



# Spoken word recognition of Chinese homophones: A further investigation

Michael C.W. Yip

School of Arts & Social Sciences  
The Open University of Hong Kong, Hong Kong SAR

## ABSTRACT

A cross-modal naming experiment was conducted to examine the effects of context and other lexical information in the processing of Chinese homophones during spoken language comprehension. In this experiment, listeners named aloud a visual probe as fast as they could, at a pre-designated point upon hearing the sentence, which ended with a spoken Chinese homophone. Results further support that prior context has an early effect on the disambiguation of various homophonic meanings, shortly after the acoustic onset of the word. Second, context interacts with frequency of the individual meanings of a homophone during lexical access. Finally, the present results pattern is clearly consistent with the context-dependency hypothesis that selection of the appropriate meaning of an ambiguous word depends on the simultaneous interaction of both sentential and lexical information during lexical access.

## 1. INTRODUCTION

The relationship between context effects and spoken word recognition has been the central focus for the study of lexical ambiguity research [1,2,3,4]. The basic theoretical question in this line of research is the time-course of the context effects during different various stages of lexical access? Answers for this question can not only reflect the underlying mechanism to our language processors but also the operation to the entire human cognitive system.

There were two contrastive hypotheses from the relevant literature: (1) exhaustive access hypothesis, which argues that all meanings of an ambiguous word will be accessed momentarily following the occurrence of the word, and the context can only help to select the appropriate meaning at a late post-access stage. This hypothesis builds on the assumption that language processing is a modular, bottom-up process in which non-lexical, contextual information does not penetrate lexical access [5]. (2) context-dependency hypothesis, which argues that the contextually appropriate meaning of an ambiguous word can be selectively accessed early on if the preceding sentence context provides a strong bias to the appropriate meaning. This hypothesis builds on the assumption that language processing is operated by an interactive approach in which information can flow both bottom-up and top-down simultaneously and that lexical access and sentential context can mutually influence one another at a very early stage [6].

Empirical examinations for the above hypotheses have been mostly conducted in English and other Indo-European

languages (e.g., Dutch and Italian). So far, only a few experimental studies were systematically examined in Chinese. Chinese is a Sino-Tibetan language that differs significantly from most Indo-European languages (e.g., in its use of lexical tones, its morphemic monosyllabicity) and it also offers many unique and interesting psycholinguistic properties in its phonological, lexical, and syntactic structures [7] to crucially investigate the issue of lexical ambiguity. According to the Modern Chinese Dictionary [8], 80% of the monosyllables (with tonal differentiation) in Chinese are ambiguous between different meanings, and 55% have five or more meanings. Taking Cantonese syllable as an example, the monosyllable *sil* has up to 14 meanings (e.g., teacher, lion, silk, corpse, poem, private, think, etc.), and this number would increase to 37 if identical syllable with different tones were considered as homophones [9]. Therefore, upon hearing the monosyllable *sil* in a sentence, do native Cantonese listeners activate all 14 or even more meanings of the single syllable simultaneously? They should if we follow the exhaustive access hypothesis in its strict sense, because according to this hypothesis: lexical access is an autonomous and capacity-free process. However, if we follow the context-dependency hypothesis, only the contextually appropriate meaning will be activated when listeners hear the syllable due to the robust context effects.

Recent studies on spoken word recognition of Chinese homophones [10,11,12,13,14] clearly provided supportive evidence to the context-dependency hypothesis. However, some critiques on the materials (properties of the visual probes) used in those experiments were proposed to challenge the data [15]. In the present study, we replicate the work, by using a cross-modal naming experiment, and take care of all the possible confounding in order to seek further evidence for making an accurate and complete picture on the relationship between the sentential context effects and other lexical factors operated in the language processing.

## 2. EXPERIMENT

### 2.1. Method

**Participants.** One hundred and forty-four native Cantonese speakers (52 male and 92 female, mean age =22.6) who reported no speech or hearing deficits participated in this experiment. All participants were students and staffs at the Open University of Hong Kong. They took part in the experiment on a voluntary basis.

**Materials and design.** Thirty spoken Cantonese homophones were selected, each with at least two different

meanings in the same tone (syllables with different tones are not considered homophones in the present study). Each homophone was embedded in three different sentences with prior context either biased to the dominant or the subordinate meaning or both. A separate group of 20 native Cantonese speakers was asked to judge the degree of constraint of the prior context on the target homophone. They were given the 60 test sentences with the prior biasing context (excluding the 30 ambiguous test sentences) but without the homophone, and were asked to fill in the word. They were told to think of a Chinese word that would naturally complete the sentence. Their responses were scored on a 1-4 scale, based on the scale proposed by Marslen-Wilson and Welsh [16]: 1 was given for a word identical to the test word, 2 for a synonym, 3 for a related word, and 4 for an unrelated word. Responses were pooled across the 20 judges, and the mean rating was 1.6. This score was above the high constraint condition in Marslen-Wilson and Welsh [16]. An effort was also made to have prior sentence context of equal length, and the average length of the test sentences, counting the target homophone, was 14 words (ranging from 12 to 17 words). More importantly, we have tried to eliminate the intra-sentential priming from any individual words within the current sentence when constructing the sentence context as much as possible.

Other than the sentential context, we carefully selected the appropriate visual probes as follows. First, all the visual probes were based on a semantic relatedness norm experiment from another separate group of one hundred native Cantonese speakers. In this simple experiment, the participants were asked to immediately think of three Chinese words that have the same or closely related meaning to each homophone, and the mostly frequent words they listed were used to be the related visual probe for each spoken homophone. The unrelated probes were randomly selected from the same source. All the visual probes in each experimental condition were matched with the same category of initial phonemes and frequency information.

Altogether, there were five variables manipulated in this experiment:

- (1) Context type: The preceding context was (a) biased to the dominant meaning of a homophone, or (b) biased to the subordinate meaning of the homophone, or (c) ambiguous.
- (2) Dominance: The visual probe was related to either the dominant meaning of the spoken homophone, or to the subordinate meaning of the spoken homophone. The dominance information was empirically based [17].
- (3) Homophone Density: A given homophone had either many potential semantic competitors (four or more alternative meanings) or few semantic competitors (two to three alternative meanings). The density information was also empirically based [9].
- (4) Relatedness: The visual probes were either semantically related to the spoken homophone or unrelated.
- (5) SOA: The visual probe occurred at three given SOA (stimulus-onset-asynchrony) relative to the spoken homophone: (a) the isolation point (IP), or (b) the acoustic offset of the homophone (OS), or (c) 300ms after the acoustic offset of the homophone (300ms). The isolation

point (IP) for each homophone was derived from other previous gating results [13]. The IP for each spoken homophone was different (average IP was 54%) and the IP indicated the lowest acoustic information percentage of the spoken word that needed for correct recognition.

An example *cheung1* (window/gun) in the three context type manipulations test sentences is given below.

- (a) Sentence context biased to the dominant meaning  
Gaan1 uk1 gam3 guk6 lei5 faai3 di1 jau2 hui3 hoi1 saai3 di1 *cheung1*.  
The room is so muggy that you should rush to open all the windows.
- (b) Sentence context biased to the subordinate meaning  
Gwan1 foh2 juen1 ga1 wa6 lei1 di1 chuen4 bo6 do1 hai6 jan1 *cheung1*.  
Military experts said that all of these are real guns.
- (c) Ambiguous sentence context  
Ngho5 yiu3 lei5 dei6 yi4 ga1 jik1 hak1 jau2 hui3 hoi1 *cheung1*.  
I order you all immediately to fire/open your guns/ the window.

The four visual probes to *cheung1* in these sentences are: *moon4* “door” (related-dominant), *daan6* “bullet” (related-subordinate), *min6* “noodle” (unrelated-dominant), and *dau6* “bean” (unrelated-subordinate).

**Experimental Apparatus.** All the test sentences were read by a female native Cantonese speaker at a normal rate and tape-recorded in a SONY DAT deck. Then, the spoken sentences were transformed and digitized into a Macintosh G3 computer. A computer program called PsyScope [18] controlled the presentation of the materials. A microphone, which was used to register listeners' vocal responses and hence calculated the naming latencies, was connected to the computer through the CMU button-box. A remote controlled SONY tape-recorder was also used and controlled by the experimenter in another partition of the experimental room to check for accuracy.

**Procedure.** Participants were divided into three groups of 48 according to the three context types. Context type was treated in a between-subject design, and the other variables in a within-subject design. Within each context condition, the 48 participants were randomly assigned to twelve groups of four. Each group randomly received an equal number of sentences for each context condition in the 2 (Dominance) x 2 (Homophone Density) x 2 (Relatedness) x 3 (SOA) mixed factorial design. This yielded a total of twenty-four different experimental conditions. The order of presentation for the sentences was pseudo-randomly arranged such that the visual probes did not consecutively bias spoken homophones. The order of presentation was counterbalanced across all participants. No participant heard the same homophone twice.

All participants did the experiment individually in a quiet experimental room. Before the experiment, the experimenter explained the task in Cantonese to the listener. First, they were told that they would be hearing a sentence through a pair of B&W speakers, and then towards the end of the sentence, they would see a Chinese character (visual probe)

appeared on the computer screen. Their task was simply to name the Chinese character aloud into the microphone as quick and accurately as possible. Before the actual experiment began, they were given a practice session in which they heard a set of separate but similar sentences. The whole experiment took about twenty minutes.

## 2.2 Results

Mean response latencies, counting from the onset of the visual probe to the vocal response, as a function of context, SOA, dominance and relatedness are presented in the following table. Errors (i.e. listeners named the visual probes with a word that is totally different from the target word) were very rare (approximately 0.07 across all conditions), and therefore data were not analyzed further in the present study.

Context type	Subordinate		Dominant	
	Related	Unrelated	Related	Unrelated
<b>Dominant</b>				
IP	803.8	813.3	761.3	800.2
OS	780.6	780.4	750.2	773.5
300msec	752.6	774.1	740.6	760.2
<b>Subordinate</b>				
IP	858.2	858.5	816.7	839.8
OS	834.7	861.7	802.6	814.2
300msec	810.6	821.3	750.1	793.7
<b>Ambiguous</b>				
IP	836.5	852.4	851.8	828.9
OS	819.2	846.4	800.4	828.1
300msec	830.1	796.7	784.6	790.3

\*Note: Because the effect of homophone density was absent in the experiment, data for this variable were not computed in the above table.

First, in the condition of ambiguous sentence context, a 3 (SOA) x 2 (dominance) x 2 (relatedness) x 2 (homophone density) repeated measure ANOVA was conducted on the response latencies to the visual probes. Results found that frequency of the individual meanings of the homophone greatly influenced the response time,  $F(1,47) = 38.56$ ,  $p < .001$ . Collapsed over levels of other variables, the mean response time to access the dominant homophone meaning was 814msec and that to the subordinate homophone meaning was 830msec. Results indicated that the dominant meaning of a given homophone would be activated first if no biased contextual information was provided beforehand. However, the post hoc comparisons (Tukey HSD test) reported that there was no difference between the dominant and subordinate homophone meanings at the IP condition,  $p > .05$  and the differences were mainly from the OS and 300msec conditions,  $ps' < .05$ . This result might propose that there was a very short period of time (probably from the IP to OS) that both meanings of the homophone would be accessed after the occurrence of the ambiguous word. This speculation was confirmed by the interaction of SOA by dominance,  $F(2,188) = 17.39$ ,  $p < .05$ . In addition, there was a main effect on the variable of SOA,  $F(2,94) = 60.33$ ,  $p < .000$ . Collapsed across other variables, the mean response time to each SOA condition was 842msec (IP), 824msec (OS) and 800msec (300msec); and the fastest response latencies usually occurred at the 300msec SOA. The SOA effect in the ambiguous context implied that the access to a correct lexical meaning appeared at a relatively late period

of time (300msec following the occurrence of the ambiguity) when the preceding sentence did not provide a biasing contextual influence. This result is consistent with Simpson and Krueger's finding [19]. And, no other effects of homophone density and relatedness were significant.

Second, in the condition of context biasing to the dominant meaning, ANOVA analyses suggested that context effects for the biased sentences to the dominant homophone meaning were stronger than the subordinate meaning,  $F(1,94) = 15.4$ ,  $p < .05$ . Consistent with the data from the ambiguous sentence condition, there was also a main effect on SOA. Collapsed across other variables, the mean response time to each SOA condition was 795msec (IP), 771msec (OS) and 757msec (300msec); and the fastest response latencies again occurred at the 300msec SOA. But the post hoc comparison (Tukey HSD test) indicated that there was no difference between the OS vs. 300msec condition,  $p > .05$  and the differences were mainly came from the comparisons between the IP vs. OS and IP vs. 300msec SOA conditions,  $ps' < .05$ . These results indicated that the processing time to access the correct homophone meaning would be shortly after the IP of the homophone and possibly within the acoustic boundary of the word [10,11,12,13,14] if there was a strong contextual biasing effect. In addition, the variable of relatedness was statistically significant at,  $F(1,47) = 27.64$ ,  $p < .05$ . This result was simply a kind of semantic priming effect that listeners usually responded to related words faster than the unrelated control words [4,20]. Again, the variable of homophone density was not significant.

Third, in the condition of context biasing to the subordinate meaning, the main effects of SOA, dominance, and relatedness were all significant, which are consistent with the earlier ANOVA analyses. Moreover, two interactions were observed on types of biasing sentences by SOA,  $F(2,188) = 10.2$ ,  $p < .05$ ; and types of biasing sentences by SOA by dominance,  $F(2,188) = 28.94$ ,  $p < .000$ . These results suggested that when the sentence context was biased to the dominant homophone meaning, the processing time would be significantly shorter than the subordinate homophone meaning in the conditions of SOA (OS) and SOA (300msec).

Finally, a separate analysis to compare the processing times in biasing and ambiguous sentence contexts was conducted, collapsed across the two biasing sentences. Results revealed a clear context effect,  $F(1,94) = 9.14$ ,  $p < .05$ . The average response latencies in biased sentence contexts were 798msec and that in the ambiguous sentence context was 822msec. These results indicated that the disambiguation of homophone meaning occurred much earlier in the biasing context than it occurred in the ambiguous context.

## 3. GENERAL DISCUSSION

The present study further examines the lexical ambiguity in sentence processing from a cross-language perspective, we used spoken Chinese homophone as a rigorous test case because of the pervasive homophony phenomenon in Chinese language. Clearly, the present results show that sentence context aids the processing of Chinese homophones at an early stage, shortly after the acoustic onset of the word and most likely within the acoustic

boundary of the spoken word. These findings further confirm that Chinese speakers could identify the appropriate meaning at an early point of time that was consistent with other relevant studies in Chinese sentence processing [10,11,12,13,14].

In addition, the present results also indicated a strong dominance (frequency) effect and its interaction with sentence context. The dominance effect took place early on, probably within the temporal acoustic boundary of the spoken word, as revealed in the fast response time in the word offset condition. This result is obviously against the assumption of modular approach: frequency effect could only occur at a late selection stage [1,3]. Furthermore, the interaction effects between context and frequency clearly implied that contextual information mutually interacted with the lexical information of the Chinese homophones together during sentence processing. These results are compatible with the reordered access hypothesis [21], which stated that importance of context and frequency effects in lexical ambiguity processing.

One point should be noted that the effect of homophone density was also absent in the present study [10,11]. Two possible explanations for this result can be made. First, it may be due to the fact that the semantic judgment norms gathered in this study are ineffective in reflecting the true nature of this variable. And the second one is that the variable of homophone density may be far from simply measuring the actual amount of semantic competitors based on the same phonological representation because listeners may not have knowledge to some potential words in their mental lexicon [22]. Ongoing experiments are being designed in our laboratory to further examine the nature of homophone density effect in Chinese sentence processing.

#### 4. ACKNOWLEDGEMENTS

This research was supported by a Research and Development Fund from the Open University of Hong Kong. I would like to specially thank Ping Li for his constructive comments to the present study.

#### 5. REFERENCES

1. 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.
2. Simpson, G. B. (1981) Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning and Verbal Behavior*, 20, 120-136.
3. Swinney, D. A. (1979) Lexical Access during Sentence Comprehension: (Re)Consideration of Context Effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645-659.
4. Tabossi, P. (1988) Accessing Lexical Ambiguity in Different Types of Sentential Contexts. *Journal of Memory and Language*, 27, 324-340.
5. Fodor, J. A. (1983) *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.
6. McClelland, J. L. (1987) The case for interactionism in language processing. In M. Coltheart(Ed.), *Attention and performance XII: The psychology of reading*. Hillsdale, NJ: Erlbaum.
7. Zhang, Y., Wu, N., & Yip, M. C. W. (2006). Lexical ambiguity resolution in Chinese sentence processing. In P. Li, L. H. Tan, E. Bates, & O. J. L. Tzeng. (eds.), *Handbook of East Asian Psycholinguistics (Vol. 1: Chinese)*, pp. 268-278. Cambridge, UK: Cambridge University Press.
8. Institute of Linguistics, The Academy of Social Sciences. (1985) *Xianda Hanyu Cidian* (Modern Chinese Dictionary). Beijing: Commercial Press.
9. Zhang, L.-Y. & Zhang, S.-Y. (1987). *Guoyin Yueyin Suoyin Zihui* (A Pronunciation Dictionary of Cantonese and Putonghua). Hong Kong: Chung Hwa Bookstore.
10. Li, P., & Yip, M. C. W. (1996) Lexical ambiguity and context effects in spoken word recognition: Evidence from Chinese. In G. Cottrell (ed.), *Proceedings of the 18th Annual Meeting of the Cognitive Science Society*, 228-232.
11. Li, P., & Yip, M. C. W. (1998) Context Effects and the Processing of Spoken Homophones. *Reading and Writing*, 10, 223-243.
12. Li, P., Shu, H., Yip, M. C. W., Zhang, Y., & Tang, Y. (2002). Lexical ambiguity in sentence processing: Evidence from Chinese. In M. Nakayama (ed.), *Sentence Processing in East Asian Languages*. Stanford, CA: CSLI Publications. pp. 111-129.
13. Yip, M. C. W. (2000). Spoken word recognition of Chinese homophones: The role of context and tone neighbors. *Psychologia*, 43, 135-143.
14. Yip, M. C. W. (2002). Access to homophonic meanings during spoken language comprehension: Effects of context and neighborhood density. In the *Proceedings of ICSLP'2002*, 1665-1668.
15. Yip, M. C. W. (2007). Lexical and sentential priming of ambiguous words. Manuscript under review.
16. Marslen-Wilson, W. D., & Welsh, A. (1978) Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10, 29-63.
17. Ho, H.-H. & Jiang, Y.-H. (1994) *Word frequency in Hong Kong in the 90's*. Research Institute of Humanities, Chinese University of Hong Kong.
18. Cohen, J. D., MacWhinney, B., Flatt, M., and Provost, J. (1993) PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments, and Computers*, 25, 257-271.
19. Simpson, G. B., & Krueger, M. A. (1991) Selective Access of Homograph Meanings in Sentence Context. *Journal of Memory and Language*, 30, 627-643.
20. Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds). *Basic processes in reading: Visual word recognition*. Hillsdale, NJ. Lawrence Erlbaum Associates. pp. 264-336.
21. Hogaboam, T., & Perfetti, C. (1975). Lexical ambiguity and sentence comprehension. *Journal of Verbal Learning & Verbal Behavior*, 14, 265-274.
22. Ahrens, K. (2002). Timing issues in lexical ambiguity resolution. In M. Nakayama (ed.) *Sentence Processing in East Asian Languages*. Stanford, CA: CSLI Publications. pp. 1-29.