Disambiguation of Tonally Unspecified Mandarin Syllables

Christopher Huff[1], *Elinor Payne*[2]

[1] Faculty of Linguistics, Philology, and Phonetics, University of Oxford and Hertford College, Oxford, United Kingdom
[2] Phonetics Laboratory, University of Oxford and St Hilda's College, Oxford, United Kingdom
christopher.huff@ling-phil.ox.ac.uk, elinor.payne@phon.ox.ac.uk

Abstract

The interpretation of written Mandarin monosyllables presented in tonally unspecified *pinyin* was investigated for 22 native speakers. Results showed a preference for high relative lexical frequency and, to a lesser extent low homophone density candidates, with some interaction between these two factors: where one candidate in a homophone cohort had a particularly high frequency, cohort density was found to play little or no role in selection. Where the frequency of the two most frequent candidates was comparable, characters from lower density tonal homophone cohorts were preferred. These results indicate a strong role for relative lexical frequency and, through the weaker homophone density effect a lesser role for tone, in Mandarin homophone disambiguation.
Index Terms: tone, Mandarin, homophony

1. Introduction

This paper investigates the role of tone, and distributional factors such as lexical frequency and density/cohort size, in the interpretation of ambiguous Mandarin monosyllables presented visually. The extent to which frequency and homophone cohort size effects interact and influence the process of lexical choice may provide insight into the role tone plays in the organisation of homophony in the mental lexicon - for example, whether two morphemes differing only in tone (e.g. 妈 *ma1* 'mother'; 马 *ma3* 'horse') are treated as phonological neighbours in the same way as 马 *ma3* 'horse' and 米 *mi3* 'rice', or as something more akin to homophones. Mandarin, having both lexical tone and extensive homophony, is an ideal language in which to explore these questions.

1.1. Homophony in Mandarin

Mandarin contains c.400 segmentally unique syllables, c.1,255 when tone is included [1], and c.6,000 characters/morphemes in modern use [2]. If the characters were divided evenly across syllables, there would be 15 characters per syllable, and about 5 characters for every syllable+tone combination. While the actual distribution of characters to syllables is not remotely this even (*yi* having some 81 characters, of which 43 are *yi4*, for example), it is immediately clear that Mandarin is rich in homophony, by any definition. The delineation as to what precisely constitutes homophony in Mandarin is of central concern to this work and at least three possible working definitions may be identified:

1) 湾 *wan1* 'bay'; 豌 *wan1* 'pea'
2) 豌 *wan1* 'pea'; 碗 *wan3* 'bowl'
3) 豌 *wan* 'pea'; 碗 *wan* 'bowl'

Arguably (1) represents the standard definition of a homophone, in which two words have 'identical' segmental strings and the same tone, while (2) represents some kind of *near* homophony, in which two words have the same segmental string yet different tones. Li and Yip have, at different times, used both definitions [3][4]. For present purposes, examples as in (1) will be termed *full homophones*, and as in (2) *segmental homophones*, with the caveat that it is not necessarily the case that such 'segmental homophones' are regarded or processed as such by Mandarin speakers.

Scenario (3) combines aspects of the first two types. As in (1) the two syllables are identical *as presented*, and there is no information that conflicts with homophony. However, as in (2) the two syllables may (or may not) have different underlying tones. Unlike (1) or (2), they are homophones only as a result of an impoverished (or underspecified) representation. Examples as in (3) will be termed *apparent homophones*.

Pinyin has a complex relationship with reading and writing, and therefore with the written representation of homophones. It is rarely used for communication except by young children (when it is written with tonal diacritics), though it is often used in typing Chinese characters, where the segments are input without tone and the intended character is selected from a generated list. There are however certain situations in which native speakers do communicate using tonally unmarked pinyin, for example in informal online communications or text messages. Encountering a scenario like (3) therefore, though undoubtedly an unusual and not entirely natural occurrence, is not wholly alien. That scenarios like (3) do occur suggests that tone is not a mandatory component in the disambiguation of homophones, despite the fact that the cohorts of potential homophone candidates are considerably enlarged when tone is excluded.

There is some evidence that tone is processed and organised by Mandarin speakers differently than segmental material, and indeed that syllables which differ only in tone are sometimes treated as homophonous. For example, the perception of tones has been found to be slower than vowels, preventing an early role for disambiguation [5]. Perhaps more importantly, there is evidence that syllables which match in segmental composition but not in tone do activate each other. Lee [6] found direct priming effects in Mandarin only when the prime and target had the same segmental structure and tone, but in mediated priming at short inter-stimulus intervals (50ms) found priming effects when the target and prime matched in segmental structure, regardless of tone - indicating activation of the *segmental homophone* cohort, even when the tones were mismatched.

Taft and Chen [7] presented two characters in sequence to Mandarin speakers, and asked them to judge whether the two were pronounced identically. Half of the pairs were full homophones, and half were not. There were three types of non-homophone - entirely different syllables (i.e. different segments and tone); syllables with the same tone as the stimulus, but a different vowel; and *segmental homophones* (same segments, different tone). They found that when the two characters were segmental homophones, responses were slower with considerably more errors than for the other two types of non-homophone. That tonal differences are so under-utilised compared to segmental differences in homophone judgements perhaps indicates a less significant role for lexical tone in distinguishing lexical items than is generally assumed.

1.2. Distributional Factors

There are a number of factors known to affect the process of lexical access, two prominent examples with particular relevance for this paper being lexical frequency and neighbourhood density [8].

Lexical frequency has been shown to affect the speed of lexical access in Mandarin, as in other languages [9][10]. High frequency words are more probable and are accessed more quickly than low frequency words. It is expected that the interpretation of ambiguous words will be affected by lexical frequency, with higher frequency characters more likely to be selected than lower frequency characters. The primary frequency statistic used throughout the experiment is *character frequency* (abbreviated to *frequency*).

The limited nature of the Mandarin syllable inventory and the high degree of homophony result in neighbourhood densities for most syllables being extremely high. Mandarin syllables are also highly differential in homophone densities, and the artificial exclusion of tone allows for potential manipulations of these densities (i.e. by artificially neutralising tonal contrasts). For these reasons, the primary density statistic in this paper is full homophone density, which is expected to play a role in the interpretation of tonally ambiguous syllables in Mandarin. Previous studies have found conflicting results of homophone density in Chinese languages - albeit with significant discrepancies in how to categorise and define homophone density [3][4][10].

Chen et al [10], for example, conducted two experiments on Taiwanese Mandarin designed to test for lexical frequency and homophone density effects. A speed/accuracy advantage was found for high frequency characters in both Naming and Lexical Decision. Both tasks also showed an advantage in speed and accuracy for high density characters, but only when these were also of low frequency. One potential problem with this result is that full homophone densities greater than 7 were defined as 'high', and less than 5 'low'. As has been mentioned, the density range in Mandarin is much greater than this definition may indicate, and can be as high as 50 or more. Thus, comparing homophone densities of e.g. 4 and 8 may not allow for a sufficiently sensitive test of differences in density.

The present study seeks to examine these factors in greater depth. An experiment was designed to investigate the effects of lexical frequency, homophone density, and tone in isolated syllables/morphemes. The motivation for using isolated syllables was to establish the extent to which these factors interact with each other in guiding lexical choice without the confound of contextual biasing factors, as well as to sidestep numerous complex issues involved in studying ‘words’ in Mandarin. The definition of the word is somewhat controversial for Mandarin, where for example studies on the segmentation of sentences into constituent words have found considerable disagreement between participants [11]. The monosyllabic morpheme/character is a more widely familiar unit.

2. Method

2.1. Participants

22 native speakers of Mandarin, 11 females and 11 males ranging from 18 to 36 years of age, were recruited from the University of Oxford, and paid to participate in the experiment. All self-identified as native speakers of Mandarin, and were familiar with the pinyin romanisation, having learnt it in elementary school as part of the national curriculum.

2.2. Materials and Elicitation

The stimuli consisted of five segmentally unique Mandarin monosyllables, varying in overall frequency and density - one high frequency (HF) and high density (HD) syllable (*yi*); one HF/medium density (MD) syllable (*shi*); one medium frequency (MF)/low density (LD) syllable (*wan*); and two LF/LD syllables (*miao* and *pang*).

This variability was included in order to test whether the predicted pattern of relative lexical frequency and homophone density effects interacted with, or were dependent upon, *absolute* values of frequency or density. For example, it may be the case that frequency is highly determinative of lexical selection only above a certain frequency threshold.

Participants were presented with each monosyllable stimulus in pinyin, one at a time, on a computer screen. Beneath each pinyin stimulus were two characters: participants were asked to select, as quickly as possible, the one which they felt corresponded best to the stimulus.

Three conditions were built into the experiment. In the first, a HF character was paired with a LF character. This was designed to provide a straight-forward examination of the lexical frequency effect, and it was expected that the HF character would be selected. In the other two conditions frequency was roughly matched, i.e. two HF characters or two LF characters were paired. Here it was predicted full homophone density would guide selection.

The full set of character combinations which participants had to choose from was generated by taking the two highest and the two lowest frequency characters associated with a given segmental syllable (e.g. from the entire *yi* syllable), and presenting all possible pairs. There are thus six possible pairs per syllable, across five test syllables, giving a total of 30 test screens seen by each participant. Participants were given a time limit within which to complete the experiment, based on a marginally sped-up average from a pilot, in order to establish moderate time pressure.

3. Results

3.1. yi (high frequency/high density)

The results for *yi* (see Table 1) show a strong preference for higher frequency characters. In all three conditions the higher frequency option of the two was predominantly chosen. Where the frequency difference between the choices was large both in absolute and proportional terms, the preference for the HF option was very nearly nearly categorical. In one combination, the difference between the two presented character options was large (c.2million), but the ratio was fairly small (3.35/1): here the higher frequency character was still preferred 91% of the time, despite having a higher full homophone density (11 compared to 7). This frequency effect appeared to hold even when the frequency difference was relatively small (722 occurrences to 19). However, participants' choice here may have been influenced by the considerably lower full homophone density of the HF option (19 compared with 43). Alternatively, it may be that what counts is the proportional frequency difference - which was quite large in this combination - rather than the absolute difference.

yi Slide		# of Responses	Freq.	Freq. Diff.	Freq. Ratio	Density
1	一	22	3050722	3050703	160,564.32	11
	劓	0	19			43
2	一	22	3050722	3050000	4,225.38	11
	沂	0	722			19
3	以	21	910627	910608	47,927.74	7
	劓	1	19			43
4	以	18	910627	909905	1,261.26	7
	沂	4	722			19
5	一	20	3050722	2140095	3.35	11
	以	2	910627			7
6	沂	18	722	703	38	19
	劓	4	19			43

Table 1. Full results for yi.

3.2. *shi (high frequency/medium density)*

As with *yi*, the higher frequency character of any given pair was most frequently chosen in all cases (see Table 2). Interestingly, in the one combination for which frequencies were close (233 compared to 52), with a large difference in homophone density (28 to 9), the higher frequency/density option was chosen 82% of the time - suggesting that although frequency was still the main determining factor in lexical choice, density may be playing a small role.

shi Slide		# of Responses	Freq.	Freq. Diff.	Freq. Ratio	Density
1	是	20	2615490	2615438	50,297.88	28
	莳	2	52			9
2	时	22	833532	833480	16,029.46	9
	莳	0	52			9
3	时	19	833532	833299	3,577.39	9
	弑	3	233			28
4	是	22	2615490	2615257	11,225.28	28
	弑	0	233			28
5	是	16	2615490	1781958	3.14	28
	时	6	833532			9
6	弑	18	233	181	4.48	28
	莳	4	52			9

Table 2. Full results for shi.

3.3. *wan (medium frequency/medium density)*

The four combinations in which the frequency difference (both absolute and proportional) between characters was large produced similar results to those reported for *yi* and *shi* (see Table 3). What is of particular interest for this syllable is the combination when frequencies were similar but, there were small differences in density. The two most frequent characters are moderately different in absolute but not relative frequency (c.9,000; ratio 1.07/1), and both have low densities (4 and 2) - which is to be expected, as the overall homophone density of this syllable is low. Nevertheless, when overall density is low and frequencies are similar, it would seem that even a small difference in density can play a decisive role in lexical selection: the lower frequency, but lower density character was chosen 91% of the time. In Slide 6 (LF/LF) the slightly lower frequency of the two was chosen 77% of the time, despite the two having the same homophone density. Here we may presume that the small magnitude difference in frequency results in frequency effects being effectively neutralised, though we can only speculate as to what precisely is driving lexical choice here.

wan Slide		# of Responses	Freq.	Freq. Diff.	Freq. Ratio	Density
1	完	20	143120	142,667	315.94	4
	豌	2	453			4
2	万	21	134220	133,599	216.14	2
	蜿	1	621			4
3	万	22	134220	133,767	296.29	2
	豌	0	453			4
4	完	20	143120	142,499	230.47	4
	蜿	2	621			4
5	完	2	143120	8,900	1.07	4
	万	20	134220			2
6	蜿	5	621	168	1.37	4
	豌	17	453			4

Table 3. Full results for wan.

3.4. *miao*

The four combinations in which the frequency differences were fairly large produced strong frequency effects (see Table 4), though slightly weaker than for *yi*, *shi*, and *wan*. There are several possible reasons for this. First, the absolute frequencies of the characters are low, even for the highest frequency characters. Second, the frequency differences are also much lower (c.20,000 in each case) than for previous syllables (with the exception of *wan* combination 5, in which a density effect was found). Where frequencies and densities are quite similar (combination 5), responses are more evenly shared (64%-36%) than in any other case. In combination 6, whilst the absolute frequency difference is very similar to that in combination 5, the proportional difference is greater, and indeed there is a greater preference for the more frequent option (73% of responses). This suggests that, particularly for syllables with a low overall frequency, proportional differences in character frequency can play an important role in lexical decision.

miao Slide		# of Responses	Freq.	Freq. Diff.	Freq. Ratio	Density
1	妙	19	20165	19972	104.48	3
	邈	3	193			7
2	描	17	20247	19959	70.3	3
	淼	5	288			7
3	妙	18	20165	19877	70.02	3
	淼	4	288			7
4	描	18	20247	20054	104.91	3
	邈	4	193			7
5	描	8	20247	82	1	3
	妙	14	20165			3
6	淼	16	288	95	1.49	7
	邈	6	193			7

Table 4. Full results for miao.

3.5. *pang (low frequency/low density)*

The combinations for *pang* are broadly divisible into three frequency-difference categories. In three of the cases frequency differences were around 30,000. For one of these, the frequency ratio was very high (c.90/1), and 100% of responses were for the HF item (despite having a higher density). For another, the frequency ratio was c.53/1, and 86% of responses were for the HF option (same density). For the final case in this category, the ratio was only 3.8/1, and the HF option attracted only 59% of responses. The second category of absolute frequency-difference is c.8,000, with a relatively high proportional difference in both combinations. For both, the responses were strongly in favour of the HF item - 86% for the combination with a frequency ratio of 14/1, and 95% for the ratio of 24/1 (with a higher homophone density of 4 to 2). The final category was a small frequency-difference (614 vs 356 occurrences, with a ratio of c.1.72/1). The lower frequency (but lower density by 2 to 4) was chosen 64% of the time.

pang Slide		# of Responses	Freq.	Freq. Diff.	Freq. Ratio	Density
1	旁	22	32268	31912	90.64	4
	滂	0	356			2
2	庞	19	8531	7917	13.89	4
	螃	3	614			4
3	庞	21	8531	8175	23.96	4
	滂	1	356			2
4	旁	19	32268	31654	52.55	4
	螃	3	614			4
5	旁	13	32268	23737	3.78	4
	庞	9	8531			4
6	螃	8	614	258	1.72	4
	滂	14	356			2

Table 5. Full results for *pang*.

4. Discussion

Across a number of syllables varying in the number and frequency of associated characters, the primary factor determining lexical interpretation (character selection) was absolute lexical (character) frequency. This finding is compatible with frequency effects found more generally, for non-homophonous selection tests.

However, there are ways in which the power of this raw frequency effect may be limited. If there is a competing character of similar proportional frequency, the absolute difference in frequency can be fairly large without having a strong effect on selection. Such a scenario naturally arises more readily amongst characters which are higher in absolute frequency, as the higher the frequencies the greater the difference between them can be without having much effect on the ratio. For very low frequency characters, such measures are likely to be less reliable. A difference of 20 occurrences between two characters in a corpus of 200,000,000 characters is not likely to be a particularly salient difference, and even if the ratio difference is 20 to 1 unlikely to be useful in predicting how speakers will respond.

Thus it would appear that it is proportional differences in frequency which count more than absolute differences, and that speakers are sensitive to such distributional statistics. Every combination which had a ratio of less than 2/1 showed an unreliable frequency effect, e.g. frequency alone was not determinative of responses. For most of these cases the absolute frequencies involved were low (>1000), but not all - the two characters in combination 5 of *wan* had frequencies of approximately 140,000 each.

There are interesting implications here for the way frequency effects shape homophone disambiguation. We might expect that as frequency increases, and lexical activation increases in speed and strength, interpretation would pattern accordingly. If the system is sensitive to differences of only several hundred occurrences, then we would expect to find that a difference of c.10,000 occurrences to result in a very strong frequency effect, no matter the relative frequencies. However this is precisely what we did not find. The apparent importance of proportional differences in frequency would indicate a role for a determining factor which is interpolated from frequencies - we call this *prominence*. A prominent character is one which stands out in terms of frequency in relation to its competitors. Raw frequency may be an important component of this, but only in relation to the whole system. A potentially interesting avenue for future research would be to examine 'prominence' in relation to the time-course of lexical activation, i.e. investigating whether this competition in interpretation due to frequency interacts with inhibition effects due to density.

The other factor investigated was full homophone density. Since full homophones include tones, a strong effect of this factor would indicate a central role for tone in the organisation of homophony in the lexicon and in the shape of 'prominence', and hence the disambiguation of homophones. We found little evidence of tone affecting the pattern of responses, even in the absence of strong frequency biases (i.e. when the frequency ratio was low), except for one case - combination 5 of *wan* showed a strong preference for a lower density yet lower frequency character. This is a special case for two reasons - it is the only relatively high frequency pair in the experiment in which the frequency ratio is low, and it is the only pair in which the ratio is low but the absolute difference is fairly large. There are only a few cases where the ratio is low *and* the densities of the two characters are different, limiting the amount of data on this particular scenario. However, it would appear that, though frequency effects dominate, when these are neutralised, density effects may become active in constructing relative prominence.

5. Conclusion

In this study it was found that lexical decisions were strongly influenced by relative lexical frequency, and to a considerably lesser extent by full homophone density (e.g. including tone). For a given syllable, the character of the highest frequency is generally chosen. If there is a relatively large difference in frequency between this character and the next most frequent character, responses will be strongly biased (and indeed were sometimes found to be categorically biased in the experiment). When the disparity in frequencies was relatively small between competitors, however, the character with the lower homophone density may be preferred. These results indicate that when a Mandarin speaker is presented with a syllable stripped of tone, the resulting cohort of possible competitors is weakly, or perhaps not at all, organised in terms of tone, but is highly structured in terms of frequency.

6. References

[1] Duanmu, San. (2007). “The Phonology of Standard Chinese.” Oxford University Press. Second Edition.

[2] Da, J. (2007). *Chinese text computing corpus*. http://lingua.mtsu.edu/chinese-computing/ copyright.html.

[3] Li, P. and M. Yip (1996). “Lexical Ambiguity and Context Effects in Spoken Word Recognition: Evidence from Chinese.” 18th Annual Conference of the Cognitive Science Society.

[4] Li, P. and M. Yip (1998). “Context effects and the processing of spoken homophones.” *Reading and Writing: An Interdisciplinary Journal* 10: 223.

[5] Ye, Y. and C. Connine. (1999). “Processing Spoken Chinese: The Role of Tone Information.” *Language and Cognitive Processes* 14(5/6): 609.

[6] Lee, C. (2007). “Does Horse Activate Mother? Processing Lexical Tone in Form Priming.” Language and Speech, 50(1):101-123.

[7] Taft, M., and H.C. Chen (1992). “Judging Homophony in Chinese: The Influence of Tones.” In H.C. Chen and O.J.L. Tzeng (Editors). *Language Processing in Chinese*. Elsevier.

[8] Hogaboam, T.W. and C.A. Perfetti (1975). “Lexical Ambiguity and Sentence Comprehension.” *Journal of Verbal Learning and Verbal Behavior* 14: 265.

[9] Peng, D., Liu, Y., and C. Wang (1999). “How is Access Representation Organized? The Relation of Polymorphemic Words and their Morphemes in Chinese.” “Reading Chinese Script.” Lawrence Erlbaum Associates, London.

[10] Chen, H.-C., Vaid, J., & Wu, J.-T. (2009). Homophone Density and Phonological Frequency in Chinese Word Recognition. *Language and Cognitive Process, 24*, 967-982.

[11] Hoosain, R. (1992). “Psychological Reality of the Word in Chinese.” In H.C. Chen and O.J.L. Tzeng (Editors). *Language Processing in Chinese*. Elsevier.