out that even clearly bimorphemic words may not be true compounds, since often the second morpheme is used in a semantically restricted way more typical of affixation than compounding. For example, the role of *kai* "open" in *zoukai* "leave" (literally "walk-open") seems to be parallel to that of *able* in *readable*; that is, it is an affix, not the head of a compound, in spite of the existence of a free morpheme written with the same character. All this suggests that even if we had separate processing models that worked well for monomorphemic forms, for affixation, and for compounding, we would still need some way of determining which model to apply in any particular case in Chinese.

A third complicating property of Chinese morphology, in some ways related to the second, is the implication from the orthography that the basic unit is the character, not the word or morpheme per se. Like a morpheme, a character is traditionally thought of as the smallest unit of meaning by Chinese readers, but like a word, it is also considered the smallest unit of discourse. Chinese speakers get very little evidence one way or the other about what constitutes a "word" in the linguistic sense: words are not separated by spaces in the standard orthography, entries in dictionaries are not for "words" but for separate characters (with the exception of rather recent attempts like DeFrancis 1996), sample usages of characters in dictionaries don't distinguish between what linguists would consider words, idioms, and syntactic phrases, and most fundamentally of all, there is no consistent word for "word" in Chinese (*zi* primarily means character, *ci* is also often used for syntactic phrases).

This traditional view among Chinese speakers is challenged by the tradition in Western linguistics to treat orthography as purely parasitic on "real language". DeFrancis (1989) even goes so far as to suggest that Chinese characters form something more like a syllabary than a set of logographs. Moreover, solid linguistic arguments can be made for the existence of words in Chinese (see e.g. Duanmu 1998). It is thus extremely tempting to declare that Chinese native speakers have simply been fooled by their writing system, and that in reality, morphemes and words are the crucial units, not characters. However, there is some experimental evidence that the morpheme/word model doesn't translate perfectly into Chinese. For example, Hoosain (1992) gave short texts of written Chinese to college students (native speakers of Cantonese), and found many disagreements on the proper way to segment the text into words. Tsai, McConkie and Zheng (1998) followed up on this with a more thorough word-segmentation task with Mandarin-speaking college students in mainland China and Taiwan. Like the earlier study, there was much variation in how readers segmented words in the text. Moreover, readers from mainland China were more consistent in their word segmentations, suggesting that judgments of wordhood can be affected by experience with alternative orthographies that do put spaces between words (i.e. Pinyin, a romanization system used in mainland China but not Taiwan). However, since cross-group correlations were quite high, the authors conclude that the variation that was found should be ascribed to varying reader thresholds for judging the status of boundaries (i.e. whether to judge them as separating morphemes, words, or phrases), not to a complete lack of the "word" concept.

There is a fourth property of Chinese that complicates the picture still further: rampant homophony. Even if tone is taken into account, Mandarin has far fewer than 1500 distinct syllables, each written with one to thirty (or more) different characters (Ho 1976). If tone is ignored (which may be justified by the surprisingly minor importance of tone in perception and lexical access; see e.g. Taft and Chen 1992, Ye and Connine 1999), the number of homophones per syllable type jumps fourfold. This makes it extremely implausible that spoken Chinese is processed morpheme by morpheme. Listeners would constantly be garden-pathing, making tentative guesses about the morphemic identity of each syllable, guesses that could not be resolved without cross-checking with tentative guesses about adjacent syllables. As Packard (1999) concludes, the basic unit in the lexical access of spoken Chinese is therefore likely to be the word. This conclusion would be quite convincing if it weren't for the fact, discussed above, that native speakers seem to disagree about what the "word" is in Chinese.

A possible way out of this dilemma is to demote the notion of "word" in the linguist's sense from its status as sole goal of the lexical parser. This is not to say that we claim that the real unit in Chinese is actually the character, as traditionally thought. In addition to the problems just mentioned for spoken word recognition, there is the robust finding of word-frequency effects in lexical access, both for written and spoken Chinese words (see references described below), suggesting that whole words are indeed listed, in some sense, in the lexicon. Instead, what we are claiming is that the Chinese mental lexicon doesn't look much like either a Western word-based dictionary or a Chinese character-based dictionary. In psycholinguistic terms, we reject both purely word-based theories of the mental lexicon (e.g. Butterworth 1983) and purely morpheme-based theories (e.g. Taft and Forster 1976). But we also reject the third commonly offered possibility, namely a lexicon with separate levels for words and for morphemes (e.g. Zhou and Marslen-Wilson 1994, 1995). If the problem is precisely how one should distinguish morphemes from words (and words from phrases), it is not solved by placing them on different levels by fiat.

We suggest that something closer to the right approach is to imagine a lexicon with entries of any size -- morpheme (written with one character or more), word, phrase -- all represented at the same level, but with links between them to indicate overlaps in form and/or meaning. To take an example from Hoosain's (1992) word segmentation experiment, the three-character string *yi* ("already") *jin* ("enter") *le* ("completive aspect") could be accessed as three lexical units (*yi*, *jin*, and *le*) and/or as two (*yijin* and *le*) and/or as one (*yijinle*), depending on how they happened to be retrieved by a particular person during production or parsed during comprehension. In other words, we suggest that psycholinguists should focus more on what Di Sciullo and Williams (1987) call the LISTEME, a form-meaning unit defined solely by its listing in the mental lexicon, regardless of its syntactic or morphological status. The degree to which Chinese speakers consider any given meaningful string to be a multisyllabic morpheme, versus a semantically opaque compound, versus a semantically transparent compound, versus a syntactic phrase, would to an important extent depend on how often that particular string has been encountered (with that meaning) in the past. This overlapping listeme approach shares basic insights with morphological theories that question the morpheme's status as fundamental atom of the lexicon (e.g. Anderson 1992, Bochner 1993) and particularly with network-type models that are concerned with the effect of frequency on morphological processing (e.g. Bybee 1985, Smith 1995, Stemberger 1998, and Rueckl and Raveh 1999).

Of course the overlapping listeme approach, as described, seems to gloss rather quickly over issues that are of deep concern to morphologists (namely, the roles of semantics, morphological structure, and syntax in the knowledge of words), but in this paper we don't intend to present this approach as a reductionist theory of everything that needs to be explained in morphology. Nevertheless, it does bring some useful insights. For example, a compound is considered to be semantically transparent if its meaning can be derived by combining the meanings of its component morphemes. This presupposes that the meanings of the morphemes themselves are clear, and this can only be the case if the morphemes sometimes appear in isolation (i.e. as words with clear meanings) or appear in a certain number of other semantically transparent compounds. Either way, the component morphemes will tend to find themselves in a variety of contexts (e.g. in novel syntactic phrases or compounds), and this in turn will tend to make the processing system treat the components of a transparent compound as separate listemes. A transparent compound should then have more of a tendency to be processed via concatenation of separately accessed listemes than an opaque compound, which should tend to be accessed as a whole listeme. If the overlapping listeme approach is correct, there should be situations where it is this difference in listeme access, rather than semantics itself, that causes differences in the behavior of transparent and opaque compounds (though see Libben 1998 for a more sophisticated analysis of the importance of transparency to compound processing).

More generally, the overlapping listeme approach predicts that the degree to which a given string is processed as a listeme depends on how predictable the component units are from each other. In a string of two Chinese characters, for instance, the more predictable the second character is from the first, the more likely one would expect a reader to access the string as a whole constituent. At one extreme, we have the so-called binding word (Taft and Zhu 1995), in which each character is entirely predictable from the other (e.g. the morpheme *qiuyin* "earthworm", written with two characters that only appear in this one word). At the other extreme, where one character is almost completely unpredictable from the other, we might expect readers to consider the string a syntactic phrase (assuming of course that it conforms to syntactic rules).

One standard measure of the mutual predictability of adjacent linguistic units employs the notion of MUTUAL INFORMATION (MI) from information theory (see Church and Hanks 1990 for basic concepts and applications to cross-word collocations in English). MI takes the probability at which the units actually cooccur (i.e. the frequency of the collocation of interest as a proportion of the total set of collocations in the corpus) and divides it by the probability expected if the two units cooccurred by chance (i.e. the product of the frequencies of each unit as a proportion of the total number of units in the corpus). This division controls for the effect of random collocations (i.e. if two highly frequent units often cooccur, this still might be due purely to chance, not because they have any particular affinity for each other). In addition, MI is transformed onto a logarithmic scale (base two), so that the midpoint (i.e. where two units are completely unpredictable from the other) has the MI value of 0. The formula for MI is thus as given below.

[INSERT FIGURE 1]

This measure is commonly used in computational linguistics, where it has proven its worth in corpus analyses that have focused on precisely the problem we discussed earlier: the segmentation of Chinese text into words. For example, using an MI of 2.5 as a cutoff point for wordhood, Sproat and Shih (1990) built a parser that was highly successful in segmenting words in a newspaper corpus, giving an accuracy rate of about 94% (as compared with word segmentations produced by linguists). Wang, Huang and Chen (1995) and Huang, Ahrens and Chen (1998) applied MI calculations to corpus analyses of more detailed questions of wordhood in Chinese, achieving similar success. As with any computational model that mimics human abilities, it is not necessarily the case that the computers and the humans are processing in precisely the same way, but results such as these are nevertheless inspiring. Could it be that Chinese readers (and listeners) deal with some of the complex problems posed by word and morpheme segmentation by performing some sort of automatic calculation of cross-unit predictability?

To make this suggestion precise and testable, we adopted MI as the measure of cross-morphemic predictability. Other ways of measuring predictability exist, however. For example, one may calculate the cross-morphemic predictability of a compound AB as the frequency of AB divided by the cumulative frequencies of all other compounds that contain A or B (as suggested to us by Dominiek Sandra, personal communication). The numerator here is parallel to that in the MI formula (except that MI uses compound frequency as a proportion of all two-character strings in the corpus). However, the denominator in this alternative formula differs somewhat from that used in calculating MI. This is because the MI denominator is the product of the (proportional) frequencies of A and B, whereas in the alternative formula the denominator is roughly the sum of these frequencies (more precisely, the sum of the frequencies of all compounds containing one or both of them).

Nevertheless, both are reliable measures of predictability. To see this, consider two artificial corpora, one consisting of a sequence of three tokens of the compound AB, i.e. L1 = ABABAB, and one containing this as a substring but also including other bimorphemic compounds, e.g. L2 = ABABABACAD. Intuitively, the cross-morphemic predictability of AB in L1 is higher than in L2. The alternative formula gives the values for AB of 3/3 = 1.0 in L1, and 3/5 = 0.6 in L2; as desired, the predictability is higher in L1. The MI formula gives a value for AB in L1 of $\log_2((3/5)/((3/6)*(3/6)))$ (since there are three tokens of AB out of the set of five possible two-character substrings of ABABAB and three tokens each of A and B out of the set of six morpheme tokens), which works out to $\log_2(2.4) = 1.26$, whereas in L2, $MI(AB) = \log_2((3/9)/((5/10)*(3/10))) = \log_2(2.2) = 1.15$. Thus as with the alternative formula, the predictability is correctly described as higher in L1. Of course, in practice it is difficult to calculate the frequency of a bimorphemic compound as a proportion of all possible two-morpheme strings in a corpus; estimating it by dividing compound frequency by the number of morpheme tokens in the corpus (which is what we did in the MI calculations in this paper) gives virtually identical results with a corpus of any realistic size. Importantly, for either formula, a key property holds: cross-morphemic predictability is positively correlated with compound frequency and negatively correlated with morpheme frequency. Teasing apart any different predictions made by different ways of calculating predictability is beyond the scope of this paper; we adopt MI because of our interest in linking research on the morphological decomposition of compounds with that on the segmentation of words from text or speech. Hence from now on we will tend to use the notions of "mutual information" and "predictability" relatively interchangeably.

As just noted, the formula for MI (and probably most reasonable alternatives) makes two specific predictions. First, we expect that high word frequency (the value in the numerator) should ease lexical access, since this will make the compound as a whole more listeme-like.

This is of course a well-established fact for words in general (e.g. Forster and Chambers 1973, Whaley 1978). However, the MI formula also makes a prediction that may seem somewhat surprising given the discussions of compound processing in the literature: morpheme frequency (represented in the denominator) should have a negative effect. That is, all else being equal, experimental participants should have greater difficulty recalling a compound word if its components are of high frequency relative to the frequency of the whole word, since this will lower the compound's MI value.

On the face of it, the effects of morpheme frequency in the experimental literature do not seem to conform well with these predictions. While all researchers find positive word frequency effects in the lexical access of compounds, some also find positive morpheme frequency effects (e.g. Taft and Forster 1976, Andrews 1986, using written English), while others find no morpheme frequency effects under apparently similar conditions (e.g. van Jaarsveld and Rattink 1988, using written Dutch). As far as we are aware, no study on European languages has reported the sorts of negative morpheme frequency effects we expect (cf. van Jaarsveld and Rattink 1988, who found negative morpheme frequency effects for nonwords in one experiment, namely higher frequencies of first morphemes led to slower "nonword" responses).

In research on Mandarin Chinese, however, the results match the model rather better. Chen (1993), Tsai (1994), Lü (1996), and Peng, Liu and Wang (1999) all found negative morpheme frequency effects in lexical decision tasks with visually presented Chinese compounds. Notably, such effects have only been found in semantically opaque compounds; even with written Chinese, semantically transparent compounds tend to show positive morpheme frequency effects. In the terminology of the overlapping listeme approach, this suggests that pure cross-morphemic predictability (independent of other factors) has its greatest effect on compounds that are already towards the "listeme" end of the continuum. Lü (1996) is particularly interesting for the purposes of this paper, since one of her theoretical variables was the number of word types in which a morpheme occurs (similar to the notion of morphological family size studied by Schreuder and Baayen 1997), and this is closely related to cross-morphemic predictability. Both morphemes with larger family sizes and morphemes of higher lexical frequency increased reaction times for the compounds that contained them (unfortunately Lü did not control for semantic transparency). However, Liang (1992) and Lee (1995) found no effect of morpheme frequency for visually presented opaque compounds (along with the usual positive morpheme frequency effects for transparent compounds). Finally, in the literature's sole non-primed lexical decision task using auditorily presented Mandarin compounds, Zhou and Marslen-Wilson (1994) found no morpheme frequency effects even for semantically transparent compounds.

To better understand this variation across languages and conditions and its relation to the overlapping listeme approach, we decided first to make the predictions of the approach precise by building an implementation on a computer (section 2). We then compared the behavior of this model with the results of an experiment examining the lexical access of spoken Mandarin Chinese compounds (section 3). We chose to look at spoken Chinese both because it has been sorely neglected in the psycholinguistic literature, and because, as we saw above, it poses especially complex problems for models of morphological processing.

2. Experiment 1: Connectionist modeling

To demonstrate how cross-morphemic predictability is expected to play a role in lexical access assuming the overlapping listeme approach, we built a simple connectionist model that does not place words and morphemes into separate levels. The model takes an entire compound word as input and gives the same compound word as output. Accuracy on this simple task is thus a measure of how well the network "remembers" the whole word.

Such a network is expected to show MI effects for purely geometrical reasons. Consider a toy lexicon with only one token each of three two-morpheme words: AB, AC and DE. Since the morpheme A is twice as common in the corpus as either morpheme in DE, the MI value for AB and AC will be lower than that for DE (specifically, MI(AB) = MI(AC) = 1.85, MI(DE) = 2.85). Now suppose that we present these compounds to a two-layer feedforward connectionist network that we train to produce an output identical to the input (e.g. inputting AB should produce AB as output), a so-called autoassociation task. Each node in the output receives information from both morpheme nodes in the input, in the sense that its activation is a function of the sum of the products of the weights of the connections with the activations of the connected input nodes. Since DE has a higher MI value than AB or AC, the nodes in the input representation for DE will cooccur more consistently than the morphemes in the other two compounds. That is, for the compound DE, there will be two consistent sources to tell the second output node that it's supposed to be E, namely both the input node for D and the input node for E. By contrast, for the other two compounds AB and AC, the only reliable source of information about the second output node is the second input node; the A node in the input is not reliable, since it cooccurs with both B and C. If accuracy of the output depends on the number of sources of information there are for each output node (which is true of the learning algorithm we use in our model), then the network must be more accurate for DE than for AB or AC.

2.1 Participants

A two-layer feedforward network was built using the *tlearn* software (Plunkett and Elman 1997), running in Windows 95. The network had 16 input nodes and 16 output nodes (since each vector represented a "compound" with two 8-node "morphemes", as described below). Thus each input node connected with every output node, but nodes did not connect with each other within layers. The learning algorithm in *tlearn* is back-propagation (Rumelhart, Hinton, and Williams 1986), which means that with just two layers, the network was essentially identical to a perceptron (Rosenblatt 1958). Back-propagation works by initializing the weights of the connections to random values, comparing the model's output with the desired output, and adjusting the weights so that the error is reduced. The initial randomization of weights allows one to test separate experimental "participants", each one with identical architecture but a different initial set of weights. Thus we created thirty participants by allowing *tlearn* to choose thirty different seed values for its randomization algorithm, with weights restricted to the range ±0.5.

2.2 Materials

The training set was built from a set of eight "morphemes" (A, B, ..., H) coded as 8-node vectors, where only one node was activated (e.g. A=[10000000], B=[01000000]). This localist representation meant that the model would not be confused by "homophones" (i.e. where two distinct morphemes shared active nodes). The morphemes were combined into 64 compound "words" represented by 16-node vectors (e.g. AB=[1000000001000000]). Since calculating MI values involves the division of word frequency by morpheme frequencies, there are two ways that a higher MI value can arise (i.e. with relatively high word frequency or with relatively low morpheme frequency), and likewise two ways to derive a lower MI value. Hence we wanted to examine the ability of the network to learn four types of words: **H:H/M** (high MI, high word frequency, mid morpheme frequency), **H:M/L** (high MI, mid word frequency, low morpheme frequency), **L:L/M** (low MI, low word frequency, mid morpheme frequency), and **L:M/H** (low MI, mid word frequency, high morpheme frequency). Words of each of these types (respectively labeled AB, CD, EF, GH) were given their appropriate properties by carefully adjusting word and morpheme frequencies of the 64 compounds, creating a total of 10033 different word tokens. Table 1 gives the number of word tokens, the number of tokens of these words' component morphemes, and the resulting MI values for the four key words AB, CD, EF, and GH.

[INSERT TABLE 1]

Note that the MI values for EF and GH are quite negative, which means that the component morphemes of these words are essentially in complementary distribution, appearing together less often than would be expected by chance. Further examination of this table shows that while we were careful to make the MI values for AB and CD very close, and likewise for EF and GH, high versus mid versus low word and morpheme frequency must be understood here solely in relative terms.

2.3 Procedure

The 30 network participants were trained in an autoassociation task (that is, inputs and outputs were identical). The 10033 tokens were presented to each participant in a different random order. Thus higher-frequency words were presented more often than lower-frequency words, so the back-propagation procedure guaranteed that their accuracy would be higher. We set the two crucial back-propagation parameters to values typical in the literature (learning rate = 0.3, momentum = 0.9). Since these parameters were the same across all tokens and participants, they cannot be responsible for our results (changing them would merely speed up or slow down learning overall).

After training was complete, we froze the final weights for each participant and then tested the four target words AB, CD, EF, and GH. Actual output vectors were compared with the desired vectors, and for each word and each participant an error value was calculated simply by summing the absolute values of the differences between the node activations of the actual and desired output vectors. (We decided that more commonly used but more complex error formulas, such as Euclidean distance, were not necessary for our purposes.)

2.4 Results

The mean error values are given in Table 2.

[INSERT TABLE 2]

We performed a one-way ANOVA with repeated measures (by participant) on the error values of the four words. There was an extremely significant effect of compound type across the four word types ($F(3,87)=1304$, $p<.001$). (The fact that this result was so significant is not itself significant, as it presumably results from the relatively low variability produced by the way we generated our participants.) Scheffé's test showed that every word had a mean error value significantly different from every other at the 0.05 level. In particular, higher word frequency resulted in lower error values when morpheme frequency was controlled (H:H/M vs. L:L/M), and higher morpheme frequency increased error values when word frequency was controlled (L:M/H vs. H:M/L). Together this meant that a higher MI resulted in lower error values, as predicted.

2.5 Discussion

The results confirm that MI does in fact play an important role in this kind of network. In particular, the effect of morpheme frequency was negative, where higher-frequency morphemes made the words they composed harder to process rather than easier, contrary to models where compounds are thought to be accessed via one or both of the morphemes (e.g. Taft and Forster 1976). We consider our results with this model important, since psycholinguists often seem to assume that if morphemes and words show separate effects, there must be separate levels for morphemes and words. Our model provides an existence proof that this isn't necessarily the case.

Of course, the effects of word frequency and morpheme frequency in the experimental literature do not seem to be as simple as predicted by our simple model. In addition to the variation across studies noted in the introduction, real morpheme frequency effects in compound processing are sometimes found to be restricted to just one position, usually the first (e.g. Taft and Forster 1976 for written English). This aspect is entirely ignored in our model; since all inputs nodes connect to all output nodes, the model doesn't even know the order of the sixteen nodes in a vector, much less the order of the two "morphemes". The model is also not designed to handle priming effects that are increasingly being reported in the literature on compound processing (see e.g. Sandra 1990, Zwitserlood 1994 for Dutch; Zhou and Marslen-Wilson 1995, Liu and Peng 1997 for spoken and written Mandarin respectively; Jarema, Busson, Nikolova, Tsapkini, and Libben 1999 for French and Bulgarian).

Since the main purpose of the model is to demonstrate how MI effects emerge naturally from an overlapping listeme approach, a much more serious concern is the decidedly mixed evidence for negative morpheme frequency effects in the experimental literature. In particular, our connectionist model is completely incapable of producing positive morpheme frequency effects, and to the extent that such effects have been reliably replicated, we gladly acknowledge the reasonable and widely accepted hypothesis that readers and listeners sometimes access compounds via one or both component morphemes. While our simple network cannot handle this, positive morpheme frequency effects are not inconsistent with the overlapping listeme approach in general, since it assumes that listemes can be of any size, including morpheme-sized, and that related listemes are linked; here the linked listemes would be compounds and the morphemes that compose them. It is an interesting open question whether a single computational model can be built to exhibit both negative and positive effects of morpheme frequency under different conditions.

However, our simple network model may help provide a framework for understanding variations between negative and neutral morpheme frequency effects. First, semantic transparency is predicted to be an inhibitory factor on reaction times, since morphemes in semantically transparent compounds tend to be less mutually predictable than in semantically opaque compounds, for reasons discussed earlier. This is precisely what has been found by some studies using written Chinese compounds (e.g. Liang 1992, Lee 1995, and Su 1998). Moreover, Tsay, Myers, and Chen (1999) show that children acquiring Taiwan Southern Min (related to and morphologically quite similar to Mandarin Chinese) split up semantically transparent compounds into separate prosodic constituents more often than semantically opaque ones. However it should be mentioned that other studies find that transparent compounds are recognized faster (e.g. Tsai 1994, Lü 1996) or find no difference in processing speed between transparent and opaque compounds (Chen 1993).

Turning to differences in results across languages and modes of stimulus presentation, recall that the morphemes in our model were specifically designed to have completely distinct vectors. In real languages, morphemes are often similar to each other phonologically, orthographically, or both. From the perspective of the model, a set of similar morphemes will tend to behave as if they are all varying tokens of a single morpheme. Thus the more such similar morphemes there are in a given language, the less the variation in true morpheme frequency will be noticeable to the model, and therefore the less of an MI effect there will be. Conversely, linguistic systems in which morphemes are very distinct in form should show stronger MI effects, including negative effects of morpheme frequency.

Now one way in which Mandarin differs from European languages (in addition to those listed in the introduction) is that its orthography clearly disambiguates morphemes, which can differ from each other in form to a much greater extent than morphemes written with a mere 26 or so letters. Following the above argument, written Chinese compounds should therefore tend to show stronger MI effects than compounds written in an alphabetic orthography. This means that morpheme frequency effects should be more strongly negative for written Chinese compounds than for compounds (written or spoken) in languages like English. This may help explain why negative morpheme frequency effects have been found in Chinese but not in other languages. However, as noted earlier, spoken Mandarin also has a high degree of homophony, even among compounds, and so we expect that spoken Chinese compounds should tend not to show strong MI effects. The fact that Zhou and Marslen-Wilson (1994) found no morpheme frequency effect in auditorily presented semantically transparent compounds in Mandarin may thus result from these sorts of compounds having particularly low overall MI values.

3. Experiment 2: Auditory lexical decision task with Mandarin speakers

We decided to test the strength of our model by applying it in a domain where we expect MI effects to be weakest, namely in auditorily presented semantically transparent compounds in Mandarin. Given the problem of rampant homophony among Chinese morphemes, we ruled out the possibility that we would find any positive morpheme frequency effects; as argued in the introduction, a processing strategy whereby compounds must be accessed via the component morphemes would be extremely counterproductive for spoken Mandarin. The question was whether the null results of Zhou and Marslen-Wilson (1994) were due to the absence of any morphological processing of compounds at all, or simply to their missing a real but small negative effect of morpheme frequency. Therefore we carried out an experiment similar to those conducted by Zhou and Marslen-Wilson (1994), but using more participants and a greater morpheme frequency range. Both of these changes were designed to bring small effects, if any existed, up to statistical significance.

3.1 Participants

Participants were native speakers of Mandarin (students at National Chung Cheng University in central Taiwan), paid for their participation. Data from seven of them ended up being rejected for analysis (four had an error rate of 25% or above in the experiment, two had technical problems with the headphones, and one turned out to be left-handed; handedness was relevant since the right hand was used to press the "real word" button on the button box). This left 54 participants whose data we analyzed.

3.2 Materials

Items were chosen to fall into the same four categories used in the connectionist model (H:H/M, H:M/L, L:L/M, and L:M/H); basic data on these categories are given in Table 3, and lists of all of our words with their statistical information are given in the Appendix.

[INSERT TABLE 3]

Zhou and Marslen-Wilson (1994) did not pretest their compounds for semantic transparency, so for our experiment, compounds in each of the four categories were pretested for semantic transparency by first choosing a set of 194 apparently transparent compounds and a set of 194 apparently semantically opaque compounds (borrowed from Lee 1995) and presenting them in a randomly ordered list to 12 native speakers of Mandarin to judge for semantic transparency (i.e. relatedness of word meaning to character meaning) on a five-point scale (1 = most transparent). We then chose the 20 items in each category with the best ratings for semantic transparency to use in the experiment. There was no significant difference in the semantic transparency scores across the four categories ($F(3,76)=0.152$, $p > .05$).

Frequency counts came from a large corpus of written Mandarin (the Academia Sinica Balanced Corpus, based mainly on texts from newspapers in Taiwan; see Chen, Huang, Chang, and Hsu 1996). The version of the corpus that we used was the one published as CKIP (1993), which has about 9.53 million word tokens (where words include compounds and other polymorphemic units analyzed as nonphrasal by the makers of the corpus) and about 14.46 million character tokens. By contrast, the corpus used by Zhou and Marslen-Wilson (1994) contained only 1.31 million word tokens and 1.81 character tokens. We defined high word frequency as over 100 tokens in our corpus (10.5 per million characters), mid word frequency as 20-60 tokens (2.1-6.3 per million), and low word frequency as fewer than 6 tokens (0.6 per million). Following Zhou and Marslen-Wilson (1994), we simply equated morpheme frequency with character frequency. Chinese characters do not correlate perfectly with morphemes, since there a small number of polysyllabic morphemes, usually foreign borrowings, that are written with two or more characters, and some characters have different meanings and/or pronunciations in different contexts. However, such cases are relatively uncommon, and since our materials contained only disyllabic semantically transparent compounds, these caveats are even less relevant. High morpheme frequency was defined as over 10000 character tokens in our corpus (692 per million characters), mid morpheme frequency as 2000-8000 tokens (138-553 per million), and low morpheme frequency as fewer than 1000 tokens (69 per million).

Perhaps due to our larger corpus, our morpheme frequency range was much wider than that used by Zhou and Marslen-Wilson (1994). Word frequency ranges were comparable; in their materials, mean high word frequency was around 95 tokens (72.5 per million words) and mean low word frequency was around 5 tokens (3.8 per million), different by a factor of 19, while for us, mean high word frequency was around 95 tokens (9.97 per million) and mean low word frequency was around 5 tokens (0.52 per million), also different by a factor of 19. However, more importantly for the theoretical issues in this paper, for Zhou and Marslen-Wilson (1994) mean high morpheme frequency was around 1400 tokens (773 per million characters) and mean low morpheme frequency around 200 tokens (110 per million), different by a factor of 7, whereas for us, mean high morpheme frequency was around 20000 tokens (1383 per million characters) and mean low morpheme frequency around 500 tokens (34.58 per million), different by a factor of 40.

As in the connectionist experiment, the materials were designed with MI in mind. In fact, as can be seen from Table 3, the MI values for the L:L/M and L:M/H categories were well below the cutoff value of 2.5 used to define wordhood in the Sproat and Shih (1990) study mentioned in the introduction, while the MI values for the H:H/M and H:M/L categories were well above. We also were careful to control the word and morpheme frequencies used to derive these MI values. Thus while there was of course a significant difference in word frequency across all four categories ($F(3,76)=15.01$, $p<.001$), there was no significant difference in word frequency across the H:M/L and L:M/H categories (both by Scheffé's test at an alpha level of 0.05, and a two-tailed unpaired t-test, $t(38)=-1.53$, $p=.13$). There was also an overall significant difference in first morpheme frequency ($F(3,76)=51.28$, $p<.001$) and in second morpheme frequency ($F(3,76)=40.73$, $p<.001$), but no significant differences across the H:H/M and L:L/M categories (both by Scheffé's tests at an alpha level of 0.05, and by two-tailed unpaired t-tests: $t(38)=0.37$, $p=.74$, and $t(38)=0.33$, $p=.75$, for first and second morpheme frequency respectively).

Since our materials were presented auditorily, another important factor was syllable frequency. Zhou and Marslen-Wilson (1994) found that higher syllable frequency slowed reaction times in a lexical decision task (where syllable frequency was calculated by adding the frequencies of all homophonous morphemes; syllables with different lexical tones were not considered homophonous). They plausibly suggest that this negative syllable frequency effect may be due to competition between similar-sounding words during the decision-making process, an explanation supported by their finding that the syllable-frequency effect only showed up in the first position, consistent with left-to-right access of phonological forms. Unfortunately, since syllable frequency normally correlates with morpheme frequency, we faced a possible confound which could give misleading support to our hypothesis. That is, higher morpheme frequency could cause longer reaction times merely because the associated syllables are also of higher frequency.

Thus we were very careful to control the syllable frequency, especially in the first position. As shown in Table 3, the frequencies of the syllables in first position were kept within 10 and 50 thousand tokens (692-3458 per million characters, which are always monosyllabic). The frequencies of the syllables in second position were harder to control given all the other constraints on our materials, and ranged between a low of 396 tokens and a high of 279354 tokens (27-19322 per million characters).

As can be seen from Table 3, the mean syllable frequencies do not pattern with the mean morpheme frequencies; even in the L:M/H category (with high morpheme frequency), the first syllable frequency is merely 1.1 times higher than the mean first syllable frequency in the H:M/L category (with low morpheme frequency). Moreover, across all four categories the mean syllable frequencies are not significantly different, either in first position ($F(3,76)=1.74$, $p >.05$) or in second position ($F(3,76)=1.59$, $p >.05$). In particular, there is absolutely no significant difference in first syllable frequency in the categories H:M/L and L:M/H, which vary morpheme frequency while keeping word frequency constant (two-tailed unpaired t-test assuming equal variances: $t(38)=-0.82$, $p=.42$). By contrast, morpheme frequency is highly significant across the four categories, in both first and second position, as noted above. In particular, mean first morpheme frequencies for the H:M/L and L:M/H categories differ by a factor of over 45 ($t(38)=-8.45$, $p<.001$). Nevertheless, across all 80 real words, syllable frequencies were significantly correlated with morpheme frequencies (first position: $r(80)=0.33$, $p<.05$; second position: $r(80)=0.27$, $p<.05$).

In addition to our controls, we do not think that syllable frequency poses a serious problem for our purposes, for the following reason. Since syllable frequency is defined as the sum of the frequencies of homophonous morphemes, controlling syllable frequency while varying morpheme frequency actually works against our hypothesis. Specifically, in the H:M/L category, first morpheme frequency is much lower than first syllable frequency (30 per million vs. 1896 per million), which suggests that the morphemes should tend to have either many competing homophones or at least a few high-frequency ones, whereas in the L:M/H category, first morpheme frequency is quite a bit closer to first syllable frequency (1373 per million vs. 2099 per million), which suggests that the morphemes should tend to have fewer homophones or at least lower-frequency ones. These expectations about the number and frequency of homophonous competitors are in fact true. The mean number of homophones of the first syllable in H:M/L compounds is 10.15, with the target morphemes showing a mean frequency ranking of 5.5 (where 1 would indicate that it had the highest frequency of all competing homophones), whereas the mean number of homophones of the first syllable of L:M/H compounds is 6.75, with the target morphemes showing a mean frequency ranking of 1.2. Both of these differences were significant (since F-tests showed that the variances were also different, we used an unpaired heteroscedastic t-test that does not require equal variances; number of homophones: $t(33)=2.13$, $p<.05$, frequency ranking: $t(20)=6.24$, $p<.001$). In short, the morphemes in the H:M/L category both have more homophonous competitors than those in the L:M/H category, and the competitors tend to be of higher frequency. This means that if competition between homophones slows down response times and increases errors, as Zhou and Marslen-Wilson (1994) have suggested in regard to their own results, we expect that the words in the H:M/L category should be *harder* to process than those in the L:M/H category. This is precisely the opposite of what cross-morphemic predictability would lead one to expect, and therefore our hypothesis is sharply distinguishable from this alternative view.

Because the task was a lexical decision task, there were also 80 nonwords. Each of these was created by taking a real semantically transparent compound word (not used in the experiment) and changing the tone of its second syllable. For example, the real word *kong1ping2* "empty bottle", which has tone 1 (a high level tone) on the first syllable and tone 2 (a rising tone) on the second, was turned into a nonword by changing it to *kong1ping4*, with tone 4 (a falling tone) on the second syllable. This procedure made the task maximally difficult, since participants had to pay close attention until the end of the word before making a decision. The benefit of this procedure is that the participants would quickly learn that even the nonwords were very similar to real words and thus perhaps adopt a strategy designed to minimize interference from real-word neighbors during lexical access of the real words as well. We hoped that we would thereby further reduce the potential problem of the syllable-frequency/morpheme-frequency confound. This is another way that our study is not a direct replication of Zhou and Marslen-Wilson (1994), since while they do not describe how their nonwords were created, it is likely that they did not follow this procedure given their interest in the role of phonological processing.

Finally, these materials (all two syllables long) were recorded by a female native speaker of Mandarin (not the second author) and digitized at a sampling rate of 10 kHz. Due to limitations in our equipment, the stimuli were recorded at a relatively fast speaking rate (approximately 200 msec for each disyllable) and were somewhat noisy, but participants did not report any problems with comprehension; their error rates were also not much larger than are typically found in experiments of this kind (see Results).

3.3 Procedure

We used an auditory lexical decision task. Participants were presented with stimuli over headphones in random order (different for each participant) and had to respond to each as real or not by pressing appropriate buttons on a button box as quickly and accurately as possible. The labels on the buttons read *zhenci* and *feici*, respectively, where *zhen* means "real", *fei* means "not", and *ci* refers to a small linguistic unit larger than one character (i.e. what a linguist would consider a word, idiom, or phrase).

Three participants were run simultaneously on separate IBM-compatible computers (Pentium I running a C-based program in DOS). A trial consisted of auditory presentation over headphones of a stimulus, followed by the participant's response (if any); if there was no response within 3 seconds, the program continued to the next trial. Trials were separated by intervals of 1.5 seconds. Before the main experiment, 16 trials were given for practice. Both the number of errors (judging real words as nonwords or taking longer than 3 seconds to respond) and response times (RTs measured from the onset of the stimulus to the pressing of a button) were recorded. RTs for nonwords and for erroneous responses to real words were not analyzed.

3.4 Results

The mean number of errors and response times (RT) from the four categories of words are given in Table 4. As can be seen, there was no speed-accuracy trade-off (which would indicate that participants became more accurate as they responded more slowly, thus hiding any interesting processing).

[INSERT TABLE 4]

To get a sense of the overall effect of MI, one-way ANOVAs with repeated measures were conducted. Unsurprisingly, there were highly significant effects across the four compound types, both for RT (by participant, $F(3,212)=10.58$, $p<.001$; by item, $F(3,76)=6.29$, $p<.001$) and for number of errors (by participant, $F(3,212)=63.44$, $p<.001$; by item, $F(3,76)=9.41$, $p<.001$). However, the more interesting question was whether MI affected speed and accuracy in the expected way. Specifically, we wanted to know if word frequency had a positive effect when morpheme frequency was controlled while morpheme frequency had a negative effect when word frequency was controlled.

Thus we performed separate analyses on H:H/M versus L:L/M, and on H:M/L versus L:M/H. We begin with analyses for H:H/M and L:L/M, where word frequency was varied while morpheme frequency was controlled. With RT analyzed by participant, an F-test found no significant difference in variance, and so we conducted a standard two-tailed paired t-test (in this case equivalent to a one-way ANOVA with repeated measures); the effect was highly significant ($t(53)=-7.60$, $p<.001$). For RT analyzed by item, an F-test also found no difference in variance, so again we conducted a standard two-tailed paired t-test, and again the effect was significant ($t(19)=-3.28$, $p<.01$). In an analysis of the number of errors, F-tests found significant differences in variance, both by participants ($F(53)=0.23$, $p<.001$) and by items ($F(19)=0.25$, $p<.01$), with greater variance for the lower frequency words. Although the paired t-test is relatively robust when the assumption of equal variance is violated, we decided to perform a weaker variant of the t-test that does not require this assumption. This two-tailed, unpaired heteroscedastic t-test showed that the effect of number of errors was significant both by participant ($t(76)=-10.53$, $p<.001$) and by item ($t(28)=-3.81$, $p<.001$).

Now we turn to the results for H:M/L and L:M/H, where morpheme frequency was varied while word frequency was controlled. With RT analyzed by participant, an F-test found no difference in variance, and so we conducted a standard two-tailed paired t-test; the effect was highly significant ($t(53)=-4.56$, $p<.001$). For RT analyzed by item, an F-test also found no difference in variance, so we conducted a standard two-tailed paired t-test; however, this time the effect just missed significance at the .05 level ($t(19)=-2.04$, $p=.056$). In an analysis of the number of errors, F-tests found significant differences in variance, both by participants ($F(53)=0.45$, $p<.01$) and by items ($F(19)=0.14$, $p<.01$); greater variance was found with the higher frequency morphemes. Again we performed a two-tailed, unpaired heteroscedastic t-test, which showed that the effect of number of errors was significant both by participant ($t(92)=-7.04$, $p<.001$) and by item ($t(24)=-3.02$, $p<.01$).

In summary, when morpheme frequency was controlled (H:H/M vs. L:L/M), a higher word frequency resulted in faster reaction times, a lower number of errors, and less variance in the number of errors, whereas when word frequency was controlled (L:M/H vs. H:M/L), higher morpheme frequency slowed reaction times, increased the number of errors, and increased variance in the number of errors. Together these results support our predictions of positive word frequency effects and negative morpheme frequency effects.

Since first syllable frequency and first morpheme frequency correlate in our materials, however, we decided to perform more stringent tests on our results. First we calculated Pearson's correlation coefficient for all syllable frequencies and response measures (RT and number of errors). For second-position syllables, the correlations were nonexistent ($r(80) = 0.09$ for RT, $r(80) = -0.05$ for errors), consistent with the finding of Zhou and Marslen-Wilson (1994) that second-syllable frequency plays no role in lexical access for spoken compound recognition. For first-position syllables, however, the correlations went in the expected directions and were statistically significant (RT: $r(80) = 0.25$, $p<.05$; errors: $r(80) = 0.29$, $p<.05$). For further analysis of the syllable frequency problem we focused on a comparison between H:M/L and L:M/H, which have the same mean word frequencies but different morpheme frequencies. Partialling out first morpheme frequency, the resulting correlations with first syllable frequency were $r(40) = 0.27$ for RT (nonsignificant, $p>.05$) and $r(40) = 0.31$ for errors (significant, $p<.05$). As a final test, we partialled out first syllable frequency in the comparison between H:M/L and L:M/H and examined the remaining correlations with first morpheme frequency; they were $r(40) = 0.00$ for RT (nonsignificant, $p>.05$) and $r(40) = 0.12$ for errors (nonsignificant, $p>.05$).

In spite of these (non)correlation patterns, recall that syllable frequencies in our materials were not significantly different between the H:M/H and L:M/L categories (i.e. the p-value was quite high by the relatively strong unpaired t-test); it is therefore not feasible to ascribe differences in responses to them solely to syllable frequency. Moreover, the above correlations were necessarily done by item, but as just noted the by-item analysis of RT between these two categories did not quite reach significance. By contrast, there was a highly significant difference in RT when analyzed by participant; in this analysis, we averaged responses to the different items in each class together, and so small variations in syllable frequency could not possibly have had any effect. There are also technical difficulties with the interpretation of the significance values associated with the correlations; Pearson's correlation coefficient assumes normal distributions for the variables, but lexical frequencies in fact do not have normal distributions because more items will be of lower frequency than of higher frequency in a random sample of types, thus seriously skewing the histogram to the right. Spearman's ranked correlation coefficient, which does not require normal distributions (but is also weaker than Pearson's), showed no significant correlations between first syllable frequency and RT, both for all items ($r_S(80)=0.26$, $p>.05$) and for items just in the H:M/H and L:M/L categories ($r_S(40)=0.31$, $p>.05$); correlations with number of errors could not be fruitfully examined due to the large proportion of tied rankings.

In summary, then, even after detailed post hoc analyses of our results, we find no reason to reject the conclusion that the negative morpheme-frequency effect is real. Nevertheless, in spite of our controls it appears that we have also uncovered the additional presence of a negative syllable-frequency effect.

3.5 Discussion

The overall results were quite compatible to the predictions of the model: MI had a positive effect on accuracy and response speed, because word frequency effects were positive and morpheme frequency effects were negative. In particular, we believe that our negative morpheme frequency effect is real, not an artifact of competition with phonologically similar words. As pointed out in the Materials section, controlling syllable frequency while varying morpheme frequency meant that words in the H:M/L category had a greater number of phonological competitors than those in the L:M/H category. The fact that L:M/H words were nevertheless harder to process shows that any negative syllable frequency effect in our results cannot be due to competition. At best, one could argue that our results were due to cross-*syllabic* predictability. We don't think that even this is correct, however, given the arguments noted in the Results section.

There are also theoretical reasons to doubt that syllable frequency was the major factor in our experiment. By definition a morpheme is a linking between form and meaning, and the processes that are activated in the lexical decision task are likely to make reference to both kinds of representation. Hence a priori a morpheme is more cognitively active than a phonological representation alone. Moreover, it has repeatedly been found that semantic representations are more active than purely phonological ones; while facilitory semantic priming is a well-established and robust phenomenon, phonological priming may have either facilitory or inhibitory effects, depending on the interstimulus interval and the task. Specifically in the domain of morphological processing, Marslen-Wilson, Tyler, Waksler, and Older (1994) found no cross-model priming with English words when prime and target were merely related in form (either phonologically or "morphologically" in a historical sense), but only when they were related via semantically transparent morphological structure. Even in the experiments of Zhou and Marslen-Wilson (1994) that ours is based on, syllable-frequency effects were quite difficult to find. In one of their experiments, syllable-frequency effects were only found in nonwords, and in another, they were only found in real words when both word and morpheme frequency were held constant. Finally, there is independent evidence for our assumptions that morphemes are accessed (as potential listemes) at some point in the processing of spoken Chinese compounds, and that syllables as purely phonological units are less important. In a set of lexical decision experiments using spoken Chinese compounds as primes and targets, Zhou and Marslen-Wilson (1995) found evidence for morphological priming that behaved distinctly from mere form priming or semantic priming. With shorter delays between prime and target, responses were facilitated by primes that shared an identical morpheme with the target, whereas responses were inhibited when the first morphemes of prime and target were homophonous or written with the same polysemous character. With longer delays, homophone and character priming effects disappeared, while morpheme priming remained, persisting even longer than word-level semantic priming. Thus it appears that even with spoken Chinese compounds, morphemes are accessed at some point, though if the overlapping listeme approach is correct not as a necessary first step in word access.

We therefore conclude that the effects of cross-morphemic predictability can be found even in the lexical access of auditorily presented Mandarin compounds, a situation where our model predicts the effects to be small. We have already given several possible reasons why Zhou and Marslen-Wilson (1994) were not able to see these effects, most notably the facts that the frequency range of our materials was much wider than theirs and that the nonwords in our experiment were designed to be as similar as possible to real words, forcing participants to adopt a strategy that minimized the interference of phonological neighbors. The relatively long RTs that we found (about 300 msec longer than those found by Zhou and Marslen-Wilson 1994) may have been a side-effect of the challenge posed by our nonwords. Given the overall long RTs (around 1 second), the size of the morpheme-frequency effect (i.e. the difference in RTs between the H:M/L and L:M/H categories, around 50 msec) comes off as somewhat small (a mere 5% or so of the average RT), which is what one would expect for spoken Chinese compounds.

4. General discussion

In this paper we have reported two findings relevant to the study of compound processing. On the theoretical side, we have demonstrated that a very simple connectionist model can produce effects of cross-morphemic predictability. If, as we have argued, cross-morphemic predictability is expected to be an important factor in segmenting morphemes from words, and words from phrases, our model provides a starting point for more sophisticated analyses of this factor. A model relying solely on cross-morphemic predictability is of course inadequate to address all issues relating to compound processing; for instance, as noted earlier, it can't deal with positive morpheme frequency effects or priming. Nevertheless, the model demonstrates that separate levels for words and morphemes aren't necessary to get separate word and morpheme effects in lexical access. Moreover, precisely the same logic leads one to conclude that negative syllable frequency effects also do not necessarily require positing a separate level for syllables. Our model gets such effects automatically, simply by overlapping representations for words and morphemes (and syllables) in an input layer, and then connecting these representations in a distributed fashion to an identical output "decision" layer.

Our model thus stands in sharp contrast to that proposed by Zhou and Marslen-Wilson (1994, 1995), who argue for a localist network with layered levels for words, morphemes and syllables, with excitatory connections between levels (i.e. between words and morphemes, and between morphemes and syllables) and inhibitory connections within each level. We feel that this model is unsatisfying, even given their own results. First, since their model was not implemented on a computer, its precise predictions are unclear. On the one hand, it might seem to predict positive morpheme frequency effects, since morphemes on one layer are linked to compounds on another, so activation should spread along the excitatory connections to speed up access for compounds with more highly activated (frequent) morphemes. On the other hand, it might predict negative morpheme frequency effects, since a higher-frequency morpheme may appear in more than one word, and these words will then compete with each other in the word layer (after all, this is how they account for their negative syllable frequency effects). On top of this, in their experiments they did not in fact find morpheme frequency effects at all, a null result that doesn't provide support for their multiple layers.

Second, they claim that syllable frequency has a negative effect by creating competition between morphemes (i.e. homophonous morphemes inhibit each other). However, despite the presence of more homophones in the H:M/L condition in our experiment, these compounds were actually easier to access than the items in the L:M/H condition. This result makes sense in our model, since given the way we designed our materials, the words with more homophones (i.e. the H:M/L words) also had higher cross-morphemic predictability (and to some extent higher cross-syllabic predictability as well) than the L:M/H words.

Turning to the empirical results of our paper, our major finding was that the predictions of our model were in fact correct, at least for spoken Mandarin Chinese: morpheme frequency has a negative effect on accuracy and response speed. Our results do not contradict those of Zhou and Marslen-Wilson (1994) since null results by their nature cannot be contradicted, and since differences in our materials and design are sufficient to explain why we found them when they didn't. Our results also replicated those of Chen (1993), Tsai (1994), Lü (1996), and Peng, Liu and Wang (1999), all of whom found negative morpheme frequency effects for Chinese compounds (when written and semantically opaque). For us the effect was quite small, just as expected given the high degree of homophony in spoken Chinese, which should reduce the effect of cross-morphemic predictability.

If the overlapping listeme approach is on the right track, negative morpheme frequency effects should not be restricted to Chinese. That they have not been reported for other languages is perhaps due to the lack of anyone systematically looking for them. By comparing the differences between our studies and those of others (other than language), we extract the following tips for those interested in seeking negative morpheme frequency effects. First, the effects may be small, so it is important that the categories of low and high frequency morphemes be as different as possible. Second, one should control, or at least understand, the role of syllable frequency (or perhaps more generally cohort size) in obscuring cross-morphemic predictability effects. This may even be important for reading studies. Third, and perhaps most important, the effect of cross-morphemic predictability may only be clearly revealed with compounds whose MI values are higher than average. This may help explain why morpheme frequency effects with transparent Chinese compounds are positive when written (Chinese orthography makes such compounds "obviously" decomposable) yet negative when spoken (homophony hides the morphological constituents and so makes semantic transparency less relevant).

Beyond our specific model and experimental results, we would like to leave the reader with more general thoughts inspired by the wordhood problem in Chinese. If Chinese speakers are right in their traditional assumption that words are fuzzy, and not the sole parsing goal in language processing, could it be that the same is true for all languages to varying extents? Could it be that readers of word-segmented orthographies have also been fooled by their writing systems, making them think that their mental lexicon is something like the word-based dictionaries they are familiar with, just as Chinese readers are fooled into thinking that their mental lexicon consists of nothing but characters?

One observation that may be of relevance is the finding that readers of English compounds with ambiguous segmentation (e.g. *clamprod*, which can be read as *clam* + *prod* or *clamp* + *rod*) generate both possible parses rather than following a uniform parsing strategy (Libben, Derwing, and de Almeida 1999). This is consistent with the overlapping listeme approach, since it suggests that the parsing mechanism is quite accepting of ambiguity, something we have claimed to be true for the deeper ambiguity between a compound parsed as a whole and as parsed into morphemes. Spoken language provides even greater challenges to a parser, since not even word boundaries are provided unambiguously in the signal (e.g. *recognize speech* sounds a lot like *wreck a nice beach*; Shillcock 1990). We suspect that our understanding of such problems would benefit from rethinking the assumption that the ultimate targets of lexical access must always be words (or always morphemes), and instead consider the possibility that the parser can be satisfied if it finds any sensible listeme (morpheme, word and/or phrase). At the very least, such a view would help integrate research on compound processing and word processing into the more general study of the processing of language in real time.

Tables and figures

Figure 1.

$$MI\,(x, y) = \log_2\left(\frac{prob\,(x, y)}{prob\,(x)\,prob\,(y)}\right)$$

Table 1 Frequency values for four word types in connectionist model experiment

Word type	H:H/M		H:M/L		L:L/M		L:M/H	
Word label	AB		CD		EF		GH	
Word frequency	1800		54		5		46	
Morpheme frequencies	A	B	C	D	E	F	G	H
	2008	2013	296	306	2019	1998	5720	5706
MI value	3.16		3.58		-5.33		-5.14	

Table 2 Results of connectionist model experiment

Word type	H:H/M	H:M/L	L:L/M	L:M/H
Mean error value	0.0426	0.3519	2.4347	2.2076

Table 3 Mean frequencies for materials used in auditory lexical decision experiment*

Word type	H:H/M		H:M/L		L:L/M		L:M/H	
Mean transparency score	1.35		1.34		1.31		1.29	
Mean word frequency	47.29		3.18		0.41		3.83	
Mean morpheme frequencies	1st	2nd	1st	2nd	1st	2nd	1st	2nd
	338	359	30	40	332	334	1373	1483
Mean syllable frequencies	1st	2nd	1st	2nd	1st	2nd	1st	2nd
	1552	1241	1896	2116	1748	2177	2099	4016
Mean cross-morphemic MI value	7.61		11.45		1.47		0.55	

*To ease comparison with other studies, word frequencies are given as a proportion of the number of word tokens out of 1,000,000 (actual corpus size: 9,529,233 words), while morpheme and syllable frequencies are given as a proportion of the number of character tokens out of 1,000,000 (actual corpus size: 14,457,534 characters), since morphemes are almost always monosyllabic and coextensive with orthographic characters. MI values were calculated using word frequencies as proportion of the number of character tokens; see text for explanation.

Table 4 Results of auditory lexical decision experiment

Word type	H:H/M	H:M/L	L:L/M	L:M/H
Mean number of errors	**1.28**	**1.00**	**4.80**	**3.04**
Mean RT (msec)	1011.46	993.56	1114.89	1042.17

Appendix

Stimuli for experiment: H:H/M items.*

Noun	ST	WF	1st MF	2nd MF	1st SF	2nd SF	MI
fan4zui4	1.1	1114	6095	5113	16378	27370	9.01349
pi1pan4	1.1	140	4075	5230	24402	7182	6.56942
cai2zhi2	1.3	126	2617	7533	27108	41074	6.52989
tu2po4	1.5	1308	4470	6198	19680	8978	9.41482
fu4zhai4	1.4	112	7381	2041	43366	11566	6.74801
zhi2ban1	2	139	6660	5168	41074	11684	5.86756
xiu1hu4	1.1	148	7684	7500	11558	17069	5.21445
xun4lian4	1.1	1417	3592	3141	14052	4498	10.8264
ting2zhi2	1	122	6777	7660	12168	41074	5.0865
zhen1ce4	1	105	5071	2573	18943	11706	6.86227
kong4su4	1.7	120	3082	4240	12417	20171	7.05271
bo2ai4	1.1	134	2268	5625	13275	7709	7.24657
ping2ku1	1	1546	4561	3850	25872	39959	10.3139
jie1ji2	1.8	129	2941	7156	36494	11928	6.46952
tu2pian4	2	100	6106	4830	19680	7205	5.61536
fan4wei2	2	1383	3581	3532	16378	110102	10.6265
gi4bu3	1.3	128	2037	3229	35429	11743	8.13622
xiao1shou4	1.1	1624	7853	6381	20225	28241	8.87206
gou4mai3	1	1627	6092	5892	13677	5892	9.35607
yan2mi4	1.3	147	7066	3656	34245	7676	6.36226

*Nouns are written in the Pinyin romanization; digits represent tone categories. ST = mean semantic transparency score (1=most transparent, 5=least transparent), WF = word frequency (number of tokens in a corpus of 9,529,233 words), MF = morpheme frequency (number of tokens in a corpus of 14,457,534 characters), SF = syllable frequency (number of tokens in a corpus of 14,457,534 characters), MI = cross-morphemic mutual information value, estimated as $\log_2((WF/14457534)/((MF1/14457534)*(MF2/14457534)))$.

For a list of the materials written in Chinese characters, please contact the first author.

Stimuli for experiment: H:M/L items.

Noun	ST	WF	1st MF	2nd MF	1st SF	2nd SF	MI
ban4zou4	1.5	25	110	871	29188	1165	11.8813
bei4ke2	1.8	20	957	349	34765	396	9.75779
ru3yi4	1.9	20	518	585	16010	279354	9.89815
yin1yuan2	1.8	24	417	854	47267	117168	9.92828
zhuo2zhuang4	1.7	36	39	619	13245	6265	14.396
bao3lei3	1.7	20	294	621	24710	2310	10.6291
zi1run4	1.5	20	609	937	29344	943	8.98504
shu1chang4	1.2	20	609	937	29344	943	8.98504
xu1wei4	1.3	25	374	625	16047	3366	10.5946
jing4pei4	1.7	58	954	510	21382	8411	10.7511
fu4xie4	1.1	32	546	95	43377	26230	13.1227
xi2mie4	1	42	129	915	17873	972	12.3288
shui4mian2	1.1	49	830	251	33867	912	11.7315
fu3xiu3	1	29	398	77	33024	996	13.7399
fu3wei4	1.2	24	483	950	33024	150093	9.56262
gu3cang1	1	28	159	624	26426	1441	11.9944
gao1bing3	1	44	246	361	36426	4380	12.8064
chu2chuang1	1.3	46	86	822	15199	4509	13.1996
chu4mo1	1	20	936	535	26503	483	9.17349
zhan4dou3	1	24	58	121	21284	1516	15.5934

Stimuli for experiment: L:L/M items.

Noun	ST	WF	1st MF	2nd MF	1st SF	2nd SF	MI
shi1san4	1.4	4	7928	2110	32327	2110	1.78955
shou3jiu4	1.1	3	3328	2351	25866	35004	2.47078
se4bi3	1	5	6966	2717	12414	16188	1.93333
xi1pan2	1.3	2	4920	6454	33681	11086	-0.1351
zhao1ling3	1.6	3	2935	5663	23066	6916	1.38379
chen2ji1	1.1	3	2146	6037	26054	126777	1.74323
xing1he2	2	3	4335	3006	15319	72139	1.73484
yin1pin2	1	5	4494	2310	47267	2941	2.79978
fei1dao1	1.3	4	3465	2330	17469	2349	2.84056
zhen4dang4	1.3	4	2533	3933	18606	28110	2.53727
bing4rong2	1	2	4868	7277	39597	19183	-0.2929
jiu3gian2	1	5	3348	4668	33234	58414	2.20957
ma3ti2	1.1	5	6476	3932	10444	36965	1.50531
chuan2huo4	1.4	5	6311	7474	15353	62946	0.61593
ye3hua1	1.2	5	2291	7458	39811	21610	2.08091
jian3ya1	1	4	5953	4814	22433	7250	1.0129
jue2ji4	1.5	5	5077	7045	32713	78222	1.0151
shang1bing1	1	3	5595	2669	27310	3484	1.53827
yan3suan4	2	5	7495	7318	12798	7457	0.3983
yan4fu2	1.8	3	5590	7057	19777	30272	0.13681

Stimuli for experiment: L:M/H items.

Noun	ST	WF	1st MF	2nd MF	1st SF	2nd SF	MI
ben3neng2	1.2	28	26639	37889	26769	37889	-1.3181
ming2li4	1	21	25124	21623	49303	98237	-0.8394
yin1guo3	1.5	45	41178	15486	47267	15681	0.02891
shou1jian4	1.2	49	14120	11783	14120	91005	2.09016
kao3zheng4	1.6	32	10049	11413	10484	19777	2.01216
qiu2cai2	1	53	14712	12526	23229	27108	2.05589
fang4kai1	1.2	22	10893	32854	10893	32862	-0.1701
wu4ti3	1.9	21	11847	17092	38243	17092	0.58437
shi1jia1	1.6	59	12164	28555	32327	72195	1.29618
ju1ying2	1	36	11395	14537	23895	19959	1.65169
wu4nong2	1.2	35	20257	12696	38243	13974	0.97638
tong1shang1	1.3	24	20141	21206	20214	27310	-0.2998
gang3wan1	1.4	23	10324	15733	10661	16288	1.03366
wu2qing2	1.1	55	25155	22393	35890	24241	0.49736
fa1wen4	1.7	22	44844	18165	45314	48851	-1.3568
lu4fei4	1	28	24029	15385	44723	18908	0.13095
guo4shi2	1.3	48	32398	61604	32398	197865	-1.5241
ren4qing1	1	57	16948	11820	36330	25965	2.04044
zeng1gao1	1.1	50	10587	36004	21253	36426	0.9233
guan2shi4	1.4	22	14201	10155	45296	319612	1.14114

Bibliography

Anderson, Stephen R. 1992. A-Morphous Morphology. Cambridge: Cambridge University Press.

Andrews, Sally. 1986. "Morphological Influences on Lexical Access: Lexical or Nonlexical Effects?" Journal of Memory and Language 25, 726-740.

Bochner, Harry. 1993. Simplicity in Generative Grammar. New York: Mouton de Gruyter.

Butterworth, Brian. 1983. "Lexical Representation." In: B. Butterworth (ed.) Language Production, Vol. II: Development, Writing, and Other Language Processes. London: Academic Press, 257-294.

Bybee, Joan L. 1985. Morphology: A Study of the Relation between Meaning and Form. Amsterdam: John Benjamins Publishing Company.

Chen, Hsuan-Chih & Ovid J. L. Tzeng (eds.) 1992. Language Processing in Chinese. Amsterdam: North-Holland.

Chen, Shao-Tsu. 1993. Hanyu Gouci Zai Yuedu Licheng Zhong Dui Yuyi Chufa Xiaogai De Yingxiang. [The Influence of Semantic Activation Effects on Chinese Morphology During Reading.] Unpublished master's thesis, National Chung Cheng University, Taiwan.

Chen, Keh-Jiann, Chu-Ren Huang, Li-PingChang, & Hui-Li Hsu. 1996. "SINICA CORPUS: Design Methodology for Balanced Corpora." Language, Information and Computation 11, 167-176.

Cheng, Chao-Ming. 1981. "Hanzi Renzhi De Licheng." [Mandarin Word Recognition.] Zhonghua Xinlixue Kan 23, 137-153.

CKIP. 1993. Xinwen Yuliao Cipin Tongjibiao: Yuliaoku Wei Ben Yanjiu Xilie Zhi Er [Corpus-Based Frequency Count of Words in Journal Chinese: Corpus-Based Research Series No. 2.] Taipei: Academia Sinica.

Church, Kenneth Ward & Patrick Hanks. 1990. "Word Association Norms, Mutual Information, and Lexicography." Computational Linguistics 16.1, 22-29.

DeFrancis, John. 1989. Visible Speech: the Diverse Oneness of Writing Systems. Honolulu: University of Hawaii Press.

DeFrancis, John (ed.). 1996. ABC Chinese-English Dictionary. Honolulu: University of Hawaii Press.

Di Sciullo, Anna-Maria & Edwin Williams. 1987. On Defining the Word. Cambridge, MA: MIT Press.

Duanmu, San. 1998. "Wordhood in Chinese." In: Packard 1998, 135-196.

Fabb, Nigel. 1998. "Compounding." In: Spencer & Zwicky 1998, 66-83.

Feldman, Laurie Beth (ed.). 1995. Morphological Aspects of Language Processing. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Forster, Kenneth I. & Susan M. Chambers. 1973. "Lexical Access and Naming Time." Journal of Verbal Learning and Verbal Behavior 12.6, 627-635.

Gong, Shuping. 1999. The Role of Cooccurrence Probabilities in the Lexical Access of Spoken Disyllabic Words in Mandarin. Unpublished master's thesis, National Chung Cheng University, Taiwan.

Ho, Kwok-cheung. 1976. "A Study of the Relative Frequency Distribution of Syllabic Components in Mandarin Chinese." Journal of the Institute of Chinese Studies 8, 275-352.

Hoosain, Rumjahn. 1992. "Psychological Reality of the Word in Chinese." In Chen & Tzeng 1992, 111-130..

Huang, Chu-Ren, Kathleen Ahrens, & Ken-Jiann. Chen. 1998. "A Data-Driven Approach to the Mental Lexicon: Two Studies on Chinese Corpus Linguistics. Bulletin of the Institute of History and Philology 69.1, 151-180.

Huang, Jong-Tsun. 1979. "Zhongwen Ci Lianxiang De Yishixing Ke Fenli Jiashe" [Time-Dependent Separability Hypothesis of Chinese Words Association.] Zhonghua Xinlixue Kan 21, 41-48.

Inhoff, Albrecht Werner, Deborah Briihl, & Jill Schwartz. 1996. "Compound Word Effects Differ in Reading, On-Line Naming, and Delayed Naming Tasks." Memory and Cognition 24.4, 466-476.

Jarema, Gonia, Céline Busson, Rossitza Nikolova, Kyrana Tsapkini, & Gary Libben. 1999. "Processing Compounds: a Cross-Linguistic Study." Brain and Language 68, 362-369.

Kehayia, Eva, Gonia Jarema, Kyrana Tsapkini, Danuta Perlak, Angela Ralli, & Danuta Kadzielawa. 1999. "The Role of Morphological Structure in the Processing of Compounds: the Interface between Linguistics and Psycholinguistics." Brain and Language 68, 370-377.

Lee, Chia-Ying. 1995. Hanyu Zuheci Yu Chengyuci Zai Xinli Cidian Zhong De Biaozheng Fangshi. [The Representation of Compounds and Idioms in the Chinese Mental Lexicon.] Unpublished master's thesis, National Chung Cheng University, Taiwan.

Liang, M. Y. 1992. Recognition processing in reading compositional and idiomatic words. Unpublished master's thesis, National Tsinghua University, Hsinchu, Taiwan.

Libben, Gary. 1998. "Semantic Transparency in the Processing of Compounds: Consequences for Representation, Processing, and Impairment." Brain and Language 61, 30-44.

Libben, Gary, Bruce L. Derwing, & Roberto G. de Albeida. 1999. "Ambiguous Novel Compounds and Models of Morphological Parsing." Brain and Language 68, 378-386.

Liu, In-Mao. 1988. "Context Effects on Word/Character Naming: Alphabetic Versus Logographic Languages." In: In-Mao Liu, Hsuan-Chih Chen, & May Jane Chen (eds.), Cognitive Aspects of the Chinese Language Vol. 1. Hong Kong: Asian Research Service, 81-92.

Liu, Ying, & Dan-ling Peng. 1997. "Meaning Access of Chinese Compounds and Its Time Course." In: Hsuan-Chih Chen (ed.), Cognitive Processing of Chinese and Related Asian Languages. Hong Kong: The Chinese University Press, 219-232.

Lü, Ching-Ching. 1996. Zhongwen Yuci Renzhi Licheng. [Chinese Word Recognition.] PhD thesis, National Tsing Hua University, Taiwan.

Marslen-Wilson, William, Lorraine Komisarjevsky Tyler, Rachelle Waksler, & Lianne Older. 1994. "Morphology and Meaning in the English Lexicon." Psychological Review 101.1, 3-33.

Packard, Jerome L. (ed.) 1998. New Approaches to Chinese Word Formation. Berlin: Mouton de Gruyter.

Packard, Jerome L. 1999. "Lexical Access in Chinese Speech Comprehension and Production." Brain and Language 68, 89-94.

Peng, Danling, Ying Liu, & Chunmao Wang. 1999. "How Is Access Representation Organized? The Relation of Polymorphemic Words and Their Morphemes in Chinese." In J. Wang, A. W. Inhoff, & H.-C. Chen (eds.). Reading Chinese Script: A Cognitive Analysis. Mahwah, New Jersey: Lawrence Erlbaum Associates, 65-89.

Plunkett, Kim & Jeffrey L. Elman. 1997. Exercises in Rethinking Innateness: a Handbook for Connectionist Simulations. Cambridge, MA: MIT Press.

Rosenblatt, Frank (1958) "The perceptron: a probabilistic model for information storage and organization in the brain." Psychological Review 65:386-408.

Rueckl, Jay G. & Michal Raveh. 1999. "The Influence of Morphological Regularities on the Dynamics of a Connectionist Network." Brain and Language 68, 110-117.

Rumelhart, D. E., G. E. Hinton, & R. J. Williams. 1986. "Learning Internal Representations by Error Propagation." In David E. Rumelhart, James L. McClelland, & the PDP Research Group (eds.). Parallel Distributed Processing Vol. 1. Cambridge, MA: MIT Press, 318-362.

Sandra, Dominiek. 1990. "On the Representation and Processing of Compound Words: Automatic Access to Constituent Morphemes Does Not Occur." Quarterly Journal of Experimental Psychology: Human Experimental Psychology 42(3-A), 529-567.

Schreuder, Robert & R. Harald Baayen. 1997. "How Complex Simplex Words Can Be." Journal of Memory and Language 37, 118-139.

Shillcock, Richard. 1990. "Lexical Hypotheses in Continuous Speech." In: Gerry T. M. Altmann (ed.) Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives. Cambridge, MA: MIT Press, 24-49.

Smith, Philip T. 1995. "Are Morphemes Really Necessary?" In: Feldman (ed.) 1995, 365-382.

Spencer, Andrew & Arnold M. Zwicky (eds.) 1998. The Handbook of Morphology. Blackwell

Sproat, Richard & Chilin Shih. 1990. "A Statistical Method for Finding Word Boundaries in Chinese

Text." Computer Processing of Chinese and Oriental Languages 4.4, 336-351.

Starosta, Stanley, Koenraad Kuiper, Siew-ai Ng, & Zhi-qian Wu. 1998. "On Defining the Chinese Compound Word: Headedness in Chinese Compounding and Chinese VR Compounds." In: Packard 1998, 347-370.

Stemberger, Joseph Paul. 1998. Morphology in Language Production with Special Reference to Connectionism. In: Spencer & Zwicky (eds.) 1998, 428-452.

Su, Y.-C. 1998. "The Representation of Compounds and Phrases in the Mental Lexicon: Evidence from Chinese." University of Maryland Working Papers in Linguistics 6, 179-199.

Taft, Marcus & Chen, Hsuan-Chih. 1992. "Judging Homophony in Chinese: the Influence of Tones." In: Chen & Tzeng (eds.) 1992, 151-172.

Taft, Marcus & Kenneth I. Forster. 1976. "Lexical Storage and Retrieval of Polymorphemic and Polysyllabic Words." Journal of Verbal Learning and Verbal Behavior 15.6:607-620.

Taft, Marcus & Xiaoping Zhu. 1995. "The Reprsentation of Bound Morphemes in the Lexicon: a Chinese Study." In: Feldman (ed.) 1995, 293-316.

Tsai, Chih-Hao. 1994. Effects of Semantic Transparency on the Recognition of Chinese Two-Character Words: Evidence for a Dual-Process Model. Unpublished master's thesis, National Chung Cheng University, Taiwan.

Tsai, Chih-Hao, George W. McConkie, & Xianjun Zheng. 1998. "Lexical Parsing by Chinese Readers." Poster presented at the Advanced Study Institute on Advances in Theoretical Issues and Cognitive Neuroscience Research of the Chinese Language, University of Hong Kong.

Tsay, Jane, Myers, James, & Xiao-Jun Chen. 1999. "Tone Sandhi as Evidence for Segmentation in Taiwanese." Proceedings of the 30th Annual Child Language Research Forum.

van Jaarsveld, Henk J. & Gilbert E. Rattink. 1988. "Frequency Effects in the Processing of Lexicalized and Novel Nominal Compounds." Journal of Psycholinguistic Research 17.6, 393-422.

Wang, Mei-Chu, Chu-Ren Huang, & Keh-Jiann Chen. 1994. "The Identification and Classification of Unknown Words in Chinese." In: Akira Ishikawa & Yoshihiko Nitta (eds.) The Proceedings of the 1994 Kyoto Conference of the Logico-Linguistics Society of Japan: A Festschrift for Professor Akira Ikeya. Tokyo: The Logical-Linguistics Society of Japan, 113-123.

Whaley, C. P. 1978. "Word Non-Word Classification Time." Journal of Verbal Learning and Verbal Behavior 17.2, 143-154.

Ye, Yun & Cynthia M. Connine. 1999. "Processing Spoken Chinese: the Role of Tone Information." Language and Cognitive Processes 14.5/6, 609-630.

Zhou, Xiaolin. & William Marslen-Wilson. 1994. "Words, Morphemes, and Syllables in the Chinese Mental Lexicon." Language and Cognitive Processes 9.3, 393-422.

Zhou, Xiaolin, & William Marslen-Wilson. 1995. "Morphological Structure in the Chinese Mental Lexicon." Language and Cognitive Processes 10.6, 545-600.

Zwitserlood, Pienie. 1994. "The Role of Semantic Transparency in the Processing and Representation of Dutch Compounds." Language and Cognitive Processes 9.3, 341-368.