ACCESS TO HOMOPHONIC MEANINGS DURING SPOKEN LANGUAGE COMPREHENSION: EFFECTS OF CONTEXT AND NEIGHBORHOOD DENSITY

Michael C.W. Yip
School of Arts & Social Sciences
The Open University of Hong Kong, Hong Kong SAR

ABSTRACT

A gating experiment was conducted to examine the effects of context and neighborhood density information in the processing of Chinese homophones during spoken language comprehension. In this experiment, listeners were presented with successively gated portions of a spoken homophone, embedded in a sentence context, and they identified the homophone on the basis of its increasing amount of acoustic information. Results indicate that 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. Third, the neighborhood density information helps to narrow down the number of candidates of the target homophone along the temporal course of spoken language processing. Finally, the results are interpreted in terms of interactive activation models of lexical processing.

1. INTRODUCTION

Context effects and spoken word recognition have been the main focus for the study of lexical ambiguity research during the last two decades [1,2,3,4]. The crucial empirical question in this line of research is how early the context effects take place during different stages of lexical access? Answers for this question can reflect the underlying mechanism to our language processors in particular and the entire cognitive system in general. However, after more than two decades of psycholinguistic research, there is still no definitive answer today.

Two contrastive hypotheses have emerged out of the relevant literature: 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 language processing is a modular, bottom-up process in which non-lexical, contextual information does not penetrate lexical access [5]. The context-dependency hypothesis 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 assumes 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].

In fact, the above hypotheses have been mainly tested in English and several other Indo-European languages (e.g., Dutch and Italian). Only a few 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 psycholinguistic properties in its phonological, lexical, and syntactic structures [7] to crucially test this issue. For example, Chinese is a language, which uses tone to differentiate different lexical items, and different tones often distinguish between different meanings associated with the same syllables. However, tonal information alone does not eliminate lexical ambiguities associated with homophones because Chinese language has a massive number of homophones on a lexical-morphemic level even with the tonal distinctions. From the Modern Chinese Dictionary [8], 80% of the monosyllables (differentiated by tones) in Chinese are ambiguous between different meanings, and 55% have five or more homophones. The single syllable yi with the dipping tone (in case of Mandarin) 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. In case of Cantonese, yi has at least four common meanings in the mental lexicon of nearly every native Cantonese speaker (e.g., easy, two, justice, abnormal, etc.) Upon hearing yi in a sentence, do Chinese speakers activate all 90 or at least four meanings of the 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 because they can use the sentence context.

In the present study, a word-gating experiment was designed to further this line of research and addressed the following questions: How much acoustic information is needed for the spoken word's recognition? Which factors determine whether a homophone will be correctly identified? What affects the speed with which listeners recognize the spoken word? Do different types of context affect the timing of homophonic meanings’ selection? And how do the various factors interact in the recognition process?
2. EXPERIMENT

2.1. Method

Participants. Eighteen native Cantonese speakers (8 male and 10 female, mean age = 29.6) who reported no speech or hearing deficits participated in this experiment. All participants were students from the cognitive psychology classes at the Open University of Hong Kong and they took part in the experiment on a voluntary basis.

Materials and design. Thirty spoken homophones (see Appendix) 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 [9]: 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 [9]. 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, I 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, e.g. prevent using any disyllabic compound words or multisyllabic compound words in the materials.

Two independent variables were manipulated in this experiment; both of them are within-subject variables.

1. Context type: The preceding semantic context was (a) biased to the dominant meaning (more frequent) of a homophone, or (b) biased to the subordinate meaning (less frequent) of the homophone, or (c) ambiguous, which refers to the context being either biased to both meanings of the given homophone or not. The frequency count of different meanings to the homophone is based on Ho and Jiang [10].

2. Homophone neighborhood density (high- vs. low-density): A given homophone had either many associated neighborhood competitors (six or above items) or a few (two items). Neighborhood competitors used in the present study were items that shared both the same segmental and tonal information, which was somehow different from the definition of our previous study [15]. There is also another variable that is nested within the Context type variable, the Dominance of the target homophone, i.e. whether the target meaning is the dominant (more frequent) or the subordinate (less frequent) meaning of the homophone.

An example of the high-density homophone coeng1 (window/gun) and the three corresponding test sentences are given below.

(a) Gaan uk gam guk nei faai di zau heoi hoi sai di coeng1.
This room is so stuffy that you should rush to open all the windows.
(b) Gwan fo cyun gaa waa le di cyun bou dou hai zau coeng1.
Military experts said that all of these are real guns.
(c) Ngo jiu nei dei ji gaa zik hak zau heoi hoi coeng1.
I order you all to go to fire/open your guns/windows.

The complete crossing of the two variables yielded a total of 6 different experimental conditions and also a total of 90 different test sentences.

Experimental Apparatus. The test sentences were read by a female native Cantonese speaker at a normal conversation rate, and were tape-recorded and then digitized into a PowerMac computer. A sampling rate of 22kHz with a 16-bit sound format was used for digitizing. The onset of the Chinese homophone was located as accurately as possible by inspecting speech waveforms and using auditory feedback. Each sentence was gated according to Grosjean [11] and Li [12]. The first gate contained all the words up to, but not including, the target homophone. The second gate consisted of the first gate plus the first 40 ms of the Chinese homophone. The third gate consisted of the second gate plus additional 40 ms, and so on, until the last gate reached the end of the word.

Procedure. Before the experiment began, the experimenter explained the task in Cantonese to the listener. Listeners were told that they would be hearing Cantonese sentences, each cut into small pieces that gradually increased in length. Their task was to identify, for each piece of the sentence, the word that would occur right after the end of the first presentation (i.e., which began after the end of the first gate). They were also told that the word would be a Chinese character. They need to write down on the answer sheet the word that they believed they were hearing and their current confidence level in selecting that word, which they were asked to rate on a 1-7 point scale (1 is least certain and 7 is most certain). They were requested to make a response at each time.

The 18 participants were randomly assigned to three groups of six. Each group randomly received an equal number of sentences for each experimental condition in the (Context) x 2 (Homophone density) design. Each listener received about 270-280 gates in the experiment (i.e., an average of 9.1 gates for each of the 30 sentences) and they heard successive gates of different sizes for each word. The order of presentation for the sentences was pseudorandomly arranged such that no one heard the same homophone twice across the six conditions.
All participants did the experiment individually. A computer program called PsyScope [13] controlled the presentation of the sentence materials. Listeners heard each sentence via two amplified speakers connected to a PowerMac computer. They pressed the computer spacebar to hear the next successive gate. The time interval between any two gates was controlled by the listener because different listeners may require different amounts of time to write down the answer [12]. The whole experiment lasted for one hour.

2.2 Results

The word-isolation results can be roughly divided into two categories depending on where the isolation point occurred: (a) before the acoustic offset of the target word, (b) after the acoustic offset of the word, i.e. never within the sentence frame. The results indicate 79% of target words belonged to the first category, 21% to the second category. Examining further the effects of different variables on the word-isolation process, percentages of the acoustic information needed for the correct identification for each word in the first category were calculated. For the words in the second category, the isolation time was replaced by the total time of the word, following Grosjean [11]. Table 1 presents the average percentage (%) of a word required to a correct identification as a function of context and homophone neighborhood density.

<table>
<thead>
<tr>
<th>Context</th>
<th>Homophone neighborhood density</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High density</td>
</tr>
<tr>
<td>Dominant</td>
<td>22.85%</td>
</tr>
<tr>
<td>Subordinate</td>
<td>28.21%</td>
</tr>
<tr>
<td>Ambiguous</td>
<td>64.70%</td>
</tr>
</tbody>
</table>

Table 1. Average percentage (%) of a word required to a correct identification as a function of Context type and Homophone neighborhood density

A 3 (Context) x 2 (Homophone neighborhood density) repeated measure analysis of variance (ANOVA) was conducted on the word-isolation data. ANOVA revealed that there were significant main effects of Context, F (2,34) = 62.71, p < .001; and Homophone neighborhood density, F (1,17) = 11.25, p < .005. The interaction between Context and Homophone neighborhood density also reached a statistically significant level, F (2,34) = 15.68, p < .001.

First of all, the main effect of context type indicated that provided a priming context, listeners can identify the homophone meaning based on less acoustic-phonetic information. On the average, only 26% of the word was needed for the recognition if the homophone occurred in a highly constrained context, compared with 55% if there was not a constrained context. The results are consistent with many other studies in spoken word recognition [2, 4, 11, 12, 14, 15, 16]. For the present set of data, the average isolation times were 106.7ms and 187.2ms for words following a biased sentence context and an ambiguous sentence context, respectively. These results match well with those estimates from Grosjean [17] and Marslen-Wilson [16]. Therefore, the results confirm the prediction that the context effect occurs immediately, shortly after the acoustic onset of the ambiguous word.

Second, the homophone neighborhood density effect proposed that the number of competitors within an identical syllable would influence the processing time to select the correct homophone. Results show that the larger the neighborhood density of a given homophone, the more acoustic information listeners needed to recognize the correct homophone. Collapsed over levels of sentence context, the mean isolation time for the high-density item was 138ms and that for the low-density item was 130ms. The difference was particularly prominent in the ambiguous context condition, approximately 26ms. The results, thus, suggest that homophone neighborhood density information can affect the narrowing down process of the access to lexical meanings [cf. 15].

Furthermore, the post hoc comparisons (Tukey HSD test) showed that the overall significant density effect mainly came from the interaction between ambiguous context and homophone neighborhood density, p < 0.05. That is, when following an ambiguous context, the more the homophone density, the more time needed to identify the correct lexical meaning. More information is needed to resolve the ambiguity since not much top-down information can help early on. Thus, listeners depend mainly on the bottom-up acoustic information. Nevertheless, the more words shared with the same sound pattern, the slower to select the right target.

Third, differences between the two biased-context types imply that the frequency of individual homophone meanings would also affect the narrowing down process. The mean isolation point was 23.4%, when the context was biased to the dominant homophone meaning and 27.6%, when the context was biased to the subordinate homophone meaning. Although the difference (4.2%) was too small to produce a significant result after conducting a post hoc test, it still pinpoints the important role of the frequency of the individual homophone meaning played during the disambiguation process [18].

3. GENERAL DISCUSSION

The goal of the present study is to understand the role of context effects, effects of homophone density and frequency of individual meanings, in the recognition of Chinese homophonic meanings during continuous speech. Results from the gating experiment provided further evidence to, first, the immediacy of context effect occurring during sentence comprehension. Second, the results implied the density of neighborhood competitors to a homophone clearly affects the isolation time for the homophone. Moreover, dominance of the homophonic meaning could also influence the selection process with reference to the present set of data. Finally, the word-isolation results match well with many other established studies in spoken word recognition [11, 12, 16].

These sort of results are best accounted for by interactive activation models of the sort in Kawamoto [19], Marslen-Wilson [16], and McClelland [6]. 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 a very early stage, rather than at a later stage at which context
effects only follow the completion of lexical access. In a connectionist perspective, the recognition of homophonic meaning can be viewed as an interactive process of constraint satisfaction: multiple sources of constraints (phonological, lexical, and contextual, etc.) 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 interaction among these sources of constraints, each of which may have different weights at a given stage [20].

4. ACKNOWLEDGEMENTS

I would like to thank Hsuan-Chih Chen, Mary Erbaugh, and Ping Li for their comments to the present study. Please address correspondence to Michael Yip, School of Arts & Social Sciences, The Open University of Hong Kong, 30 Good Shepherd Street, Homantin, Kowloon, Hong Kong SAR. Email: myip@ouhk.edu.hk.

5. REFERENCES