Reading and Writing: An Interdisciplinary Journal **10:** 223–243, 1998.
C.K. Leong & K. Tamaoka (eds.), Cognitive Processing of the Chinese and the Japanese Languages, pp. [69–89]

Context effects and the processing of spoken homophones

PING LI[1] & MICHAEL C. YIP[2]
[1] *University of Richmond, Virginia, USA;* [2] *Chinese University of Hong Kong, Hong Kong, PR China*

Abstract. This study examined the role of context effects in the processing of homophones in Chinese and Chinese-English bilingual speech. In Experiment 1, Chinese speakers were presented with successively gated portions of a homophone in a sentence context, and they identified the homophone on the basis of its increasing acoustic information. In Experiment 2, Chinese-English bilinguals were presented with a cross-language homophone in a sentence context, and they named a visual probe that had or did not have phonological overlaps with the homophone. Results indicate that prior sentence context has an early effect on the disambiguation of various homophone meanings, shortly after the acoustic onset of the word, in both monolingual and bilingual situations. The results are accounted for by interactive activation models of lexical processing, in which the recognition of a homophone is a result of the interactions among phonological, lexical, and contextual information at an early stage.

Key words: Processing Chinese homophones, Context effects, Gating paradigm, Interactive activation models, Multiple constraints

Introduction

Imagine the following scenario for a second language learner of Chinese. The learner begins with a Chinese-English dictionary in her pocket. Her dictionary is organized according to the phonetic transcription system, the Pinyin, which starts with *a* and ends with *zuo*. But she soon discovers that this dictionary is very 'user-unfriendly', because unlike an English dictionary, the same entry, for example, *yi*, occurs in the dictionary for a multiple number of times, referring to different meanings or different words, for example, 'hundred-million', 'skill' and 'easy'. The learner is told that if she learns to read Chinese characters, the problem will go away because Chinese characters differentiate these different meanings or words. So she learns Chinese characters, which turns out to be very helpful. She can now use the dictionary more effectively since the multiple entries of *yi* are accompanied by different Chinese characters. The learner is satisfied, until she realizes that this problem still exists in the speech: she can use no comparable characters in the speech to differentiate the various meanings of *yi* that she hears. She therefore asks her language teacher how to solve this problem. The teacher has never thought that this can be a problem. He gives the learner his simple and straightforward answer:

you know immediately that *yi* refers to 'hundred-million' if you hear the prior sentence context as being related to money, 'skill' as being related to talent, and 'easy' as being related to task difficulty.

Although our imagined teacher has not given any serious thought in answering what to him is an uninteresting question, his answer is relevant to one of the central debates in sentence processing: when speakers see or hear an ambiguous item that has multiple meanings, do they use prior sentence context to help them eliminate irrelevant meanings right away, or do they first activate all possible candidate meanings in the mental dictionary, irrespective of the prior context? It seems that, after more than two decades of psycholinguistic research, there still is no definitive answer.

Two hypotheses, one in direct contrast to the answer from our imagined naive language teacher, and one similar to it, have emerged in the last twenty years or so (Onifer & Swinney 1981; Small, Cottrell & Tanenhaus 1988; Simpson & Krueger 1991; Swinney 1979; Tabossi 1988). The exhaustive access hypothesis argues that all meanings of an ambiguous word will be accessed momentarily following the occurrence of the word, and that semantic context can only help to select the appropriate meaning at a post-access stage. This hypothesis assumes that sentence processing is a modular, bottom-up process in which non-lexical, contextual information does not penetrate lexical access (cf. Fodor 1983). By contrast, the context-dependency hypothesis argues that the contextually appropriate meaning of an ambiguous word can be selectively accessed early on if sentence context provides a semantic bias to the appropriate meaning. This hypothesis assumes that sentence processing is an interactive process in which information can flow both bottom-up and top-down and that lexical access and sentential context can mutually influence one another at a very early stage (McClelland 1987).

The above hypotheses have been mainly tested in English and several Indo-European languages (e.g., Dutch and Italian). However, they have not been, to our knowledge, systematically examined in Chinese, a language that offers unique features in crosslinguistic studies of sentence processing (Li 1994, 1996a, 1998; Li, Bates & MacWhinney 1993). Chinese differs significantly from most Indo-European languages in its phonological, lexical, and syntactic structures. For example, Chinese involves a tonal system, and different tones distinguish between different meanings associated with the same syllables. But tonal information alone does not eliminate lexical ambiguities associated with homophones: Chinese has a massive number of homophones on the lexical-morphemic level even with tonal distinctions. In English and other Indo-European languages, homophony is a relatively low-frequency event. In Chinese, homophony is extensive. In this study, we are particularly concerned with phonological and lexical properties of Chinese such as those associated with the extensive homophony, and the processing consequences thereof.

According to the Modern Chinese Dictionary (Institute of Linguistics, 1985), 80 percent of the monosyllables (differentiated by tones) in Chinese are ambiguous between different meanings, and 55 percent have five or more homophones. The single syllable *yi* with the dipping tone has up to 90 homophones (e.g., skill, justice, benefit, discuss, intention, translate, hundred-million, etc.), and this number would increase to 171 if identical syllables with different tones were considered as homophones. Upon hearing *yi* in a sentence, do Chinese speakers activate all 90 or more meanings of the syllable? The exhaustive access hypothesis predicts that they should, because lexical access is a modular, autonomous, and capacity-free process. The context-dependency hypothesis predicts that they activate only the contextually appropriate meaning with aid from prior sentence context.

Recently we have explored the processing of Chinese homophones in a number of studies. Using cross-modal and gating paradigms, Li (1998) and Li & Yip (1996) examined the effects of sentence context on Cantonese Chinese speakers' access and selection of homophone meanings. The cross-modal experiment showed that context effects can occur immediately following the occurrence of the homophone, and the gating experiment showed that listeners can recognize the appropriate meaning with less than half of the acoustic information of the homophone. These experiments indicate that Chinese speakers are sensitive to the contextually biased meaning at an early stage, probably within the acoustic boundary of the spoken homophone. The results point to a much earlier context effect than what has been previously assumed (e.g., about 1.5 seconds following the occurrence of the ambiguous word, e.g., as in Onifer & Swinney 1981). It seems that Chinese listeners, to cope with the extensive ambiguity created by massive homophony, must rapidly disambiguate alternative homophone meanings during sentence processing. Our results provide evidence for the context-dependency hypothesis that ambiguous meanings of a word may be selectively accessed at an early stage according to prior sentence context (Simpson 1981; Simpson & Krueger 1991; Tabossi 1988).

In this paper, we extend our work in homophone processing along the following lines. First, as discussed above, lexical tones in Chinese can differentiate alternative meanings associated with the same syllable and thus reduces the potential number of homophones, but it does not eliminate homophony. Some interesting questions arise there: how does tonal information interact with sentence context to disambiguate homophone meanings? When does tonal information start to play a role in differentiating alternative meanings for the syllable? Does sentence context outweigh lexical tone during sentence processing to produce garden-path effects? Experiment 1 was designed to answer these questions. Second, to understand homophone processing across situations, we examined 'cross-language homophones' in bilingual lexical processing. Cross-language homophones refer to homophones that share phonological similarities across two or more languages. For example, word pairs like *cite-sit, note-knot,* and *pique-pick* are considered French-English homophones because they sound similar and sometimes result in misinterpretations in code-switched conversations (Grosjean 1988).[1] Experiment 2 was designed to determine the role of context effects in bilingual speakers' processing of Chinese-English homophones such as *lok-lock, sik-sit,* and *fei-fail* in code-switch situations.

Experiment 1

In this experiment we adapted the gating paradigm of Grosjean (1980, 1988) to investigate the role of context and tonal information in homophone processing. The gating paradigm has been applied to the study of monolingual and bilingual spoken word recognition over the past decade (Cotton & Grosjean 1984; Grosjean 1980; Grosjean 1988; Tyler & Wessels 1985) and has proven to be particularly useful in assessing the amount of phonetic-acoustic information needed for the correct identification of a word. In the gating task, listeners are presented with fragments of a word, one at a time in increasing duration, until the whole word has been presented. At each presentation, listeners are required to identify the word being presented on the basis of the information up to that point. Our adaptation involves asking listeners to identify homophones that either match or do not match the sentence context. Homophones that match and those that do not match differ in the particular lexical tone that they carry: the former carry the correct tone, and the latter the incorrect tone.

Participants. Twenty native Cantonese Chinese speakers (6 men and 14 women, mean age = 19.85) who reported no speech or hearing deficits participated in this experiment. All participants were students at the Chinese University of Hong Kong. They took part in the experiment as a laboratory requirement for credit in an introductory psychology course.

Materials and design. Sixty spoken homophones (see Appendix 1) were selected. Each homophone was preceded by a sentence context that was biased toward one of the meanings of the homophone. The homophone either matched the sentence context or did not match because of an incorrect tone, creating a total of 2 $\times$ 60 sentence materials. A separate group of 18 native speakers were asked to judge the degree of constraint of the prior context on the target homophone. They were given the test sentences with the biasing context but without the target homophone, and were asked to fill in a word that they thought would naturally complete the sentence. Their responses were scored on a 1 to 4 scale, based on the scale used by Marslen-Wilson and Welsh (1978): 1 was given for a word identical to the correct homophone, 2 for a synonym, 3 for a related word, and 4 for an unrelated word. Responses were pooled across the 18 judges, and the mean rating was 1.5. This score was above the high constraint condition in Marslen-Wilson and Welsh (1978). The average length of the test sentences, counting the homophone, was 15 words (ranging from 12 to 18 words).

Two independent variables were manipulated in this experiment. (1) Match to context: the homophone carried either the correct tone that fit the sentence context, or the incorrect tone that did not fit the context. (2) Homophone density: high density – half of the homophones were associated with five to six tones, and low density – half of them with one to two tones only. The number of tones associated with each homophone was determined on the basis of the LSHK Cantonese Romanization Scheme (Linguistic Society of Hong Kong, 1994). Li and Yip (1996) introduced the notion of homophone density and defined it as the number of lexical meanings within a given homophone. Here we use the notion somewhat differently to refer to the number of lexical tones associated with a particular syllable. In other words, our notion of homophone here is broadened to refer not only to identical syllables with a specific tone (as in Li & Yip 1996), but to identical syllables with different tones. This broadened notion of homophone enables us to examine how tonal information can help to narrow the range of semantic candidacy and how it interacts with context effects.

Procedure. The test sentences were read by a native Cantonese speaker at a normal conversation rate, first tape-recorded and then digitized into a PowerMac computer. A sampling rate of 22kHZ with a 16-bit sound format was used for digitizing. Each homophone was gated and presented to listeners as follows (see Grosjean 1980, 1988; Li 1996b). Listeners heard a gated portion of the homophone, one at a time in increasing length, until the whole homophone was presented. The first gate contained the preceding context up to, but not including the homophone. The second gate contained the first gate plus the first 40 msec of the homophone, and so on, until the last gate reached the end of the homophone. The presentation of successive gates was controlled by the PsyScope program (Cohen, MacWhinney, Flatt & Provost 1993). Listeners were asked to identify, for each presentation, the last word of the sentence that was being presented (all homophones occurred at the end of the sentence). They wrote down the answer in Chinese characters and then pressed the computer spacebar to hear the next gate.

The 20 participants were randomly assigned to three groups of seven, seven, and six each. Each group received an equal number of sentences in the 2 (match to context) $\times$ 2 (homophone density) design. Each listener received 40 sentences with about 300 gates in total (i.e., an average of 7.7 gates for each sentence). The order of presentation for the sentences was pseudo-randomly arranged such that no participant heard the same homophone twice across the four conditions.

Data analysis. The dependent variable was the amount of acoustic information that listeners needed to arrive at the identification point: the point at which listeners correctly identify the homophone and do not subsequently change their minds (Grosjean 1980). This correct identification was expressed as percent of homophone needed, that is, the identification time divided by the total length of the homophone.

Results and discussion

Listeners in this experiment identified the homophones at one of the two locations: (a) before the acoustic offset of the homophone, (b) after the acoustic offset of the homophone, that is, never within the sentence frame. The results indicate that 80 percent of the homophones were correctly identified at point (a), and 20 percent at point (b). All the homophones in (b) were those that did not match the sentence context (i.e., occurred with the incorrect tone). Figure 1 presents the mean percent of homophone needed for the correct identification of the homophone, as a function of match to context and homophone density. Following Grosjean (1980), we used the total length of the homophone as the percent needed for those homophones that were identified after their acoustic offset.

A 2 $\times$ 2 (match to context by homophone density) ANOVA on the data in Figure 1 revealed several interesting results.

First, there was a main effect of match to context [$F(1,76) = 1814.37$; $p < 0.01$], indicating that the listeners could identify the homophone with much less acoustic signal if the homophone had the correct tone for the context than if the homophone had an incorrect tone. Collapsed over levels of homophone density, listeners needed on the average only 33 percent of the word for the homophone that matched the context, but 82 percent of the word for the homophone that did not match. This result showed a clear context effect at an early stage of word identification. It is consistent with our previous results from both cross-modal and gating studies in which listeners

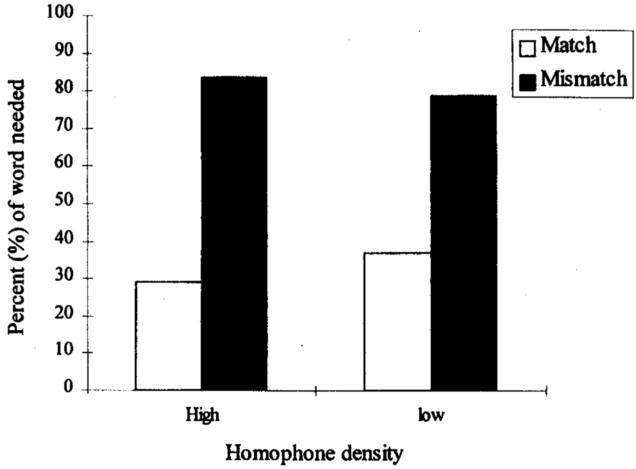

Figure 1. Percent of word needed as a function of match to context and homophone density (Experiment 1).

need only less than half of the acoustic information for the recognition of Chinese homophones (Li 1998; Li & Yip 1996).

Second, there was no main effect of homophone density [$F(1,76) = 2.40$; $p > 0.05$], indicating that whether a homophone was associated with many tones or with only one or two tones had little effect in general on the amount of acoustic signal required for the identification of the homophone. The listeners needed a similar amount of acoustic information to identify the homophone for both kinds of items. Collapsed over levels of match to context, high density items required 56 percent of the homophone for successful identification, while low density items required 58 percent.

Third, there was a significant interaction between match to context and homophone density [$F(1,76) = 32.62$; $p < 0.01$]. This interaction shows that the high density homophones required less acoustic information than the low density items when the homophone matched the context, but the reverse was true when the homophone did not match the context. However, this interaction went in a direction somewhat different than we expected, because we would expect that the more tones a homophone is associated with, the more acoustic information it would require for identification.

To understand this interaction more clearly, we examined all the word candidates that listeners proposed during various stages of the identification process. Recall that listeners wrote down their answers in response to each auditory gate of the homophone. Thus, there could be multiple candidates proposed by listeners at successive gates. Analysis of the erroneous candidates as well as the correct targets provides a window for tracking the paths followed by listeners at successive points with increasing proportion of the acoustic signal. It was clear from this analysis that sentence context had a more important role than the lexical tonal information in disambiguating homophone meanings. Figure 2 presents a typical profile of the lexical candidates that listeners proposed for the high and the low density homophones that matched the sentence context. It shows that high density items elicited more uniform responses early on (i.e., responses that tended to fall within the semantic range of the sentence context), whereas low density items elicited more diverse responses. This difference was unlikely due to the inherent properties of the homophone, and might simply be due to the different degrees of contextual constraint of the sentence. We reason so because the successful identification was entirely contextual driven at this early point, and no inherent properties of the homophone per se, for example, tonal information could have an effect yet.

Figure 3 presents a typical profile of the lexical candidates that listeners proposed for the high and the low density homophones that did not match the sentence context. In contrast to the results in Figure 2, listeners proposed a larger variety of candidates for the target homophones across various stages of the identification process. More important, the high density items again elicited more uniform responses (but this time the responses tended to fall within the phonological structure of the target syllable), whereas the low density items elicited more diverse responses. This is easy to understand: high density homophones were associated with more lexical tones, and thus more items sharing the same syllable became activated, given the inappropriate contextual information; low density homophones, on the other hand, had fewer choices within its phonological structure, and thus the processing system was forced to search through phonologically irrelevant items.

Most interesting is that for both the low and the high density items, listener were initially misled to believe that the target was a word with the same syllable but a different tone, because the sentence context was biased to that meaning (e.g., taking *zi6* 'word' for *zi2* 'purple/paper', and *kwong4* 'crazy' for *kwong3* 'mine/expand'). As the acoustic signal unfolded, they had to make a switch to the syllable with the right tone. It is interesting to see at which point they could make such a switch, because this point would clearly reflect the interaction between sentence context and tonal information and the relative role of each variable. Figure 3 shows that the listeners did not start to make the switch until the seventh to eighth gate (240 to 280 msec of the word) for the high density items, and the fifth to sixth gate (160 to 200 msec of the word) for the low density items. This result revealed that (1) listeners were misled by an inappropriate context early on, before any tonal information could be detected in the speech; (2) tonal information did not have

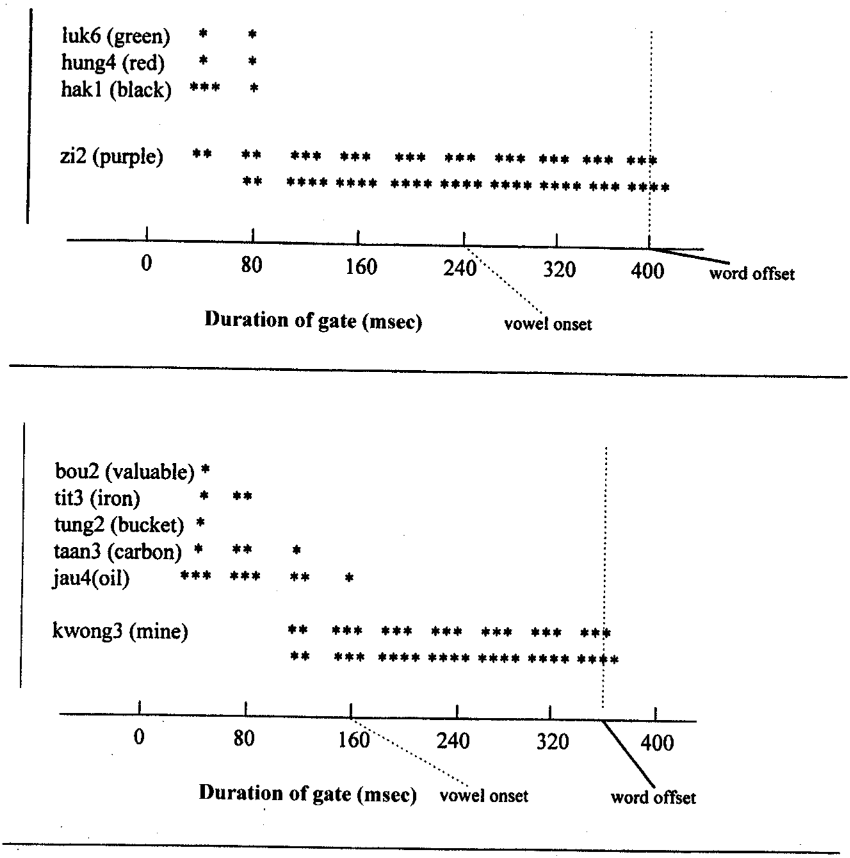

Figure 2. Profile of the lexical candidates that listeners proposed for the high density *zi2* (upper panel) and low density *kwong3* (lower panel) that matched the sentence context (Experiment 1). On the horizontal axis is the duration of gates (in 80-msec increments, i.e., two gates), and on the vertical axis are the proposed candidates. Phonetic transcriptions of the Chinese words are based on the LSHK Cantonese Romanization Scheme. The asterisks indicate the number of listeners who proposed the candidates.

any influence on processing until there was sufficient syllabic information. In other words, lexical tone, a suprasegmental phonological unit, was void until a sufficient amount of its segmental carrier, the syllable, was there; (3) in most cases, the point at which listeners switched to the right tone corresponded to the onset of the vowel, as indicated in Figure 3; and (4) the strong effect of sentence context could persist beyond listeners' detection of the tonal information, so that some listeners continued to propose the candidate with an incorrect tone that fit the sentence context.

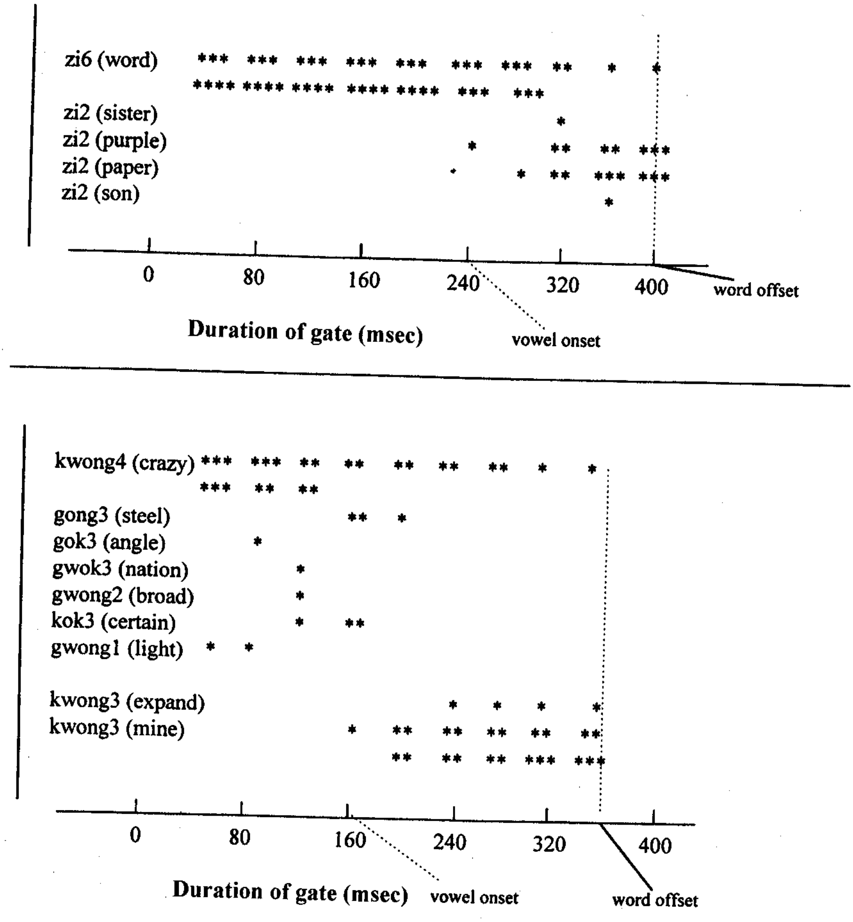

Figure 3. Profile of the lexical candidates that listeners proposed for the high density *zi2* (upper panel) and low density *kwong3* (lower panel) that did not match the sentence context (Experiment 1).

In this experiment, the high density items were not only associated with more lexical tones, but also associated with more semantic competitors within the same tone. By contrast, the low density items were associated with fewer lexical tones and at the same time with fewer semantic competitors within the same tone. The consequence of this discrepancy explains why the high density items required more acoustic information for identification than the low density items in the mismatch sentence contexts. Figure 3 shows that listeners proposed the same sound *zi2* with many different meanings (e.g., sister, purple, paper, and son), but they proposed only two items for *kwong3* (expand and mine), given the limited number of alternatives in the latter case. It shows that when the sentence context cannot guide the selection of phonologically correct alternatives, the more alternatives there are, the less likely listeners can hit on the right answer. This is exactly what happened in our experiment, as shown in Figure 3.

In sum, results from Experiment 1 indicate that context plays a significant role in the processing of Chinese homophones from early on. Consistent with our previous results (Li 1998; Li & Yip 1996), the experiment shows that when prior sentence context is semantically biased toward a specific meaning of the homophone, Chinese listeners can identify the appropriate meaning with less than half of the acoustic-phonetic information of the homophone. This experiment also demonstrates an interesting interaction between sentence context and lexical tonal information: the role of tonal information shows up relatively late during the temporal course of homophone processing, when sufficient amount of acoustic information of the syllable becomes available and usually at the onset of the vowel. Tonal information interacts with, and is often outweighed by sentence context, resulting in contextually driven interpretation of the homophone (e.g., misguiding listeners to a garden-path of word identification, as seen in Figure 3).

Experiment 2

In Experiment 1 we studied homophone processing by native Chinese speakers in a gating task, in which listeners were presented with homophones that varied in their match to prior sentence context. In this experiment, we seek further evidence to expand our vision of homophone processing by examining Chinese-English bilinguals' processing of cross-language homophones.

In Experiment 1 we used the gating task developed for spoken word recognition (Grosjean 1980). The gating task has also been successfully used for bilingual word recognition in code-switched situations (Grosjean 1988; Li 1996b). However, researchers have debated whether gating results reflect on-line processes or reflect only off-line processes (Cotton & Grosjean 1984; Grosjean, Dommergues, Cornu, Guillelmon & Besson 1994; Tyler & Wessels 1985). In this experiment, we used a cross-modal task, a variant of that used by Swinney (1979), and Seidenberg, Tanenhaus, Leiman, and Bienkowski (1982). In the cross-modal task, listeners hear an auditorially presented sentence followed by a visual probe at a given SOA, and are required to name the visual probe or make a lexical decision as soon as possible. Naming instead of lexical decision was used here, because (a) naming, in contrast to lexical decision, involves no listener's metalinguistic knowledge, and (b) lexical decision might be susceptible to post-lexical processing strategies, especially in studies of context effects (Forster 1981; Simpson & Krueger 1991). Cross-modal naming allows us to examine the access of different auditory candidates compatible with the speech signal, without explicitly manipulating the signal as in gating.

Participants. Twenty Chinese-English bilinguals who reported no speech or hearing deficits participated in this experiment. They were students at the Chinese University of Hong Kong. All of them used both Cantonese and English on a daily basis: Cantonese was their language with families and friends, and English had been their major language of education (at least up to the time of the experiment). They had all used English for over 10 years by the time of the experiment. They took part in the experiment as a laboratory requirement for credit in an introductory psychology course. None of them had taken part in Experiment 1.

Materials and design. Eight English nouns and eight English verbs were selected as the test words. These are words frequently used as code-switches in Cantonese speech (see Appendix 2). They all begin with a CV syllabic structure that shares with a Chinese CV syllable. To determine the possible effect of frequency on cross-language homophones, we compared the number of words that contain these initial CV structures in Chinese and English by examining the Longman Active Study Dictionary and the Cantonese Frequent Words Pronunciation Dictionary (Ho 1992). The Cantonese to English ratio is 1 to 0.62, indicating that these CV structures are slightly more popular in Cantonese than in English [the difference is not statistically significant, $t(30) = 1.56$, $p > 0.05$]. It shows that these CV syllables are homophonous across the two languages, especially if they are considered in the spoken language along a temporal resolution.

All the test words were embedded in a Cantonese sentence that had either a biasing context or a neutral context. The sentences with the test words were read by a bilingual speaker at a normal rate, and were digitized into the computer as in Experiment 1. The test words were pronounced in Cantonese phonetics, as is often done in natural code-switching (see Li 1996b, and footnote 1).

The visual probe was one of the following kind: (a) the same English word as the auditory test word, (b) a Chinese counterpart that shares the CV syllable as the test word, (c) an English word that shares no phonological overlap with the test word, and (d) a Chinese word that shares no phonological overlap with the test word. We selected these four kinds of visual probes for the following reasons. First, comparison between naming latencies to (a) and to

(b) would allow us to see if the auditory information from the test word similarly activates both English and Chinese candidates for the bilingual listener. Following Marslen-Wilson, Tyler, Waksler, and Older (1994), we assume that the auditory signal of the test word activates in the mental lexicon the lexical candidates compatible with the signal so far, which should in turn facilitate listeners' immediate responses to these candidates when they are presented in the visual form. Second, listeners' naming latencies to (c) and (d) would provide a baseline against which the effects of (a) and (b) could be compared.

Thus, the design of the experiment involved 2 levels of prior sentence context (biasing vs neutral), 2 levels of cross-language homophony (homophonous vs nonhomophonous), and 2 levels of the language of the visual probe (English vs Chinese). Sixteen Chinese words were also selected as fillers and intermixed with the test words during experimental presentation. The Chinese fillers were included to prevent listeners from identifying the nature of the test words and thus to prevent possible specialized processing strategies.

Procedure. During the experiment, listeners saw a fixation point on the computer screen for 500 msec. Immediately after this they heard on a pair of headphones a sentence with the test word, and then saw the visual probe at 150 msec SOA relative to the onset of the test word. The SOA was determined on the basis of the data from Li (1996b), at a point in time when the acoustic signal was still ambiguous between English and Chinese (and between various word candidates in each of the two languages). The visual probe appeared in the center of a high-resolution Apple monitor, roughly 8 × 8 cm in size, until the listener named it. As in Experiment 1, the PsyScope program (Cohen et al. 1993) controlled the presentation of the test materials. Listener's vocal responses would trigger the internal oscillator of the CMU button-box, a timing device connected to the computer interfaced with PsyScope.

The 20 participants were randomly assigned to four groups of five. Each group randomly received an equal number of sentences for each condition in the 2 × 2 × 2 design, and no listeners heard the same word twice under the same condition. The order of presentation of the sentences was pseudorandomly arranged so that the test words and fillers were interspersed.

Data analysis. The dependent variable was listeners' naming latencies to each of the visual probes. Naming latencies were automatically recorded by PsyScope, calculated from the onset of the auditory word to the onset of the listener's vocal response. Naming accuracies were also measured, and the overall error rate for this experiment was 2 percent.

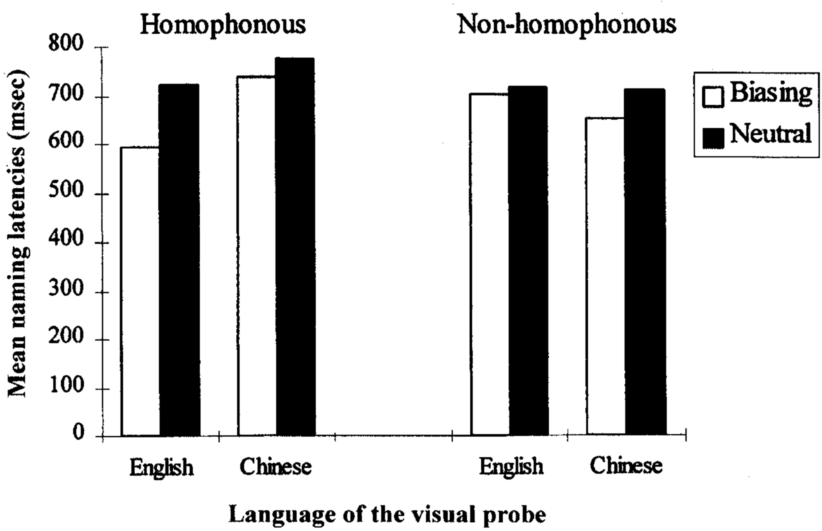

Figure 4. Mean naming latencies as a function of context, homophony, and language of the visual probe (Experiment 2).

Results and discussion

Figure 4 presents the mean naming latencies as a function of context, homophony, and language of the visual probe. A $2 \times 2 \times 2$ (context by homophony by language) ANOVA was conducted on the data in Figure 4. The analysis revealed a number of interesting results.

First, there was a significant main effect of context [$F(1,72) = 4.23$; $p < 0.05$], showing that prior sentence context significantly facilitated listeners' identification of the cross-language homophone, which in turn facilitated their naming of the target visual probe. Collapsed over variables other than context, listeners needed 733 msec to name the visual probe when the context was neutral, but they needed only 673 msec to name the same probe when context was biased.

Second, there was no main effect of either homophony [$F(1,72) = 0.13$; $p > 0.05$] or language of the visual probe [$F(1.72) = 1.33$; $p > 0.05$], showing that in general the naming of a visual probe did not matter whether or not the visual probe had phonological overlaps with the auditory homophone, or whether it was in English or in Chinese.

Third, there was a significant interaction between homophony and language of the visual probe [$F(1,72) = 4.50$; $p < 0.05$]. This interaction shows that when the visual probe overlapped with the auditory homophone phonologically, the English visual probes elicited significantly faster responses than the Chinese probes (on the average 659 vs 757 msec, Tukey HSD, $p < 0.05$).

This pattern was apparently due to listeners' early identification of the auditory homophone, which facilitated responses to the English visual probes, compared to the Chinese visual probes. In contrast, when the visual probe did not overlap with the auditory homophone phonologically, the English and the Chinese visual probes elicited more similar response latencies (on the average 713 vs 684 msec, Tukey HSD, n.s.). This pattern was predicted: if there was no phonological overlap, the visual probe should receive no facilitation from the auditory homophone, irrespective of the language in which it occurred. This result also shows that the non-homophonous visual probes did serve as good baseline items for the homophonous visual probes.

With this interaction in mind, it is easy to understand that in Figure 4, the critical comparison for context effects involved the differences in naming latency between the English and Chinese visual probes in the two context situations, under the homophonous condition only. The naming latency difference between English and Chinese in the biased context was much greater (596 vs 742 msec) than that in the neutral context (719 vs 774 msec). This indicates that prior sentence context helped the bilingual listeners to recognize the English target word more quickly and in turn they named it faster than its Chinese counterpart.[2]

In sum, results from Experiment 2 provide us with new information about the role of context effect in bilingual word recognition. They indicate that bilingual Chinese-English listeners can use prior sentence context early on to recognize a code-switched word that is pronounced in native Chinese phonetics, which subsequently facilitates their naming responses to English visual probes. As we set the SOA at 150 msec, the results suggest that listeners can use prior contextual information around 150 msec of the spoken word. Sentence context apparently had an effect at this early point, to distinguish the contextually appropriate word in English from the contextually inappropriate counterpart in Chinese.

General discussion

This study examined the role of context effects in the processing of spoken homophones in two experiments. In Experiment 1, Chinese speakers were presented with successively gated portions of a homophone in a sentence context, and they identified the homophone on the basis of its increasing acoustic information. The results indicate that sentence contexts influence the processing of Chinese homophones from early on, shortly after the acoustic onset of the word: when the homophone matches with sentence context, Chinese speakers can identify the appropriate meaning with less than half of the acoustic-phonetic information of the homophone. The results also indicate that lexical tonal information plays its role relatively late, usually at the onset of the vowel of a syllable, and that tonal information interacts with sentence context, leading to purely contextually driven interpretations of the lexical item. In Experiment 2, Chinese-English bilinguals were presented with a cross-language homophone in a sentence context, and they named a visual probe that had or did not have phonological overlaps with the homophone. The results show that prior sentence context significantly influences Chinese-English bilinguals' recognition of cross-language homophones, within the acoustic boundary of the word. Context helps bilingual listeners select the appropriate words at an early point when the acoustic signal is still ambiguous between Chinese and English and between various lexical candidates in the two languages.

Results from this study add new information on the operation of context effects in both monolingual and bilingual situations, and on the interaction between context and tonal information in homophone processing in Chinese. Consistent with our previous studies, our data support the context-dependency hypothesis that ambiguous meanings of a word may be selectively accessed from early on according to prior sentence context (Simpson 1981; Simpson & Krueger 1991; Tabossi 1988). In contrast, our study indicates that it is unlikely that Chinese speakers would exhaustively access all meanings of a homophone without using contextual information initially to constrain the access. Chinese speakers, faced with the extensive ambiguity created by massive homophones in the language, seem to have at their disposal a processing system that can rapidly disambiguate alternative homophone meanings during sentence comprehension. Such a processing system must be contextually driven early on to be able to operate efficiently.

There has been evidence in spoken word recognition that English speakers can identify a one-to-three syllable word in sentence context within about 200 msec, usually half or less of the acoustic signal of the word (Grosjean 1980; Marslen-Wilson 1987). According to Marslen-Wilson (1987), in English, there would be an average of 40 words still compatible with the available stimulus at 200 msec, when only the initial two phonemes are heard. In a bilingual situation, the problem may be even worse if lexical items are considered outside of context, because the number of lexical candidates compatible with 200 msec of a cross-language homophone may be even larger. Results from our study indicate that listeners can identify the correct meaning with only 33 percent of the homophone in the right context (Experiment 1), and they can successfully respond to the target visual probe when only 150 msec of the auditory homophone was heard (Experiment 2). The 33 percent or 150 msec is insufficient acoustic information of a word. An examination of the acoustic waveforms of the 16 test words in Experiment 2 reveals that at

150 msec, all words except two included only the initial consonant plus some information of the vowel, at which point the word is still ambiguous between Chinese and English (and between various lexical candidates in each of the two languages). It is thus hard to imagine how speakers could recognize a word with only its minimal acoustic information, if they do not use contextual information from early on.

Tonal information in Chinese differentiates alternative meanings associated with the same syllable and thus reduces the potential number of homophones, although it does not eliminate homophony. We show in this study, however, that the role of tone in homophone processing is limited relative to the role of sentence context. Lexical tone can help the listener to disambiguate homophone meanings only when sufficient amount of the acoustic signal of the homophone is available, usually at the onset of the vowel in a syllable. Initially, only sentence context guides (or misguides) the word identification process. Later on, tonal information helps listeners to select among various candidates. In addition, tonal information does not always help. In some cases, listeners have detected the physical properties of the tone associated with the syllable, but context effects persist through the entire spectrum of the homophone, leading listeners to adhere to their incorrect identification. This pattern shows that context may initially override the physical properties of the lexical items during perception, leading to a garden-path of interpretation.

In short, results from the present study suggest that the successful recognition of spoken homophones depends on the interactions among the contextual, lexical, and phonological information in the sentence from early on. These results are best accounted for by interactive activation models of the sort in Kawamoto (1993), Marslen-Wilson (1987), McClelland (1987), and McClelland and Elman (1986). In these models, information processing flows both bottom-up and top-down, rather than strictly bottom-up, and lexical access and sentence context mutually influence one another at an early stage, rather than a stage at which context effects follow the completion of lexical access. These interactive models are largely inspired by or built on connectionist mechanisms that involve distributed representation, degrees of activation, and adaptation of connection strengths among processing units for phonological, lexical, syntactic, and semantic information of the sentence (Rumelhart, McClelland & the PDP Research Group, 1986).

In a connectionist perspective, the processing of spoken homophones can be viewed as an interactive process of constraint satisfaction: multiple sources of phonological, lexical, and contextual constraints either converge to facilitate the activation of relevant meanings, or compete to inhibit their activation. Thus, the product of processing at any stage is a result of the interactions among these sources of constraints, each of which may contribute different weights at a given time. Our goal in this line of research is to provide a comprehensive picture of the interactions among these various constraints, including context effects, homophone density effects, effects of lexical tones, and effects of the frequency of homophone meanings in the temporal course of the processing of spoken homophones.

Acknowledgments

The preparation of this paper was supported by a Faculty Research Initiation Grant from the School of Arts and Sciences, University of Richmond to the first author, and in part by a postgraduate scholarship from the Chinese University of Hong Kong to the second author. We would like to thank Elizabeth Bates, Brian MacWhinney, and Francois Grosjean for insightful discussions on the experiments presented here. Thanks also go to Sylvia Pang and Cynthia Wong for their assistance to the experiments.

Notes

1. One may argue that these are not true homophones, because there are subtle phonological differences between them (e.g., the vowel in *pique* versus that in *pick* differ in quality). However, it is important to consider what characterizes cross-language homophone for the bilingual speakers and listeners in bilingual conversations. Cross-language homophones occur mainly in code-switching situations. In these situations, (1) the bilingual speaker does not always follow phonological prescriptions to pronounce a code-switched word exactly as the word should be pronounced by a monolingual native speaker in the target language, and thus the phonologically defined subtle differences may not be there in the bilingual speaker's output. In fact, bilingual speakers often pronounce a code-switched word in the phonetics of their base/native language (Grosjean 1988; Li 1996b); (2) bilingual listeners may also not be sensitive to the subtle differences in comprehension, because misinterpretations or misunderstandings of cross-language homophones can also occur on the part of the listener. Grosjean (1988) and Li (1996) show that upon hearing cross-language homophones, bilingual listeners may propose homophonous counterparts in the base language (French or Chinese); (3) in real-time processing, cross-language homophones are ambiguous between the two languages at various temporal locations of the acoustic spectrum (especially in the perspective of the cohort model of Marslen-Wilson 1987). For example, the CV structures contain many pairs that share initials and vowels in Chinese and English, such as *lok-lock, sik-sit,* and *fei-fail.* Examination of these pairs can provide important information about context effects and lexical access in bilinguals.
2. The responses were slightly faster for the English than the Chinese visual probes even in the neutral context situation. This difference might be because the target word was not fully adapted to the Chinese phonetics when the reader pronounced it, which provided cues to the listener for identifying it as English. See Li (1996b) for the role of language phonetics in bilingual word recognition.

Appendix 1. Chinese homophones used in Experiment 1

Low density items	aap3	ai2
	ai3	caak3
	caak6	caang1
	caang4	faa1
	faa3	gap1
	gap3	goek3
	gwat1	gwat6
	hoe1	jeng2
	jeng4	kwong3
	kwong4	mat6
	nai4	niu5
	niu6	ok3
	syut3	teng1
	teng5	tit3
	tuk1	uk1
High density items	ci3	ci4
	faan6	fan1
	fu2	fu3
	jan4	jau4
	jau6	ji1
	ji5	jing4
	joeng4	jyu4
	jyu6	jyun4
	jyun6	sau2
	seoi2	seoi3
	si1	si4
	sin3	tou3
	wai1	wai6
	wan4	wu4
	zi2	zi6

Note: the number following each syllable indicates the correct tone that matches the sentence context.

Appendix 2. Cross-language homophones (English targets) used in Experiment 2

Nouns		Verbs	
bike	size	cut	sit
book	sense	fail	solve
boss	gift	lock	take
case	list	sell	talk

References

Cohen, J., MacWhinney, B., Flatt, M. & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments, *Behavior Research Methods, Instruments, and Computers* 25: 257–271.

Cotton, S. & Grosjean, F. (1984). The gating paradigm: A comparison of successive and individual presentation formats, *Perception and Psychophysics* 35: 41–48.

Fodor, J.A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: The MIT Press.

Forster, K. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing, *Quarterly Journal of Experimental Psychology* 33: 465–495.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm, *Perception and Psychophysics* 28: 267–283.

Grosjean, F. (1988). Exploring the recognition of guest words in bilingual speech, *Language and Cognitive Processes* 3: 233–274.

Grosjean, F., Dommergues, J., Cornu, E., Guillelmon, D. & Besson, C. (1994). The gender-marking effect in spoken word recognition, *Perception and Psychophysics* 56: 590–598.

Ho, K. (1992). *Cantonese frequent words pronunciation dictionary*. Chinese Division, Institute of Education in Language, Hong Kong Education Department.

Institute of Linguistics, The Academy of Social Sciences (1985). *Xianda Hanyu Cidian [Modern Chinese Dictionary]*. Beijing: Commercial Press.

Kawamoto, A.H. (1993). Non-linear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account, *Journal of Memory and Language* 32: 474–516.

Li, P. (1994). Understanding the time course of sentence comprehension: A sentence gating study in Mandarin Chinese. In Chang, H-W., Huang, J-T., Hue, W. & Tzeng, O. (eds.), *Advances in the study of Chinese language processing*, Vol. 1 (pp. 303–323). Taipei: National Taiwan University Press.

Li, P. (1996a). The temporal structure of spoken sentence comprehension in Chinese, *Perception and Psychophysics* 58: 571–586.

Li, P. (1996b). Spoken word recognition of code-switched words by Chinese-English bilinguals, *Journal of Memory and Language* 35: 757–774.

Li, P. (1998). Crosslinguistic variation and sentence processing: The case of Chinese. In D. Hillert (ed.), *Sentence processing: A cross-linguistic perspective*. San Diego, CA: Academic Press (in press).

Li, P., Bates, E. & MacWhinney, B. (1993). Processing a language without inflections: A reaction time study of sentence interpretation in Chinese, *Journal of Memory and Language* 32: 169–192.

Li, P. & Yip, M. (1996). Lexical ambiguity and context effects in spoken word recognition: Evidence from Chinese. In G. Cottrell (ed.), *Proceedings of the 18th Annual Conference of the Cognitive Science Society* (pp. 228–232). Hillsdale, NJ: Erlbaum.

The Linguistic Society of Hong Kong (1994). *The LSHK Cantonese Romanization Scheme*, Jyutping. LSHK Newsletter, No. 17.

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

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

Marslen-Wilson, W. D. & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech, *Cognitive Psychology* 10: 29–63.

McClelland, J.L. (1987). The case for interactionism in language processing. In M. Coltheart (ed.), *Attention and performance* Vol. 12: *The psychology of reading* (pp. 3–36). Hillsdale, NJ: Erlbaum.

McClelland, J.L. & Elman, J. (1986). Interactive processes in speech perception: The TRACE model. In: McClelland, J.L., Rumelhart, D.E. & the PDP Research Group (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, Vol. 2 (pp. 58–121). Cambridge, MA: The MIT Press.

Onifer, W. & Swinney, D.A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias, *Memory & Cognition* 9: 225–236.

Rumelhart, D., McClelland, J. & the PDP Research Group (1986). *Parallel distributed processing; Explorations in the microstructure of cognition*, Vol. 1, Cambridge, MA: The MIT Press.

Seidenberg, M.S., Tanenhaus, M.J., Leiman, J.M. & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing, *Cognitive Psychology* 14: 538–559.

Simpson, G. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity, *Journal of Verbal Learning and Verbal Behavior* 20: 120–136.

Simpson, G.B. & Krueger, M.A. (1991). Selective access of homograph meanings in sentence context, *Journal of Memory and Language* 30: 627–643.

Small, S., Cottrell, G. & Tanenhaus, M. (1988). *Lexical ambiguity resolution: perspectives from psycholinguistics, neuropsychology, and artificial intelligence*. San Mateao, CA: Morgan Kaufmann Publishers.

Swinney, D.A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects, *Journal of Verbal Learning and Verbal Behavior* 18: 645–659.

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential contexts, *Journal of Memory and Language* 27: 324–340.

Tyler, L. & Wessels, J. (1985). Is gating an on-line task? Evidence from naming latency data, *Perception and Psychophysics* 38: 217–222.

Addresses for correspondence: Ping Li, Department of Psychology, University of Richmond, Richmond, VA 23173, USA
Phone: (804) 289–8125; Fax: (804) 289–8943; E-mail: ping@cogsci.richmond.edu