A Study of Mandarin Homophony

Christopher L. Huff, Hertford College

Thesis submitted for the degree of Doctor of Philosophy in General

Linguistics and Comparative Philology

Michaelmas 2017

A Study of Mandarin Homophony

Christopher L. Huff, Hertford College

Thesis submitted for the degree of Doctor of Philosophy in General Linguistics and Comparative Philology

Michaelmas 2017

An investigation into the distribution of homophones was conducted using a large written corpus (Da, 2007). Both tone and homophony were found to be systematically uneven, with a tendency toward clustering. Syllables with the full complement of tones have, on average, higher homophone densities than those with fewer tonally-specified syllables, indicating a beneficial role of increasing homophone density in lexical processing. Three experiments were conducted to test this hypothesis. In Experiment 1, participants were shown tonally-unspecified *pinyin* and asked to produce a character. Participants gave the most frequent character matching the *pinyin*, except when two characters of differing homophone density but similar lexical frequency were in direct competition (e.g. the two highest frequency *wan* characters were associated with different tones). In this case, it was found that the slightly *lower frequency*, but *higher homophone density* character was selected, indicating a facilitative role of homophone density. In Experiment 2, participants were shown tonally-specified *pinyin* followed by characters, and asked to judge whether the two matched. Character frequency (not the cumulative frequency of all homophones) was found to be the primary determinant of response time, with higher frequency characters responded to faster than lower frequency characters regardless of homophone density. However, increasing homophone density was found to speed responses for all frequencies. High frequency characters were always faster than low frequency characters, but low frequency characters with high homophone density were faster than low frequency characters with low homophone density. In Experiment 3, participants read aloud characters embedded in context sentences. High frequency words were more reduced (i.e. shorter in duration) than low frequency homophone mates. High frequency characters with many homophones were shorter than high frequency characters with few homophones, while the opposite held for low frequency characters.

Total Thesis Length: 61,668 words.

Table of Contents

Chapter I. Introduction

The focus of this thesis is Mandarin homophony. The impetus for this work lies in an experience in attempting to speak Mandarin, now some years ago, in which it was noticed that the accurate production of tone (a somewhat challenging obstacle for the novice) when speaking in full sentences was not necessarily essential for reasonably successful communication. A cursory look through a Mandarin dictionary revealed a high degree of homophony for most morphemes, and perhaps a cue as to why tone might not be strictly necessary - as most morphemes, even when properly produced, are ambiguous (i.e. have a number of homophones) at the syllable level, it appeared to be the case that Mandarin speakers are very much used to dealing with uncertainty at the morphemic level. The loss of tonal distinctions (as experienced by an L2 learner and due, in the author's case at the time, to incompetence) expanded the pool of potential morpheme-candidates; but as other factors (for example, context) are required to resolve ambiguity anyway, the result was not as catastrophic as it might otherwise have been. Homophony was evidently not what it might superficially have appeared, its extensive nature not the imagined monolithic obstacle to lexical access, and its presence in the lexicon not simply a binary imposition of ambiguity. Furthermore, its status in the lexicon could not be thought of or dealt with simply as the sporadic homophony of, say, English (or a multitude of other languages) writ large. The aim of this thesis is an in-depth investigation into homophony in Mandarin - how extensive homophony is in practice, how it is structured, and how the properties of the Mandarin lexicon affect how Mandarin 'words' are stored, accessed and produced.

In **Chapter II**, a general outline of Mandarin is given, as well as an introduction to a number of features of the language which may impact on homophony. 'Mandarin' as a term has variable meaning, as it may refer to one of the two branches of Sinitic (Northern/ Mandarin branch), a sub-group within this branch (the Mandarin language/dialect family),

and to a particular language, e.g. Standard Chinese, and the potential impact of this terminological uncertainty on linguistic research is discussed. This chapter also investigates the Mandarin syllable inventory and restrictions on morpheme structure. There are only around 400 unique syllables not including tonal distinctions, and only around 1,255 when tone is considered - compared to around 9,000 in English (Duanmu, 2007). In addition, Mandarin morphemes are almost exclusively monosyllabic, further restricting the number of phonologically unique morphemes. These two interacting linguistic restraints result in an unusually high degree of homophony. Most Mandarin morphemes have homophones, even when tone is specified.

In Chapter II, we established that Mandarin is structurally prone to extensive homophony, and discussed some of the causes of that homophony. In **Chapter III**, we examine the distribution of Mandarin homophony in detail. Mandarin has an average of around 3.2 homophones per tonally specified syllable. An examination of the actual distribution of homophony in a large corpus (Da, 2007), winnowed to exclude extremely rare (low frequency) characters in an approximation of the lexicon of a modern-day literate Mandarin speaker, shows that homophony is systematically uneven, with a marked tendency toward homophone 'clustering'.

Of the very few unique base syllables in Mandarin (disregarding tone), many do not have fully utilised tone paradigms - 59 of the 399 base syllables only have one associated tone. Tonal contrasts, which we might expect to help minimise homophony in Mandarin, are grossly under-utilised. In addition, the distribution of homophony is dramatically uneven. The more tones associated with a unique base syllable (e.g. *luo* + tone 1, *luo* + tone 2, etc.), the more homophones associated with each tonally specified syllable. There are 33 unique base syllables in the corpus with only one associated character. For example, the character 谬 'absurd' is the only character associated with the syllable *miù*, and there are no characters associated with *miū*, *miú*, or *miǔ*. If these syllables had full tonal paradigms, there could be

132 characters with no homophones at all. Yet there are syllables such as *ji*, which has all four tones, and an astonishing 13.5 full homophones per tonally specified syllable. **Chapter III** also discusses some of the special considerations that must be taken into account when examining homophony in a tonal language such as Mandarin. These are both conceptual and terminological. For example, what sort of relationship should we expect tonal minimal pairs (such as *pào* 泡 'bubble', and *páo* 咆 'roar') to have? Are they classic lexical neighbours, as they differ in just one property? Or are they more akin to 'true' homophones, as they are segmentally identical? And how shall we refer to them in any case?

In **Chapter IV**, we ask: on what level are two words homophonous? Do homophones share a unique phonological representation, or is each lemma independently represented? There is evidence that tone is not equivalent to segmental information in lexical processing tasks such as naming, same-different judgements, vowel/tone monitoring, etc., in either speed or accuracy (Taft and Chen, 1992; Ye and Connine, 1999; Cutler and Chen, 1997; Lee, 2007). Does this indicate that tone is not as critical to lexical identity as segmental information? Are segmental homophones in some sense 'closer' than phonological (segmental) neighbours? The effect of homophone density on lexical access and production may provide information regarding these questions. Is having higher full homophone density 'beneficial', 'harmful', or neutral (or some combination, depending on a frequency and density interaction)? There are two aspects to this - 1) locus of frequency effects; and 2) type of homophone processing model – differing assumptions about which lead to quite different predictions. For any model assuming an abstract phonological form is the locus of frequency (i.e. all homophones share one phonological representation), increasing homophone density should have a positive effect on lexical access (through frequency inheritance), particularly, and arguably uniquely, for low frequency words, which benefit most clearly from the cumulative frequency of neighbours (high frequency words are already high in frequency, and increase in cumulative frequency only marginally by adding low frequency neighbours). We should find clear differences in

response times for cumulative frequency of all full homophones. If all homophones share a single phonological representation, all homophones have effectively the same frequency (i.e. frequency inheritance). Exhaustive-access of homophones would imply that all meanings are activated simultaneously, as frequency is not a factor in choosing between lemmas that share a phonological form. If frequency is associated with each meaning independently however (usage-based, though also compatible with abstract representational models where frequency is associated with the lemma), then increasing homophone density will have some effect in exhaustive access - all forms activate, but higher frequency lemmas are faster/stronger than lower frequency lemmas.

Finally, there is evidence that words conventionally considered to be homophones (e.g. *time*/*thyme*) can be distinct in production, particularly in terms of duration, as a function of lexical frequency (Gahl, 2008). These differential effects on the production of homophones recall findings of *lexical frequency effects* on the comparative durations in non-homophones (cf Bell et al, 2009; Bybee, 2001), with more frequent words more likely to shorten and undergo other phonetic reduction effects. They also mirror certain *neighbourhood density effects* on production, for both duration (cf Kilanski, 2009; Gahl et al, 2012) and other variables such as degree of vowel dispersion (cf Wright, 1997, 2004; Munson and Solomon, 2004; Munson, 2007) and degree of coarticulatory effects (e.g. Scarborough, 2009). A usage-based model might claim that 'homophones' have independent lexical representations, and are thus in some sense not homophones at all. It is unknown whether a distinction in duration such as this is utilised in perception and thus in lexical access more generally, however, and whether such pairs are processed in word retrieval as homophones, or merely as near-homophones. The extent of homophony in Mandarin makes it extremely unlikely that each word in a homophone cohort of dozens can be sufficiently distinct in terms of duration, or any other phonetic variable, to give rise to dozens of distinct emergent categories. However, given Gahl's (2008) findings, it is at least possible that *some* of the conventionally-defined

homophones in Mandarin are phonetically distinct from each other in production and potentially also in perception, and thus in possession of an autonomous representation.

In **Chapter V**, an experiment was conducted (Experiment 1) to test for effects of lexical frequency and homophone density. Experiment 1 provided tonally impoverished *pinyin* stimuli, and asked subjects to write the first matching character which occurred to them. It was predicted that the highest frequency character for a given base syllable (e.g. *miao*), by dint of having the strongest lexical representation, or at least the strongest 'phonological form to lemma association', would be activated most quickly, and therefore be most likely to be given as a response, and this was generally confirmed, across a wide range of frequencies. It was also predicted that a higher full homophone density would be of benefit in terms of speeding lexical access, particularly for lower frequency characters, and this was at least partially supported by results.

Chapter VI reports a further experiment (Experiment 2) that refines the scope and purpose of Experiment 1. In Experiment 2, a tonally specified *pinyin* stimulus (e.g. pàng) was presented, followed by a single character that participants had to judge as either matching or mismatching. For matching sets of *pinyin* stimulus and character, it was predicted that responses would be conditioned primarily by lexical frequency, and secondarily by full homophone density. Higher frequency was expected to speed responses generally, while increasing full homophone density was expected to speed responses to lower frequency characters. For matching pairs, a general frequency effect was indeed found, corresponding to lexical frequency (and not cumulative full homophone frequency, indicating a weak, or non-existent frequency inheritance effect). However, against expectations, a speeding effect of full homophone density was found for all characters, regardless of frequency. For mismatching pairs, responses differed according to type of mismatch. When the segments of the target failed to match those of the *pinyin* stimulus, regardless of tone, participants were both faster and more accurate in rejecting the pair as a match. Characters which either completely

mismatched the stimulus, and those which mismatched in segments but had matching tone, were indistinguishable in response times and accuracy. However, when pairs mismatched only in tone, response times were slower, indicating greater difficulty in rejecting the pairs as homophonous. This is in accordance with expectations and previous research, which has indicated that segments are more critical in determining lexical identity than tones.

Chapter VII reports an experiment conducted (Experiment 3) to investigate whether lexical frequency and homophone density, which were found to affect lexical access for visually presented characters in Experiments 1 and 2, might also affect fine phonetic detail in production. It was expected that higher frequency, higher density characters would experience greater 'reduction', as such characters generated the fastest response times in Experiment 2, and also given previous studies on homophony in English (Gahl, 2008) and the effects of density on phonological neighbours (Gahl et at al, 2012). Syllable duration was chosen as the variable for examination as having the greatest likelihood of exhibiting reliably measurable indices of phonetic reduction. Ten full homophone pairs were tested, each with one high and one low frequency character, and varying in full homophone density. Participants read the target characters in biased context sentences and in neutral context 'carrier' sentences. Unlike earlier studies which found robust effects only in contextual speech, systematic effects were found only in the neutral, carrier sentence task. High frequency words were systematically shorter on first repetition than low frequency effects, for both high and low density cohorts, and also underwent fewer shortening effects with repetition. However, density appeared to interact with frequency: high homophone density further boosting the phonetic reduction of high frequency words, but further inhibiting phonetic reduction in low frequency words. This result echoes findings by Munson and Solomon (2004), who found low frequency, high density words undergo the least amount of vowel reduction (i.e. greater vowel dispersion).

Finally, in **Chapters VIII (Discussion) and IX** (**Conclusion**) we discuss the findings of our analysis of the distribution of homophony in Mandarin (**Chapter III**) and of Experiments 1-3

(**Chapters V**, **VI** and **VII**) and draw conclusions for our understanding of homophony, how it is stored and the process of lexical access. In particular, we discuss i) the divergent findings on the effects of homophone density on lexical access and production (with density adding a boost to lexical access across different frequencies, but interacting with frequency for production); ii) the ways in which our findings on the effects of frequency and density diverge, to some extent, from those reported for phonological neighbours. We seek to explain some of these divergences as an artefact of the different nature of the tasks involved in perception/lexical access vs lexical access/production, and also the different nature of each of these two types of process for homophones vs phonological neighbours; and iii) our findings on the role of tone in the phenomenon of homophony in Mandarin. The pressures involved in lexical access from a homophonous input (finding a needle in a haystack) are different from those involved in producing an unambiguous output (making the needle in said haystack more visible). Likewise, the process of evaluating competitors in a phonological neighbourhood (different forms) is different from that of evaluating competitors in a homophone cohort (same, or quasi-identical, form).

We also discuss the findings on tone and the implications for its role in homophony, and more widely for the nature of homophony in general, the extent to which homophony is categorically distinct from a phonological neighbourhood, and accounts of phonological representation that can plausibly capture the findings we present. Most models of lexical organisation and representation were not designed with extensive homophony in mind, and a number of suggestions are put forward for how we might go about incorporating the findings of this thesis into more general models. We also discuss potential further investigations necessary for building such a model.

Chapter II. Introduction to Mandarin

This work is concerned with the Mandarin Chinese language. As a number of topics to be discussed in later chapters require some knowledge of the basics of Mandarin, this chapter is primarily devoted to providing a brief overview of Mandarin's place in the broader Sinitic language family, and the phonetic and phonological structure, tone, grammar, and orthography of Mandarin.

1 Sinitic and the Chinese Language(s)

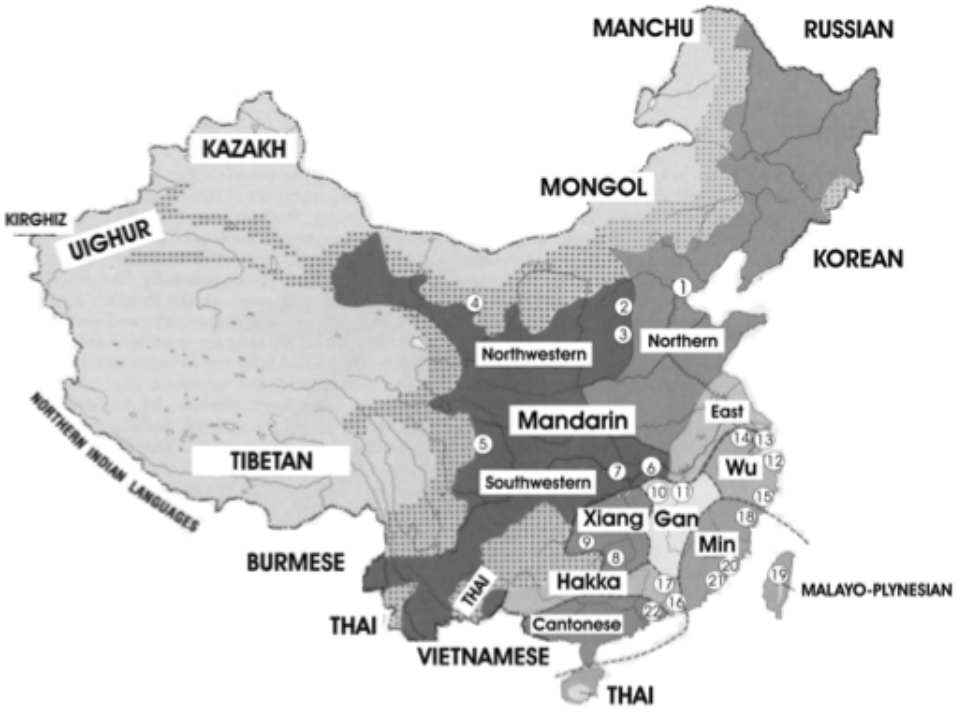


Figure 2.1. Language Map of Sino-Tibetan. The seven 'dialect families' of Sinitic are shown, along with a number of constituents. ***Mandarin***: 1) Beijing, 2) Taiyuan, 3) Wuhan, 4) Ningxia, 5) Chengdu; ***Xiang***: 8) Changsha, 9) Shuangfeng; ***Gan***: 10) Anyi, 11) Nanchang; ***Wu***: 12) Ningbo, 13) Shanghai, 14) Suzhou; ***Hakka***: 16) Liancheng, 17) Meixian; ***Min***: 18) Fuzhou, 19) Taiwan, 20) Xiamen; ***Yue***: 22) Guangzhou. Map taken from Hamed and Wang (2006).

Some care is required when using either 'Mandarin' or 'Chinese' in regard to language, as the two are variably employed. In the broadest sense *Chinese* is a synonym for Sinitic, a language family part of the greater Sino-Tibetan family (LaPolla 2010; Chao, 1968b; Norman, 1988), and is used to refer to the language family generally, or to its constituents, such as Cantonese (Li and Yip, 1998). Sinitic can be broadly divided into two main branches, a Southern branch and a Northern/Mandarin branch, each of which consists of a number of 'dialect families', or sub-groups which in turn consist of a number of related dialects/languages. There are around seven major dialect families, the exact number dependant on the specific source: for example, Duanmu (2007) and Hamed and Wang (2006) both give 7; Tang and van Heuven (2009) give 8; and Chao (1943; 1968b) is undecided as to whether it is 8, 9, or 10. A map marking out a number of the major dialect families and their approximate areas is given in Figure 2.1.

The Northern/Mandarin branch of Sinitic is the larger of the two, both in terms of the number of speakers and in geographical range, with about 70% of the population of the People's Republic of China speaking a language/dialect within this branch (Duanmu, 2007; Norman, 1988). Mandarin as a term is itself not without ambiguity and classificatory controversy (c.f. Tang and van Heuven, 2009), and can variously refer to one of the two major divisions of Sinitic (the Northern/Mandarin branch just mentioned), to a sub-group within this binary division (the Mandarin dialect/language family within the greater Northern/Mandarin family), and to a language within the Mandarin 'dialect family' corresponding to what is also known as Standard Chinese (Coblin, 2000).

There have been standardised forms of Chinese, in some sense, in use since at least 700 BC (Duanmu, 2007). These come in two distinct types - written standards, and spoken standards. The most well documented is the written standard *wenyan*. In the period roughly from the beginning of the Spring and Autumn Period (770 B.C.E.) to the end of the Han dynasty (220 C.E.), the written and spoken varieties of Chinese were largely coextensive. The written language of this period was frozen, and was used as the literary standard until 1919

with relatively little change (Norman, 1988). It is important to note that this literary language, after its formation period, was not generally spoken. Individuals pronounced the characters in their own modern dialects (Norman, 1988).

Historical standard spoken languages are more difficult to track, as they were rarely written down. In a sense China had no standard national language until the 20th century, apart from *wenyan,* which was not spoken (Chao, 1961). Particular languages were employed as 'court languages', though these were known only by 'a tiny minority of officials and merchants' (Norman, 1988). These 'court languages' have been known by an assortment of names - as *yayan* 'refined speech' in the Spring and Autumn period (722-482 BCE); *tongyu* 'common speech' in the Han dynasty (206 BCE-220 CE); and *guanhua* 'language of officials' (or 'Mandarin') since the Ming Dynasty in the 14th century (Duanmu, 2007).

It is commonly assumed that *guanhua* has been based on the language of Beijing since the early part of the 13th century, when that city first became the capital during the Yuan dynasty (Norman, 1988; see commentary by Li, 2004). However, relatively recent research by Coblin (2000) into Western missionary records has contradicted this assumption. Missionaries to China wrote detailed accounts of both the current spoken and written language varieties they encountered, for use as teaching and learning aids for future missionaries. Though it is true that the political capital of China moved from the southern city of Nanjing to the northern city of Beijing in the 13th century, Nanjing remained the cultural centre of China for hundreds of years. As such, the variety spoken in Nanjing remained the prestige variety, and the court language, until the end of the 18th century, when the Beijing variety gradually took over as the court standard (Coblin, 2000).

In 1919 an attempt was made to create a national language, at least in terms of standard character pronunciation, in the form of the *Guoyin Zidian* (Chao, 1961). This work is an artificial amalgam of the speech of a number of dialects, though still heavily based on the classical literary speech (Norman, 1988; Chao, 1961). In order to provide a pronunciation

guide for teachers of this newly minted language, Y.R. Chao was asked to make phonograph recordings for language teachers, making him the only speaker of the national language of China for a period of around 13 years. This artificial language failed to gain traction, and in 1932 the standard was changed to be largely (but not entirely) coextensive with the Beijing dialect in pronunciation, differing from it in various respects (such as far less /ɚ/ suffixation, different phonotactic constraints, as well as lexical differences) (Norman, 1988; Chao, 1961). This variety of Mandarin is known as *guoyu* in Taiwan and *putonghua* in mainland China, and as Mandarin, Standard Mandarin, or Standard Chinese in English. With only minor alterations, it is still the official language of the Republic of China and the PRC, and its use is promoted in schools and in the national media (Norman, 1988). In concert with spoken language reform, the written standard language was modified to closely resemble the standard spoken language, and *wenyan* was largely abandoned (Norman, 1988).

Despite the official status of Standard Mandarin, only around 53% of the population of China are capable of speaking it in any capacity, and of these only 20%, or around 130 million people, are considered 'fluent' (Duanmu, 2007). Figures for the number of speakers of 'Mandarin' in the 800 million range are frequently interpreted as relating to a particular Mandarin language (presumably the standard), when in fact they refer to speakers of all of the languages/dialects in the Northern/Mandarin branch, which vary considerably in terms of inter-intelligibility. Often the Northern/Mandarin varieties are at least partially mutually intelligible with each other, though not always (Tang and van Heuven, 2009). One striking example of this variability is in tonal inventory, which varies from as few as three to as many as six tones in the Mandarin languages/dialects (Norman, 1988), with Standard Mandarin having four.

It is relatively common to refer to all varieties of Sinitic as 'Chinese dialects'. There are cultural and political reasons for this, though it is perhaps more important that all varieties of Sinitic use essentially the same writing system (some populations, such as those in Taiwan

and Hong Kong, use different versions of certain characters, i.e. 'traditional' characters), in which each symbol represents a 'word', but contains unreliable phonetic information. Speakers of different Chinese languages who would be incapable of communicating using spoken language are perfectly capable of written communication. It would certainly be possible to write English using the Chinese script that would be intelligible to a Chinese speaker, as the characters convey meaning and not sound. This is currently the case for written Japanese, which in general outline uses Chinese characters for content words, and appends phonetic symbols to these characters to indicate morphology.

It should be remembered that what are frequently referred to as dialects in Chinese can be as divergent phonologically, lexically, and in grammar as French from Spanish or Italian, Dutch from English or German, or Spanish from Rumanian (Chao, 1968; Norman, 1988). It is of critical importance in performing research on any of these languages/dialects to keep this in mind as, for example, the inclusion of French and Spanish speakers in an experiment intended to examine Spanish, is in jeopardy of drawing erroneous conclusions. This research will centre on a variety of Chinese - a particular dialect of the Mandarin language, in the Mandarin family of the Mandarin branch of Sinitic, generally known as Standard Mandarin.

2 Phonetics and Phonology of Mandarin

2.1 Consonants

Mandarin has 25 consonant phonemes. Table 2.1 provides an illustration of the phonological consonant inventory. It should be noted that there is some disagreement as to the status of the three palatal obstruents - Duanmu (2007), for example, treats them as a combination of dental plus [j], and does not include them in his phoneme inventory.

	Labial	Dental	Retroflex	Palatal	Velar
Stop	p p^h	t t^h			k k^h
Affricate		ts ts^h	tʂ $tʂ^h$	tɕ $tɕ^h$	
Fricative	f	s	ʂ ʐ	ɕ	x
Nasal	m	n			ŋ
Liquid		l			
Glide	(w)			(j ɥ)	(w)

Table 2.1. The consonants of Mandarin, based on Wan and Jaeger (1998) and Duanmu (2007).

Mandarin makes little use of phonetic voicing in consonantal contrasts, with aspiration being contrastive for stops, for all places of articulation. It is worth pointing out that a relatively large proportion of syllables have voiceless onsets because of this. For a tonal language this is significant, as pitch may only be phonetically implemented on the voiced portion of a syllable (Chao, 1968a). For most syllables with a consonant in the onset, tone implementation is restricted to the rhyme.

2.2 Vowels

Mandarin is generally described as having five vowel phonemes (Wan and Jaeger, 1998, 2003; Norman, 1988; Duanmu, 2007). The vowel inventory given in Table 2.2 contains six vowels following Duanmu (2007), including a retroflex central vowel, though the status of [ɚ] is complex. The 'retroflex vowel' has limited distribution, occurring only as an independent syllable or as a coda suffix. The vowel [ɚ] could also be transcribed as [ɻ], in which case [ɻ] would need to be added to the consonant inventory - either replacing [ʐ], or as a variant, with [ʐ] in the onset and nucleus and [ɻ] in the coda (Duanmu 2005, 2007).

	Front	Central	Back
Close	i y		u
Mid		ə (ɚ)	
Open		a	

Table 2.2. The vowel phonemes of Standard Chinese, based on Wan and Jaeger (1998) and Duanmu (2007).

The mid and open vowels have a number of variant realisations. The mid vowel can vary in backness and in rounding due to the environment (e.g. [o] in open syllables after labials; [e] before [i]; and [ɛ] in open syllables after palatals). The open vowel varies only in backness (e.g. [ɑ] in closed syllables before [u] or [ŋ]; [æ] in closed syllables before [n] and after [j]; and [a] in closed syllables, before [n] or [i] and not after a palatal). Two of the high vowels, [i] and [u], can occur as the second vowel of a diphthong by combining with the mid and open vowel, i.e. [əi], [əu], [ai], and [au] (Duanmu, 2007).

In addition to the six vowel phonemes given above, Mandarin has five consonants which can function as 'syllabic consonants' (Duanmu, 2007). These include the three nasals [m, n, ŋ], which tend to occur almost exclusively as interjections, e.g. [m] ('yes') and [hŋ] (a grunt of contempt). The other two, [ʐ] and the allophone [z] (sometimes called 'apical vowels') are more common, e.g. [ʐʐ] 'sun' and [sz] 'silk' (Duanmu, 2007).

2.3. Phonotactics and Syllable Structure

All consonants except [ŋ] can occur in onset position, while only [n, ŋ] can occur in coda position. The maximal syllable is CGVX, where C is a consonant, G a glide, V a vowel, and X a long vowel, the second vowel in a diphthong, [ɚ], or one of the two allowed nasals.

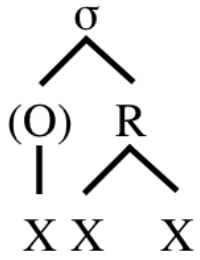


Figure 2.2. Maximal syllable structure in Mandarin (after Duanmu, 2007). The onset is maximally composed of a consonant and a glide, while the rhyme slots are maximally filled by a long vowel, the two elements of a diphthong, vowel + [ɚ], or vowel + coda nasal.

There are 63 possible CG combinations in Standard Mandarin - 25 total consonants, minus the velar nasal, minus the three glides, leaves 21 consonants which could theoretically combine with each of the three glides. Of these 63 potentially possible combinations, only 26 occur in Standard Mandarin. Duanmu (2007) counts the three palatal obstruents as dental obstruent/glide combinations, giving 22 total consonants (minus the velar nasal and the three glides), combining with the three glides for a total of 54 possible combinations. Of these, he counts 29 possible (the 26 I have given, plus the three palatal obstruents, which he considers CG combinations in themselves). The CG combinations of Standard Mandarin are given in Table 2.3.

	Cʲ	**Cʷ**	**Cᵚ**
p	pʲ	-	-
pʰ	pʰʲ	-	-
m	mʲ	-	-
f	-	-	-
t	tʲ	tʷ	-
tʰ	tʰʲ	tʰʷ	-
n	nʲ	nʷ	nᵚ
l	lʲ	lʷ	lᵚ
ts	-	tsʷ	-
tsʰ	-	tsʰʷ	-
s	-	sʷ	-
tʂ	-	tʂʷ	-
tʂʰ	-	tʂʰʷ	-
ʂ	-	ʂʷ	-
ʐ	-	ʐʷ	-
tɕ	-	tɕʷ	-
tɕʰ	-	tɕʰʷ	-
ɕ	-	ɕʷ	-
k	-	kʷ	-
kʰ	-	kʰʷ	-
x	-	xʷ	-

Table 2.3. CG combinations in Standard Mandarin. Based on Wan and Jaeger (1998) and Duanmu (2007).

The palatal series of consonants is in complementary distribution with the dental sibilants, the retroflexes, and the velars, and is generally considered to be an allophonic variation from one of these series before high front vowels. Arguments have been made for all three series (Norman, 1988), though historically they derive from the velars and dentals (Chao, 1968a).

There are approximately 400 unique syllables in Standard Mandarin, when not including tonal distinctions. Of the syllable types, C(G)V is the most common in a large corpus of Chinese (Da, 2007) at around 55% of the total, C(G)VN 42%, V 2%, and VN 1%. Details of the distribution of syllables by type are given in Table 2.4.

Syllable Type	No. of Syllables	No. of Characters	Frequency	Characters per Syllable	Frequency per Syllable	Frequency per Character
C(G)V	219	1,967	624,445.11	8.98	2,851.35	317.46
C(G)VN	166	1,296	316,841.82	7.81	1,908.69	244.48
V	9	165	49,608.18	18.33	5,512.02	300.66
VN	5	55	9,490.14	11	1,898.03	172.55

Table 2.4. Distribution of Standard Mandarin syllables by type and frequency (occurrences per million characters).

The two most common syllable types differ radically in frequency, with C(G)V *nearly twice as frequent* in the corpus. There are more C(G)V characters, on average just over one character more per syllable - but each of these characters is also on average 73 occurrences per million more frequent.

Second, though there are only 9 syllables of the V type, each of these syllables has on average over 18 characters - by far the most. These characters are also, on average, nearly as high in frequency as the C(G)V characters. In contrast, the other limited syllable type, VN, has 11 characters per syllable, but these are on average the least frequent characters in the corpus. Open syllables are generally more frequent in Mandarin than closed syllables.

The V syllable type contains the second most frequent character in the corpus: 一 (*yī*, 'one'), at 15,844 occurrences per million. This character alone accounts for roughly a third of the frequency of the entire syllable type. The syllable *yi* also has the second highest number of associated characters of any syllable in the corpus at 49, almost 30% of all V type

characters. The V syllable type also includes *yu*, which has the fourth most characters at 42. These two syllables account for 55% of the characters of the entire syllable type.

There are approximately 400 unique syllables in Mandarin, disregarding tone. Given the four lexical tones, we might expect to find around 1,600 unique syllable + tone combinations, but due to a number of gaps there are only around 1,255 (Duanmu, 2007). As all but a handful of morphemes in Mandarin are monosyllabic, this results in considerable homophony (more on this in ***5 Word Formation***).

By way of comparison English, for example, has closer to 9,000 unique syllables (though not all of these form words) (Duanmu, 2007), and a large number of polysyllabic morphemes - both of which contribute to an enormously greater inventory of unique combinations for forming morphemes. Mandarin also contrasts with other languages with small syllable inventories. Hawaiian has an even more restricted syllable inventory than Mandarin. With only 8 consonants and 5 vowels (plus long vowels and diphthongs), and a strict (C)V(V) syllable structure, there are only around 200 total syllables (Elbert and Pukui, 1979). Unlike Mandarin, the vast majority of morphemes are two syllables (Krupa, 1971), avoiding the pressure for extensive homophony.

3 Tone

Mandarin has four lexical tones, as well as a 'neutral tone' (or weakly stressed syllables) (Wan and Jaeger, 1998; Duanmu, 2007; Chao, 1968b). Syllables with the neutral tone cannot be produced in isolation, as the content of the tone is determined by the tone of the preceding syllable. In Standard Mandarin (and northern varieties generally), neutral tone or weakly stressed syllables may occur predictably, as in second syllable reduplication (妈妈 *māma*, 'mother') and grammatical suffixes (你们 *nǐmen*, 'you (pl.)'), in particles (了 *le*, completion particle), as well as in lexically specified positions (e.g. 天气 *tiānqi*, 'weather' in Beijing; *tiānqī* in Yangzhou (Li, 2006)). In the Mandarin varieties spoken in southern China, however,

the neutral tone is more restricted, and occurs primarily on grammatical suffixes and particles (Chao, 1968a; Norman, 1988).

The four tones are represented by diacritics in *pinyin* (i.e. *mā, má, mǎ, mà*), though in English they are also referred to by number (T1, T2, T3, T4) or trajectory on the Chao scale, where '5' is high and '1' is low in a speaker's range. The four tones of Standard Mandarin in the Chao scale are: T1 (55), T2 (35), T3 (213), and T4 (51). The neutral tone is unmarked in *pinyin*, but can be marked as T5.

A sample sentence in Mandarin, exhibiting the four lexical tones, as well as the neutral tone, is given below. Note that ordinarily characters are spaced uniformly, but are presented here so that the character-morpheme-syllable correspondences are clear.

Character	妈	妈	买	了	两	幅	油	画
Pinyin	*mā*	*ma*	*mǎi*	*le*	*liǎng*	*fú*	*yóu*	*huà*
Translation	mother	mother	buy	(perfective)	two	(measure word for paintings)	oil	painting
Gloss	Mother has bought two oil paintings.							

Typical pitch contours for the four lexical tones (spoken in isolation by a single male native speaker) are demonstrated in Figure 2.3.

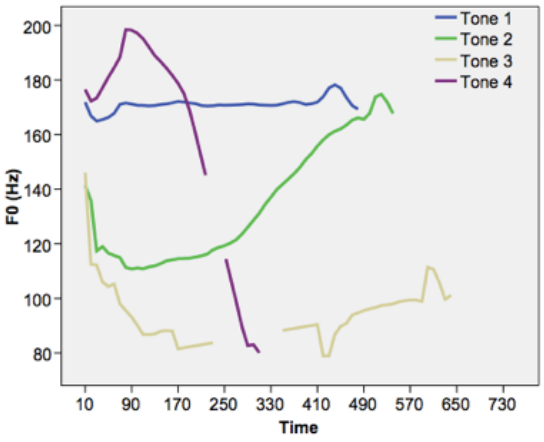


Figure 2.3. The four tones of Mandarin, produced by a native male speaker (from Lai and Zhang, 2008).

Pitch height and change are the primary correlates of tone in Mandarin, though there are systematic duration and amplitude-contour cues as well (Whalen and Xu, 1992; Ye and Connine, 1999; Chen and Massaro, 2008; Kuo et al, 2008; Liu and Samuel, 2004; Norman, 1988).

Both duration and amplitude have been found to provide sufficient cues for tonal identity in the absence of pitch. Liu and Samuel (2004) presented monosyllables with pitch excised artificially, as well as naturally produced whisper. They found that 40% of responses correctly identified tone in naturally produced whisper, compared to 36% in artificial whisper (with chance at 25%), and attributed the slightly more accurate responses to exaggerated durations in naturally produced whisper (i.e. trading relations).

Amplitude contour varies systematically by tone (see Figure 2.4). When neutralising both pitch and duration cues, Whalen and Xu (1992) found tonal identification to be possible by amplitude contour alone. There is redundancy in cues to tonal identity, in that a number of factors work together to cue tonal contrasts.

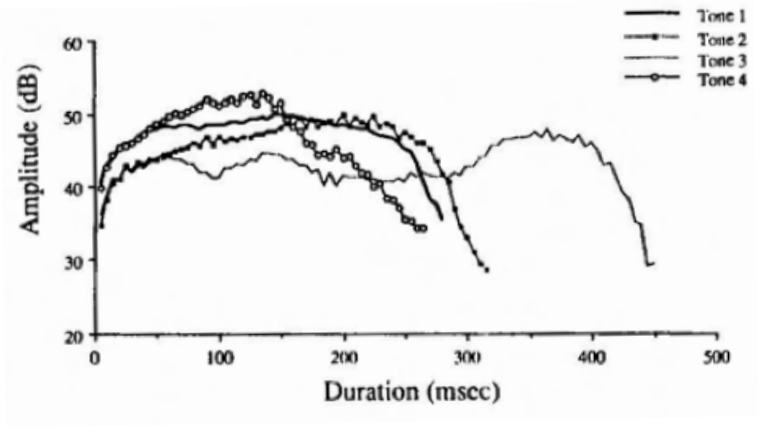


Figure 2.4. Amplitude contours for the syllable /ba/ (from Whalen and Xu, 1992).

The four lexical tones are briefly described below. Information about pitch height and change comes primarily from Chao (1968a). Duration is taken from Chen and Massaro (2008), and amplitude from Whalen and Xu (1992).

Tone 1 - the 'level tone'

Tone 1 begins towards the upper range of an individual's pitch register, and stays approximately level through the end of the syllable. Tone 1 is the third longest in mean duration, and has a stable (flat) amplitude in the upper range.

Tone 2 - the 'rising tone'

Tone 2 starts in the middle of a speaker's register, then rises through the end of the syllable, and is the second longest in duration. Tone 2 has an increasing amplitude through its duration, rising at its peak to approximately the level of tone 1.

Tone 3 - the 'falling-rising tone'

Tone 3 is similar in contour to tone 2 - though this is highly dependent on context. In isolation or phrase-final position, it begins in roughly the same range as tone 2 and falls, before rising again. In most other contexts, the rising portion of the tone is omitted. Creakiness is often associated with tone 3, and it is the longest in duration. Tone 3 has a stable (flat) amplitude in the lower range through most of its duration, and is marked with a rise in amplitude in its final third.

Tone 4 - the 'falling tone'

Tone 4 starts high in an individual's pitch register, and falls rapidly, and is the shortest in duration. Tone 4 has the highest peak amplitude, followed by a rapid amplitude decrease.

In terms of distribution across syllables, the four tones are relatively evenly spread. Of the approximately 1,255 syllables (including tone), 337 are T1, 255 are T2, 316 are T3, and 347 are T4. Of the 402 unique syllables (disregarding tone), only 178 carry all four lexical tones. The majority of Mandarin syllables do not have a complete tonal paradigm - 130 syllables

have some combination of three tones, 59 have just two tones, and 35 syllables are associated with only one tone (Duanmu, 2007).

Tone in Mandarin is often described as critically distinguishing words. A common example given uses the syllable *ma*: *mā* means 'mother', *má* 'hemp', *mǎ* 'horse', and *mà* 'scold'. This is potentially misleading - the syllable *mǎ* certainly can mean 'horse', but it can also mean 'ant', 'agate', 'stack', and 'mammoth', for example, while the syllable *mà* can mean 'scold', 'headboard', 'mark', and 'a sacrifice at the beginning of a military campaign'. A more complete listing of the possible meanings associated with *ma* is given in Table 2.5.

Character	Pinyin	Gloss
嘛	ma	(a modal particle)
吗	ma	(question tag)
妈	mā	ma/mamma
嬤	mā	ma/mamma
麻	má	(to have) pins and needles/tingling/hemp/numb/to bother
蟆	má	toad
蚂	mǎ	ant
马	mǎ	horse/horse chess piece/Surname
玛	mǎ	agate/cornelian
码	mǎ	a weight/number/yard/pile/stack
犸	mǎ	mammoth
骂	mà	scold/abuse
杩	mà	headboard
祃	mà	a sacrifice at the beginning of a military campaign
罵	mà	scold
唛	mà	mark

Table 2.5. Abridged listing of *ma* morphemes compiled from Da's (2007).

As can be seen in the table, the specification of tone does not necessarily uniquely identify a 'word', or a morpheme/character. This feature - the ambiguity of a syllable in isolation - is a key challenge for the processing of homophones in Mandarin, and is discussed more fully in **Chapters III** and **IV**.

4 Grammar - Morphology, Syntax, and Grammatical Particles

Mandarin is an SVO language, with more than 90% of sentences in both written and spoken language following this template (Chen, 1995). Mandarin has no inflectional morphemes (Chen, 1995), and a small number of affixes. According to Zhang and Peng (1992) there are only ten - four prefixes (the ordinal prefixes 初 *chū* and 第 *dì*, a prefix for names 阿 *ā,* and 老 *lǎo*) and six suffixes (the nominal suffixes 子 *zǐ*, 儿 *ér*, 头 *tóu,* 巴 *bā*, and 者 *zhě*, and a suffix to adjectives/adverbs indicating 'the state of', 然 *rán*). Words containing an affix make up less than 4% of all words in Mandarin (Zhang and Peng, 1992).

Classifiers are frequent in Mandarin, where they are often known as 'measure words'. There are some 900 classifiers in Mandarin, which fall into a number of categories (such as nominal classifiers for count and mass nouns, verb classifiers, and others) (Zhang, 2007). Numeral classifiers are obligatory before a noun (Zhang, 2007), and like other Mandarin classifiers can provide a certain amount of disambiguation for an ambiguous word. For example:

Character	这	是	一	座	水	塔
Pinyin	*zhè*	*shì*	*yī*	*zuò*	*shuǐ*	*tǎ*
Translation	this	is	one	MW (buildings)	water	tower
Gloss	This is a water tower.					

Character	这	是	一	只	水	獭
Pinyin	*zhè*	*shì*	*yī*	*zhī*	*shuǐ*	*tǎ*
Translation	this	is	one	MW (animals)	water	otter
Gloss	This is an otter.					

The two sentences are phonetically identical except for the classifiers - the two words ‘tower’ and ‘otter’ are homophones in Mandarin. The homophones are distinguishable in this example by the classifiers with which they occur - 座 (*zuò*) is used only with buildings, while 只 (*zhī*) is used with (most) animals. The situation is similar to the use of the word ‘pair’ or ‘cup’ in the English phrases ‘a pair of trousers’ and ‘a cup of tea’, though the use in Mandarin is obligatory for all nouns.

5 Word Formation

5.1 Characters

The most widely familiar unit in general use by Chinese speakers, and still in frequent use in Chinese linguistics, is the character - not, as is common in European languages such as English, the word (Chao, 1968a). Characters may only represent one syllable, and except for the rare instances where a morpheme is polysyllabic (and therefore requires more than one character to write), also represent a single morpheme (Chao, 1968a; Norman, 1988). It is in this sense that Mandarin is called a monosyllabic language - though there are many compound words composed of several characters/syllables/morphemes, morphemes themselves are almost always one syllable.

Words, as a concept, are more problematic for Mandarin linguistics (Hoosain, 1992; Duanmu, 2007; Chao, 1968a). The concept of the word in descriptions and analysis of Chinese is relatively recent, entering usage in the early 20th century (Duanmu, 2007). Words are not marked in writing, as they are in English. Each morpheme is represented by a character, and characters are evenly spaced. Punctuation was also absent until the early 20th century language reforms, when a punctuation system based on European usage was adopted (Hu, 2007). Word boundaries are not reliably signalled in Mandarin speech, through either stress or tone, as they are sometimes in English (i.e. ‘green house’ versus ‘greenhouse’) (Duanmu, 2007). Though there are tone sandhi processes that occur only within word

boundaries in certain Chinese languages, such as the Wu dialects, the equivalent processes tend to apply both within and across word boundaries in Mandarin (Chao, 1968a). One exception is when three T3 syllables occur in a row. If none of the syllables combine to form a word, the sandhi occurs on the first and second syllables (T3-T3-T3 > T2-T3-T3 > T2-T2-T3). If the second and third syllables constitute a word, then the sandhi occurs only on the second syllable (T3 - [T3-T3] > T3-[T2-T3]) (Sproat et al, 1996).

Judgements concerning word delineation are also variable between speakers. In a study on written Cantonese conducted by Hoosain (1992), in which participants were asked to segment sentences into constituent words, the responses were found to be in substantial disagreement. This was attributed in the study to a number of potential factors - a relative unfamiliarity in talking about 'words' generally for Cantonese speakers, such talk centring on morphemes/characters, as it does for Mandarin speakers as well (Chao, 1968a); the influence of an awareness in most of the literate population of written Classical Chinese, in which most morphemes could occur independently; and the fact that a significant proportion of the population speak more than one Chinese language, languages differing in which morphemes can occur independently (for example, in Cantonese 桃 'peach' is a free word, while in Mandarin it is not, and must occur as 桃子 (peach+noun suffix) meaning just 'peach'). Similar word segmentation variability has also been found in Mandarin. Sproat et al (1996) found that Mandarin subjects were in agreement about word segmentation around 75% of the time, which they noted was surprisingly low. The apparent instability of the word in Mandarin, in conjunction with the strong Chinese bias towards the character/morpheme as the basic lexical item, will have implications for language processing and homophony. These concerns are addressed in **Chapters III** and **IV.**

5.2 Polymorphemic Words

Though nearly all morphemes in Mandarin are monosyllabic, there are many words composed of two or more morphemes. There is a relatively small class of affixes, which combine with free morphemes to form words. Some of these, such as the plural suffix 们 -*men*, are minimally productive. In general, only pronouns and 'human nouns' can take the plural suffix (Li, 1999), e.g. 我们 *wǒ men*, 'I'+'plural marker', meaning 'we'. Most nouns are not marked for number, and number is inferred from cues in the phrase or from context, e.g. 三盏灯 *sān zhǎn dēng* 'three MW lamp', meaning 'three lamps'. Other suffixes are very productive, such as 子 *-zi,* which Norman (1988) calls a noun 'word-formative' suffix. This suffix attaches to a noun, to form a noun, e.g. 桌子 *zhuō-zi* 'table+noun suffix', meaning just 'table'. The suffix itself is devoid of meaning, and its only purpose appears to be to make a monosyllabic noun into a disyllabic noun.

The most common type of polymorphemic word is, however, the compound (Chao, 1968a). Mandarin compounds consist of at least two non-affix morphemes, with two being the most common, though three (and sometimes more) morpheme compounds also exist (Norman, 1988). There are many possible combinatory structures (i.e. noun+noun; adjective+adjective; adjective+noun; verb+object; verb+adjective complement, etc.). The constituent morphemes of compound words in Mandarin often tend to retain some part of their independent meanings, which contribute to the meaning of the compound in a (variably) transparent way (Hoosain, 1992). Two examples of this can be found in one word, the humorously straightforward name of the comic character Popeye: 大力水手 *dà lì shuǐ shǒu* 'big strength sailor', where the word 'sailor' itself is composed of two morphemes (water-hand). There is debate as to how such compounds are processed in Mandarin - for example, whether *shuǐ shǒu* 'sailor' is treated as a single unit, or as a series of independent pieces. This topic is discussed in **Chapter IV**.

Compounding is quite common in modern Mandarin, with some sources giving figures as high as 80% of the lexicon (Peng et al, 1999; Taft et al, 1999), and it is the most common method for creating new words (Norman, 1988). This is in apparent contrast to Classical Chinese, which is very nearly devoid of polymorphemic words (Norman, 1988). There is some debate on whether Classical Chinese texts accurately reflect the spoken language of the period, however. The apparent monosyllabic nature of the classical language may at least in part be due to the fact that it was originally carved on bone or shell, and that every attempt was made to reduce labour, in a sort of parallel to the terseness of telegrams (Duanmu, 2007). It is also possible that compounding is related to phonological changes that have occurred in the intervening centuries. Modern Mandarin has many fewer syllables than the classical language, due to historical processes such as the loss of all non-nasal coda consonants, consonant clusters, as well as a number of tonal distinctions (Norman, 1988), one effect of which is a large increase in homophony. It is possible that the subsequent increase in compounding has served as an aid to clarity as other sources of syllable disambiguation were lost.

One notable type of compound consists of two semantically near morphemes, which combine to form a 'redundant' compound word with the meaning of one of the constituents (i.e. 豌豆 *wān dòu*; 'pea'+'bean', meaning 'pea') (Duanmu, 2007; Hoosain, 1991). Somewhat like the set of nouns which occur with the noun suffix *-zi*, this results in a class of disyllabic words where there is no obvious semantic need.

'Redundant' compounds, and compounds generally, may be a response to the increase in homophony - *wān* may mean 'pea' already, but it also *could* mean 'bay', 'crooked', or 'cut', among other things. Though *wān dòu* may very well still have homophones, these are certainly fewer than *wān* on its own. This difficulty was less prevalent in Classical Chinese, as any given syllable/morpheme would have had fewer homophones, and it is no difficulty at all in writing, where each morpheme is represented by a unique character.

The process of disyllabification in some way mimics a common occurrence in spoken conversations. A word which may be ambiguous in a certain utterance is repeated, either in a different context, or in a compound, to specify which meaning of that particular syllable was intended. An example of this, where the potentially ambiguous word is repeated in a longer (less ambiguous) stretch of speech to clarify the intended meaning of a monosyllable, is given below.

Character	我 姓 卫，保 卫 的卫。
Pinyin	wǒ xìng wèi, bǎo wèi de wèi.
Translation	I surname Wei, protection *poss. Wei.
Gloss	My surname is Wei, 'Wei' as in the word 'protection'.

In this case, *wèi* (this individual's surname) is ambiguous. There is another *wèi* surname in use in modern China, 魏. This individual's actual name, 卫, is less common, and so she has specified that her particular *wei4* name is the same as the character occurring in the word 'protection'. The situation is analogous to someone stating that their given name was [ɛlənə], then specifying that it was spelt 'Elinor' and not 'Eleanor'.

5.3 *Number of Compounds*

Compounds are sometimes reported as making up more than 80% of Mandarin words (Peng et al, 1999; Taft et al, 1999), however the reported percentages vary. For example, The Modern Chinese Frequency Dictionary (Institute of Language Teaching and Research, 1986) reports that 73.6% of word *types* are disyllabic compounds, while 12% are monosyllabic. When *tokens* of the corpus were counted, 64.3% were found to be monosyllabic and 34.3% disyllabic.

The precise ratio of monosyllables to disyllables does vary considerably depending on the source material, with monosyllabic words comprising 50% of the tokens in newspaper articles, and disyllabic words 48.5% (Institute of Language Teaching and Research, 1986),

indicating that the monosyllabic/disyllabic ratio may be dependent upon register. It also means that the sources used in compiling such frequency counts are extremely important if an accurate picture of Mandarin usage is to be drawn. Unfortunately, there are systematic biases in the most commonly cited frequency dictionaries, such as *The Modern Chinese Frequency Dictionary*, which draws from a number of potentially unsuitable sources. For example, the corpus includes political speeches by non-native Mandarin speakers, such as several given by Mao Zedong (a native speaker of Xiang who was far from proficient in Mandarin; see Chang and Halliday, 2005) more than a decade before modern 'Standard Chinese' was created (see Institute of Language Teaching and Research (1986) for a full listing of corpus sources).

Though compounding is unquestionably highly prevalent in Mandarin, considerable care is required in compiling accurate statistics for a number of reasons. First, it has been noted that the number of disyllabic words fluctuates significantly depending on the type of writing. Second, as has already been discussed, word segmentation is frequently controversial in Mandarin, and is a potential source of inconsistency across studies - if there is not substantial agreement in word segmentation, the counts are unreliable. Finally, and perhaps most importantly, the numbers reported are often given as applying to *Chinese*. As briefly discussed, the so-called 'dialects' of Chinese vary in a large number of ways, not least in terms of what morphemes may occur as monosyllabic words, and those which must join into compounds.

This makes it vitally important that the materials used to compile frequencies are properly balanced, that word segmentation is carried out in a consistent and theoretically justifiable fashion, and that all texts used were composed by native speakers of the particular language for which the calculations are intended. This last point may prove difficult to implement, as the meanings of characters are accessible to speakers of all Sinitic languages. It also makes it important to cite frequencies as applying to the 'dialect' for which they were collected.

6 Orthography

Mandarin is written with two orthographic systems. The first, and by far the most prevalent, is the system of Chinese characters. The second is the romanisation system known as *pinyin*, which is chiefly used in teaching the pronunciation of characters. These two systems will now be briefly discussed.

6.1 Characters

Chinese characters have been in use for at least 3,000 years, and though characters do change over time, early forms are largely consistent with modern characters (Duanmu, 2007). Characters are a form of morphemic writing, each character representing one syllable. With a very small number of exceptions, every morpheme is monosyllabic (Duanmu, 2007), and polysyllabic morphemes must be written with multiple characters. Characters are always written with uniform spacing (Hoosain, 1992), i.e. two characters in a compound word will be the same distance apart as two monomorphemic words. Word boundaries are therefore not represented in the writing system, though various European punctuation is now in regular use, such as commas, question marks, and full stops (Hu, 2007).

In Classical Chinese most characters could occur independently, unlike in modern Mandarin with its many compounds words, leading to the long-held assumption that Classical Chinese largely consisted of monosyllabic words (Chao, 1968a). This assumption has recently been called into question, and it is likely that the apparent terseness of Classical Chinese is, at least in part, due to the amount of effort required to carve characters into shell and metal, the principal medium for writing at that time (Duanmu, 2007).

A large number of characters have more than one meaning. Frequently the multiple meanings of a character will appear related to each other in some (usually semantic) way, i.e. 婉 (*wǎn*), which means either ‘graceful’ or ‘tactful’. These two meanings are distinguished in compounds, for example: 婉约 (*wǎn yuē*; meaning ‘*wǎn* restrain’, i.e. ‘graceful’; and 委婉

(*wěi wǎn;* meaning 'devious *wǎn*', i.e. 'tactful'). Opinion varies as to when two meanings such as these should be considered as two separate words written with the same character, and when they should be considered as two senses of the same word. In a certain sense, characters such as 婉 have no meaning outside of compounds in modern Mandarin, or they have at best a kind of vague, ambiguous meaning.

This point is further illustrated in the borrowing of polysyllabic words from other languages, which can have interesting consequences. The standard transcription of the name 'Christopher' is 克里斯多夫 (*kè lǐ sī duō fū*; meaning 'gram inside this much man'). But these characters were only used because they matched the sounds of the intended syllables, and (clearly) not as a semantic translation.

There are well over 50,000 characters, when archaic and rare characters are considered (Liu et al, 1989), though only around 6,000 characters are in general use (Da, 2007; Norman, 1988). Estimates of how many characters are required to be an accomplished reader of Mandarin are generally around 4,000 (Norman, 1988; Zhang and Peng, 1992), though the number of known characters increases with education.

Given the restriction in Mandarin that virtually all morphemes are monosyllabic and the number of unique syllables in Mandarin, if the 6,000 characters cited for common modern usage were divided evenly across all syllables there would be 15 characters (i.e. morphemes) per syllable, disregarding tone, and about 5 characters for every syllable+tone combination. The actual distribution of characters is, however, not remotely this even. For example, there are 81 characters which are pronounced *yi*, 51 of which are pronounced with the same tone, *yì*, and the most frequent 15 syllables (disregarding tone) account for more than 1,000 characters (Da, 2007; Duanmu 2007). The uneven spread of homophony across syllables will be a critical component of this study. Whether the number of homophones associated with a particular word affects lexical access and word recognition is investigated in **Chapters V**, **VI**, and **VII**.

6.2 Radicals

There are two types of characters. Simple characters are those which are composed of only one element, called a *radical*, for example 口 (*kǒu*; 'mouth'). The other type are 'compound' characters, which are composed of more than one radical. Compound characters are composed of between one and several radicals. The addition, subtraction, and placement of radicals is capable of signalling differences in meaning by forming new characters. Radicals may be simple characters in themselves, and thus capable of standing alone, or they may be restricted to occurring only as a component in a compound character.

There are two broad classes of compound characters - 'semantic' compounds and 'phonetic' compounds, both of which are composed of radicals. Semantic compounds are generally composed of two radicals, both of which may contribute in some way to the meaning of the whole. Phonetic compounds are also generally composed of two radicals, one of which may signal the meaning of the character and which typically occurs on the left, and the other potentially signalling something about the pronunciation, and which is typically on the right (Feldman and Siok, 1999). These two functions of radicals are sometimes referred to as *semantic radicals* and *phonetic radicals* (or *radicals* and *phonetics*) respectively, though it should be noted that some radicals can function as both phonetic and semantic indicators. Examples of these types of compound-character are given in Figure 2.5, with 口 (*kǒu*; 'mouth') serving as a phonetic radical in one, and as a semantic radical in the other.

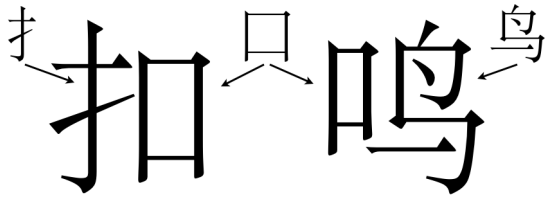


Figure 2.5. Illustration of a phonetic compound (*kòu*, left), and a semantic compound (*míng*, right).

The character on the right is the semantic compound character 鸣 (*míng*; 'chirp'). It is composed of two radicals: on the left side, 口 (*kǒu*; 'mouth'), and on the right side is 鸟 (*niǎo*; 'bird'), both of which contribute to the meaning of the whole character (and neither of which contribute to the pronunciation). The character on the left side of Figure 2.5 is a phonetic compound, 扣 (*kòu*; 'detain'). It is also composed of two radicals: on the right side of the character is 口 (*kǒu*; 'mouth'); and on the left side is扌(*shǒu*; 'hand'). This is a phonetic compound because one of the radicals contributes to pronunciation (i.e. *kǒu*), differing only in tone, while the other contributes to the meaning of the character (i.e. 扌 which does not occur independent of compound characters).

Phonetic compounds make up the majority of characters in modern usage at approximately 85%, with simple characters and semantic compounds making up the rest (Perfetti and Tan, 1999; Shu and Anderson, 1999). It should be noted that the phonetic component of a phonetic compound character provides accurate pronunciation information only some 25-35% of the time (Shu and Anderson, 1999; Feldman and Siok, 1999). In cases where the phonetic radical does not actually contribute pronunciation information, it simply contributes to the unique 'look' of a character. Phonetic radicals are an occasionally useful, if somewhat unreliable, cue to pronunciation in Modern Mandarin. These phonetic indicators were more reliable in the past, and have been rendered ineffective by language change (and therefore changes in pronunciation of the characters) over a great deal of time (Norman, 1988).

6.3 Pinyin

Hanyu Pinyin ('Chinese Spelling System', usually referred to as simply *pinyin*) is the standard romanisation system used for transcribing Mandarin in the People's Republic of China. There are a number of other spelling systems for transcribing Sinitic languages, some of which are used to transcribe Mandarin outside of the PRC (e.g. the Mandarin Phonetic

system in Taiwan) (Chen and Yuen, 1991), and others which are used to transcribe other Sinitic languages which may have different vowel qualities and tones than Mandarin, and therefore different romanisation requirements (e.g. *Jyutping* for Cantonese in Hong Kong) (Wong and Law, 2008).

Pinyin marks lexical tone with diacritics which roughly mimic the contour of the tone: T1, which is relatively flat in pitch (fēn); T2 which is a rising pitch (fén); T3, which is falling then rising (fěn); and T4, which is a falling pitch (fèn). These symbols are simplifications of the tone marking system devised by Y.R. Chao (1930), and those used in the IPA (Duanmu, 2007). A full table of the symbols used in pinyin, along with IPA equivalents, is given in Table 2.6.

Consonants	**Pinyin**	b	p	m	f	d	t	n	l	g	k	h	j	q	x	zh	ch	sh	r	z	c	s	w	y
	IPA	p	p^h	m	f	t	t^h	n	l	k	k^h	x	tɕ	$tɕ^h$	ɕ	tʂ	$tʂ^h$	ʂ	ʐ	ts	ts^h	s	w	j
Vowels	**Pinyin**	i	u	ü	a	o	e	ai	ei	ao	ou	an	en	ang	eng	er								
	IPA	i	u	y	a	o	ə/ɛ	ai	ei	au	ou	an	ən	aŋ	əŋ	ɚ								

Table 2.6. Pinyin and IPA comparison table.

Pinyin is a relatively modern innovation, and stems from the language reforms of the early 20th century. It is taught to children to aid in learning the pronunciation of characters, and to foreign students for the same reason. It is also a common input method for typing on a computer (Tan et al, 2013) - pinyin is typed (without diacritics), and a list of possible characters is generated. The user then selects the appropriate character (see Figure 2.6). Though most Mandarin literate speakers in the PRC know pinyin, it is rare for adults to make much use of it outside certain restricted contexts (like teaching and typing).

Figure 2.6. An example of pinyin-based typing of characters on a computer.

7 Summary

Mandarin (and the Chinese languages generally) presents a number of features in terms of lexical organisation that are linguistically distinct. Cross-linguistically, the number of unique (segmentally-defined) syllables in Mandarin is low, at around 400 (compared to English, at around 9,000). This low syllable inventory is mitigated to some extent by the use of tone. Mandarin makes use of what is called 'lexical tone', where variations in pitch (primarily, but also in duration and amplitude contour) are used as phonological distinctions within a given syllable. With the four tonal distinctions used in Mandarin, there are still only around 1,255 total unique syllables, however. In addition, in contrast to other languages with very small syllable inventories (e.g. Hawaiian), polysyllabic morphemes are virtually prohibited, resulting in an extremely small number of morpheme-level distinctions. As a result, there is extensive homophony in Mandarin. Estimates based on dictionaries (e.g. Duanmu, 2007) give a figure of roughly four homophones *for every morpheme*. With this in mind, it is perhaps not surprising that users of the Chinese languages have persisted in a non-phonological script. Mastering the thousands of characters necessary to read and write a morpheme-based script like the Chinese languages use is laborious, but certainly helpful in avoiding ambiguity in

languages where extensive homophony is the norm. We may legitimately wonder, however - if such measures are necessary in writing, how is this ambiguity resolved in spoken language?

It has been noted that homophony is not equally distributed in Mandarin (Duanmu, 2007). Some variation in homophony across syllables is to be expected, of course, as language is not 'intelligently designed' (though Standard Mandarin has origins that are less standard than most). Mandarin, however, has a number of syllables with dozens of homophones, and scores of syllables with none, and this is potentially curious. To understand how Mandarin users navigate a language with such extensive homophony, it is first important to understand the exact nature of that homophony. And it is possible that some clues as to how speakers store and process homophones will be found in the way homophony is organised, both structurally (in terms of homophone distributions over syllables), and in terms of usage. This will be the subject of **Chapter III.**

Chapter III. Homophony in Mandarin

In **Chapter II** it was argued that the deeply homophonous nature of Mandarin results primarily from two interacting linguistic constraints - strict phonotactic constraints, which result in a comparatively small pool of possible syllables; and a strong morphological prohibition on polysyllabic morphemes. Thus, there are very few unique syllables in Mandarin at around 1,200 (compared to around 9,000 for English), and virtually no polysyllabic morphemes (Duanmu, 2007). This results in an unusually high degree of homophony when compared cross-linguistically, some aspects of which are acknowledged in the existing literature - for example, that the mean number of around four homophones per syllable/morpheme is fairly high, and that there are concentrations of homophony in some syllables (Duanmu, 2007). However, beyond these general broad-stroke observations, the distributional properties of Mandarin homophony, in terms of both structure and usage, have received little attention. If we conceive of the Mandarin lexicon as a kind of landscape, by examining these properties we are able to sketch out the 'topography' of this landscape, and assess the extent to which this shapes lexical processing and production. Thus, the purpose of this chapter is to map out the precise distributional nature of Mandarin homophony, as a basis for examining native speaker behaviour (word recognition and production) in subsequent chapters. In particular, we examine first of all the extent to which homophony is distributed – in both incidence and usage (density and frequency statistics) across the pool of possible syllable/morphemes, and the nature of any 'clustering'; and secondly, the specific role that tone has in this distribution and in any clustering, and the implications this may have for the question of whether tone is an intrinsic part of homophony or separately processed.

The chapter will begin with a discussion of the special considerations that must be taken into account when examining homophony in a tonal language such as Mandarin. These are both conceptual and terminological. For example, what sort of relationship should we

expect tonal minimal pairs (such as *pào* 泡 'bubble', and *páo* 咆 'roar') to have? Are they classic lexical neighbours, as they differ in just one property? Or are they more akin to 'true' homophones, as they are segmentally identical? And how shall we refer to them in any case?

Once the terminological and conceptual groundwork has been laid, we proceed to the main purpose of this chapter - mapping the contours of homophony in Mandarin. As the goal was to provide an analysis of the distribution of homophony for real (not idealised) speakers, the analysis in this chapter was carried out on a large written corpus (Da, 2007), modified to exclude extremely rare (i.e. low frequency) characters. This was done as an approximation of the expected lexicon of a modern-day literate Mandarin speaker. Most Mandarin dictionaries contain far more characters/words/morphemes than the corpus used here. Indeed, the corpus itself contains far more characters than even an educated speaker would reasonably be expected to have knowledge of. Homophone densities cited in the literature, both averages across the Mandarin lexicon and for particular syllables, are typically higher than those given here (e.g. Duanmu, 2007). A more detailed discussion of this is given later in this chapter.

As the bulk of this thesis explores homophony through the written medium, we also examine the relationship between words/morphemes/syllables and characters. Section 2 constructs a 'typical' Mandarin syllable based on mean values from the corpus for characteristics such as tone density and homophone density, and compares that to actual syllables (i.e. how representative is the 'average'?) across a number of metrics: 1) number of tones per unique segmental string (typical and actual tone density); 2) number of associated characters per unique syllable (and whether and how this is affected by tone density); 3) distribution of tones (are some tones more common than others, and why might this be?); 4) distribution of full/segmental homophone density (how many syllables show mean homophone density, and what homophone density do syllables typically have?).

1 Background

There is considerable variation in the application of terminology and approaches between researchers working on Chinese languages, particularly with regard to work in homophony (c.f. Li and Yip 1996, 1998 for variation in terminology across publications, even for the same authors). This variation springs largely from a number of theoretical factors, such as what precisely should count as a homophone, how tone is represented, as well as disagreement as to what constitutes a 'word' in Mandarin. In addition, there are several concepts developed for the first time in this thesis which require labels. Therefore, a supplementary aim of this chapter is to lay out an internally-consistent terminology, which will be used in exploring the extent and nature of homophony in this and further chapters.

1.1 Types of Mandarin Homophony

A source of one of the more pervasive terminological inconsistencies, and one of the primary research questions of the current study, is what precisely constitutes a homophone in Mandarin. Of particular concern is the potentially critical distinction between three scenarios:

1) two lexical items with 'identical' segmental and tonal composition,

   e.g. *pào* 炮 'cannon' and *pào* 泡 'bubble'

   wān 湾 'bay' and *wān* 豌 'pea'

2) two lexical items with 'identical' segmental composition and any tone (i.e. regardless of tone, but tone is specified),

   e.g. *wān* 豌 'pea' and *wǎn* 碗 'bowl'

3) two lexical items with 'identical' segmental composition, which artificially lack tone (i.e. regardless of tone, and tone is *not* specified),

e.g. *pao* 抛 'throw' and *pao* 跑 'run'

wan 万 '10,000' and *wan* 弯 'crooked'

Scenario (1) is what might conventionally be considered a homophone - two words with identical phonological form which have different meanings (Crystal, 2008; Gahl, 2008) - though the typical definition does not usually include tone, as it is normally used in the context of non-tonal languages. Lexemes in this relationship will be termed ***full homophones*** in this thesis. Questions about the nature of such homophony have been raised by recent empirical research (e.g Gahl, 2008) revealing phonetic *non*-identity between such pairs and thus, in frameworks proposing phonetically-rich representations, implications also for the phonological representation of homophony. If there are systematic differences in the production of purported homophones, and if phonological representations are phonetically rich, it could be argued that true homophony does not, theoretically, exist. The precise nature of our definition of homophony depends at least in part on how we conceive of the phonological representation. At this point we remain agnostic, and will return to a more substantive discussion of this issue at the end of this section.

Scenarios (2) and (3) are in effect sub-types of a single scenario, where tone is not essential to a judgment or definition of homophony. Scenario (2) is the strong version of this position and, unless we expressly allow mis-matching of information, is arguably not homophony at all. The two syllables 豌 (*wān*; 'pea') and 碗 (*wǎn*; 'bowl') more closely resemble phonological neighbours in the classic paradigm, as the two differ in only one phonological aspect, although it is not necessarily the case that segmental and tonal information are equally weighted for these purposes. A judgment of homophony in this scenario would require the listener to ignore information, and it is plausible that tonal

information is deemed sufficiently irrelevant to the judgment of homophony that it can be ignored without consequence. For present purposes, two syllables in a relationship such as demonstrated in scenario (2) (i.e. *wān* 'pea' and *wǎn* 'bowl') will be called ***segmental homophones,*** with the caveat that it is not necessarily the case that segmental homophones are processed as homophones by Mandarin speakers, though there is a precedent in the literature for labelling these as homophones (c.f. Li and Yip, 1996), and research (e.g. Cutler and Chen, 1997; Taft and Chen, 1992; Ye and Connine, 1999) indicating that this could indeed be the case, as will be discussed presently.

Scenario (3) is the weak version of the position that tone is not critical to homophony, and shares with scenario (1) the requirement that the two syllables in question are identical *as presented* – albeit artificially, in *pinyin* lacking tonal diacritics, or in speech modified or synthesised so as to be tonally ambiguous. This study, focused as it is on production and visual perception, does not directly address the contribution of varying acoustic indicators of tone in auditory perception, such as pitch, duration, and amplitude. It should be possible, however, to induce tonal ambiguity through acoustic manipulation, such that listeners are unable to determine whether a heard item is of tone two or three, for example. The *intended* morphemes associated with the two syllables may (or may not) have different tones, but since this information is suppressed it at least does not add conflicting information to the pair of stimuli being judged for homophony. Focusing on syllables with different *intended* tones, take the two syllables 豌 (*wān*; 'pea') and 碗 (*wǎn*; 'bowl') from the previous example. If these two syllables are presented devoid of any tonal information (e.g. *wan)* they *appear* identical in the paradigm given, at least when written in *pinyin*, but only because the stimuli are impoverished or underspecified. Syllables which are related to each other in this way will be termed ***apparent homophones,*** whether the intended associated lemmas match in tone or not.

1.2 Defining Homophony: the role of tone

As already mentioned, what are here called '***segmental homophones***' are not, at least conventionally speaking, homophonous insofar as a pair of syllables such as 豌 (*wān*; 'pea') and 碗 (*wǎn*; 'bowl') differ in one phonological aspect. We may plausibly question whether the (tonal) difference between *wan1* and *wan3* is equivalent (phonologically, phonetically, or psycho-linguistically) to the (segmental) difference between *wān* and *yān*. This is of critical importance since lexical tone is canonically considered to be part of the phonological form, although within non-linear frameworks it can at least be conceived of as *separate* from the segmental string (Goldsmith, 1999; Ye and Connine, 1999). For theories of abstract representation, a gradient judgment of homophony would then be explained as activating either both segmental and tonal tiers (full homophony) or just the segmental tier (near homophony). For theories of phonetically-rich representation (e.g. Pierrehumbert, 2000; Bybee, 2001) a gradient judgment would then have to be able to differentiate between exemplar clouds which acoustically match/over-lap in a holistic way (i.e. do not discriminate between the phonetic exponents of the hypothetical segmental and tonal tiers) and exemplar clouds which acoustically match/over-lap only for the phonetic exponent of the hypothetical segmental tier[1]. This potentially causes a problem for Exemplar Theory, as it is not immediately clear how these different kinds of phonetic information could be discriminated without some form of abstraction being involved. In theory we could posit a differentiation between F_0 and all other acoustic information; however, this is overly simplistic, since we know that the phonetic expression of tone is not limited to F_0. On the other hand, an exemplar approach is, more generally, very suited to a gradient evaluation of homophony that is *scalar* (i.e. not a series of different categories, but a continuum of greater or lesser homophony). Thus, the exemplar clouds associated with e.g. *wān*; 'pea' and 碗 *wǎn*; 'bowl' would overlap

[1] Hypothetical in the sense of that part of the phonetic exponent that corresponds to the output of the segmental tier in abstract frameworks.

to a significant degree, thereby triggering a judgment of likeness; likewise the exemplar clouds associated with syllables which are phonological neighbours, e.g. *wān* and *ān*, would also overlap to a significant degree, thereby also triggering a judgment of likeness. Any distinction between these would depend on the degree of likeness involved, which may be quantitatively different (i.e. the difference between two tones may be phonetically less marked than the difference between two segments), hence phonological neighbours may result as a different category from segmental homophones, through phonetic accident rather than phonological analysis.

The extent to which tone is essential to the judgment and definition of homophony thus has a number of important implications. It raises the question as to how exhaustive the phonological representation has to be in order for a judgment of homophony to be triggered, and goes to the heart of what we conceive of as the phonological representation. It also has important consequences for the *extent* of homophony. If tone is as important as segmental composition in terms of identity, it will greatly affect our calculation of both homophone density and neighbourhood density - homophone density will be *relatively* smaller, as homophone identity is more particularly constrained. Neighbourhood density will greatly increase - it is already relatively dense, and will now include all three tonal neighbours. A key question addressed in this thesis is whether frequency and density properties, known to influence lexical processing of non-homophones through phonological neighbourhood effects (e.g. Luce and Pisoni, 1998), also exert an influence in the processing of homophones, and if so whether these effects are comparable. Do the distributional properties of the 'segmental homophone neighbourhood' function in any way analogous to those of lexical neighbourhoods observed in other languages, such as English? Does the classic definition of lexical neighbourhood even make sense for Mandarin, where such neighbourhoods can constitute large swathes of the lexicon?

If we assume that tone is essential to the phonological form, and fully integrated, at some level, with the segmental string, then tonal identity is a pre-requisite for homophony. There is evidence, however, that tone is processed and organised differently than segmental material in Mandarin, and that syllables which differ only in tone are regarded by native speakers as closer to each other than syllables which differ segmentally. Ye and Connine (1999) investigated tone and vowel monitoring, for auditorily-presented monosyllables and syllables in four-syllable context. In their Experiment 1, subjects listened to monosyllables and monitored for the tone-vowel combination *á*. Ye and Connine (1999) found that monitoring decisions were made the fastest for syllables differing only in the vowel (*n.b.* stimuli mismatched for both vowel and tone were not tested). Syllables containing the correct vowel, regardless of tone, were responded to at the same speed - those containing the tone-vowel combination *á* were accepted, and those with the correct vowel but incorrect tone (e.g. *a4*) were rejected. These results indicate what Ye and Connine (1999) call a 'temporal asymmetry' in the processing of segmental and tonal information. Vowel information is available relatively quickly compared to tonal information, so syllables which do not contain the target vowel are rejected quickly. Syllables which contain the target vowel are presumably identified as doing so early in the process as well, but as the tonal information is relatively late in becoming available (Mandarin tones incorporate pitch contours, and have different Isolation Points, some later than others - see Lai and Zhang, 2008) there is a delay in deciding whether the tone matches or not. This accounts for the fact that full matches and tonal mismatches take approximately the same amount of time to accept and reject, respectively.

Ye and Connine (1999) performed an additional experiment, their Experiment 2, in which they examined the effect of context on the temporal asymmetry of tone and segmental information. Participants were asked to monitor for either the vowel 'a' or tone 2 (as opposed to the conjunction *á* in Experiment 1). Two levels of context were examined - a 'neutral' context, in which the target syllable was the last in a meaningless concatenation of four

syllables; and a highly biasing context of four-syllable lexicalised idiomatic expressions known as *chengyu*, in which the target was the final syllable. *Chengyu* are typically opaque idioms with a specific meaning, often allusions to stories (Li and Hin Tat, 2014). For example: 叶公好龙 '*yè gōng hào lóng*', literally meaning 'Lord Ye loves dragons.' The idiom is a reference to a story about Lord Ye, who is famous for his love of dragons. When he is actually visited by a dragon, however, he is terrified. The idiom 'Lord Ye loves dragons' has a set meaning, which is roughly 'shallow or insincere love of something that, if understood, would be feared or hated'. In order to provide maximal context, Ye and Connine (1999) only included *chengyu* that, if given the first three syllables, the fourth was completely predictable.

The authors found that both vowel and tone monitoring were faster for idioms than for neutral context. This was to be expected, as given the initial three syllables, the fourth was entirely predictable, and a certain amount of pre-activation indicated. For neutral context, vowels were monitored faster than tones, as in Experiment 1. However, for idiom context, tones were monitored faster than vowels. The authors attribute this inversion in response times to what they claim is the 'privileged role of tone' in Mandarin speech processing, though this is only observable when there is pre-activation generated by the speech context of lexicalised idioms.

There are at least two potential difficulties with the experimental setup, which may indicate that drawing the conclusion that there is a 'privileged role of tone' in processing is inappropriate. First, the context of *chengyu* is not necessarily 'more realistic' speech. *Chengyu* are lexicalised idiomatic expressions, and the ones used in Ye and Connine's experiment were selected such that, given the first three syllables, the fourth (target) was completely predictable. We would expect pre-activation of the fourth syllable to be very strong in this case - perhaps more similar to the context provided by the early portion of a compound word than to that provided by a phrase. Second, and more importantly, the stimuli themselves may be contributing to the observation of a tone advantage in *chengyu*. For the

chengyu, participants explicitly know what the target fourth syllable will be given the first three syllables (e.g. the syllable will be *bá*). They are then asked to determine whether this syllable contains a tone 2 for the tone monitoring task, or the vowel 'a' in the vowel monitoring task, and are found to be faster in detecting the tone. The experimental design, however, counts any syllable that contains the *letter* 'a' in the pinyin romanisation of that syllable as a match. Examples of 'matches' for 'a' from the experiment include: *bái*, *jiáo*, *báo*, *mián*, *huán*, and *jiá*. Even if we assume that all *pinyin* instances of 'a' correspond to the phoneme /a/, there are still multiple vowels to sift through in each syllable - but only one tone. If a diphthong like in *bái* has independent phonemic status, the vowel identification task is further complicated. Adult Chinese rarely write in *pinyin*, and the average Chinese speaker is not often confronted with the fact that there is an 'a' in the word *bái*, as users of more phonemic orthographies are.

Given the potential difficulties of identifying the letter 'a' over identifying a tone in the experiment, we might expect to find a tone advantage generally. Instead, Ye and Connine (1999) found a vowel advantage for all conditions except where the target syllable was known in advance of actually hearing it (i.e. in *chengyu*), indicating that the vowel advantage effect may be fairly robust. It may be that the 'tone advantage' is simply an artefact of orthographic-letter monitoring being more difficult than tone-monitoring given the experimental design. The 'tone advantage' is only observed in the *chengyu* condition because it is only the conjunction of the difficulty of the vowel monitoring task with the strong pre-activation provided by the *chengyu*, that allows the inherent weakness of tone to overcome the inherent vowel advantage.

The temporal asymmetry, whichever way it is directed, suggests that tonal and segmental information are processed independently. That syllables with the wrong tone take longer to reject than those with the wrong vowel may be solely due to the nature of the two phenomena. The entirety of a monophthongal vowel is not necessarily required for

identification. For a language with contour tones like Mandarin, more input is required - to distinguish, for example, the high level tone 1 from the high falling tone 4. If the temporal asymmetry is due solely to a difference in the amount of acoustic information required for identification, it is unclear why, in addition to taking longer to reject, tonally mismatched syllables also result in a greater number of errors - perhaps judgements are made prematurely. It is at least possible that tonal information not only requires more processing time, but is also processed differently, such that words differing only by tone are more difficult to reject as homophones.

In an earlier study, Taft and Chen (1992) presented two characters sequentially, and asked Mandarin speakers to identify whether the two were homophones. There were three types of mismatch: 1) segmental homophones (same segments, different tone); 2) same tone but different vowel; and 3) total mismatches. Response times were slowest, and error rate highest, for segmental homophones. Responses to pairs in which any segment differed - both those that differed only in the vowel, and those that were total mismatches - were relatively faster and more accurate. The authors note that we might expect syllables only differing in tone to be more similar to each other acoustically/phonetically than those differing by a segment, and that this is what is causing homophone judgements of tonally mismatched syllables to be slowed/impaired when compared to segmentally mismatched syllables. We would then expect total mismatches to be faster/easier to reject as homophones than vowel-only mismatches, as total mismatches would be more acoustically/phonetically distinct from each other. That syllables matching in all aspects except the vowel, and total mismatch syllables, are both rejected as homophones relatively quickly and with few errors, while those differing only in tone are at best slow and often incorrectly judged, points away from simple acoustic/phonetic similarity as the only element affecting judgements, and to the possibility of a fundamentally weaker role for tone in perception.

In another experiment (their Experiment 3), Taft and Chen (1992) showed Mandarin speakers a single character, and asked whether that character had any homophones. Four types of characters were included in the experiment: 1) characters with at least one full homophone, and at least one segmental homophone; 2) characters with at least one full homophone, but no segmental homophones; 3) characters with no full homophones, but at least one segmental homophone; and 4) characters with no full or segmental homophones. For characters with no full homophones, errors were significantly higher when characters had segmental homophones. This reflects the finding from Taft and Chen's experiment 1 that segmental homophones are harder to reject as homophones, and is further evidence that tone is less critical than segmental information to Mandarin speakers in making homophone judgements. For characters with full homophones, errors were significantly *lower* for characters with segmental homophones. Taft and Chen (1992) attribute this to a combination of the experimental design, and the way native Mandarin speakers make homophone judgements.

Taft and Chen (1992) tentatively concluded that tones are poorly represented in the underlying phonological representation of a syllable. It at least appears from Taft and Chen's (1992) experiments generally that Mandarin speakers sometimes ignore, or struggle to successfully integrate, lexical tone into homophone judgements, causing segmental homophones to be sometimes considered by native speakers as full homophones. When a character has no full homophones, a mistake of this type registers as an error (e.g. the subject responds that the character has a full homophone when it does not). When a character does have a full homophone, this type of mistake is either not possible (i.e. the character has no segmental homophones), or results in a false positive. Perhaps the same number of segmental homophone judgement errors are committed when characters have full homophones - they just do not register as errors in the experiment.

It is also possible that the presence of segmental homophones is contributing in some way to activation. If so, the characters with both full and segmental homophones are likely to

have more characters which 'count' as homophones, and contribute to activation - therefore, characters with full and segmental homophones are likely to be responded to more quickly and more accurately than characters which only have full homophones. It would be interesting to see if this is indeed the case - whether it is simply the size and/or cumulative frequency of all segmental homophones that matters, or whether full homophones and segmental homophones behave differently (affect characters to which they are related, differently).

Cutler and Chen (1997) performed a series of experiments to examine the role of tone in lexical processing in Cantonese. In their experiment 1, two syllables were presented together to Cantonese subjects, who were asked whether the two syllables formed a word. The second syllable for targets was altered by changing the tone, the rhyme, or both the rime and the tone. When only the tone was altered the error rate was higher than for any other condition, indicating that tone errors are more easily overlooked than segmental errors. Cutler and Chen's (1997) experiment 2 asked Cantonese subjects to judge whether two auditorily-presented open syllables were homophones. Syllables could mismatch in the consonant, vowel, tone, or any combination of the three. Again, the tone-mismatch condition registered the highest number of errors, and in addition the slowest response times. It is more difficult to decide whether two syllables are identical when the only difference between them is tone.

Replicating the experiment with non-tonal speakers (Dutch) resulted in a similar pattern, indicating that the difficulty in differentiating syllables by tone is driven by acoustic difficulty. The experiments of Cutler and Chen (1997) were conducted on syllable pairs in isolation, where tonal identity is most distinct, and found that tone was problematically implemented in homophone judgements. In more common contexts, tonal identity is more variable (Cutler and Chen, 1997), potentially making such judgements even more difficult.

There is also evidence that syllables which match in segmental composition but not in tone activate each other. For example, Lee (2007) 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 *segmental homophones* was taking place, even when the tones were mismatched.

To summarise the findings of these studies, it has been shown that the perception of tones is slower and more error prone than for vowels (Ye and Connine, 1999; Cutler and Chen, 1997; Taft and Chen, 1992; Lee, 2007). Speakers are more likely to count segmental homophones as full homophones than they are to count segmental phonological neighbours as full homophones, and they take longer in deciding. Tonal information is not fully available as quickly as segmental information, and cannot be used for word disambiguation early in the recognition process. Tone is not weighted equivalently to segmental information in Mandarin. This has implications for the definition of the relationship between homophones and near homophones, 'homophone density', and neighbourhood density, and how these affect lexical processing in Mandarin. There appears to be an intermediary level between full homophones and phonological neighbours, here called *segmental homophones*, which are segmentally identical, but differ in tone. For an exemplar model, the phonetic overlap in tokens for 豌 (wān; 'pea') and 碗 (wǎn; 'bowl') will be extensive, but phonetic identity is compromised by different F_0 patterns.

1.3 Apparent Homophones

Research suggests that the processing of ***segmental homophones*** is distinct from that of both full homophones and of conventional phonological neighbours in Mandarin (Ye and Connine, 1999; Cutler and Chen, 1997; Taft and Chen, 1992; Lee, 2007), and while the differential treatment of tone can be presumed to play a role, it is not entirely clear how. To gain better insight, it may be helpful to investigate tonally impoverished stimuli, e.g. through the medium of *pinyin* without tonal markings. This medium is admittedly somewhat unnatural,

although not entirely so; the author has himself witnessed native speakers communicating with tonally unmarked *pinyin* on occasion, for example in online communications or text messages when only English input was available. That native speakers can, at least in certain contexts, communicate without specified tone raises further questions about the role of tone, and its detachability, in lexical representation and processing[2]. If ***segmental homophones*** trigger a degree of homophony judgment, then ***apparent homophones***, which further lack any overt mismatching of tone, are to be expected to also do so. Very little, if any, research has been done on tonally unspecified Mandarin, and so we include this category in our Experiment 1.

1.4 Defining Homophony: levels of identity

Full homophony implies identity, at some level, between two lemmas. As has already been mentioned, the precise nature of our definition of homophony depends at least in part on how we conceive of phonological representation. Within a theory of abstract representation, identity would entail the same sequence of segments, features, gestures etc., allowing also for a separate sequence of identical tones if the representation is auto-segmental. Within a theory of phonetically-rich representation, such as an exemplar-based model, homophony implies a large degree of overlap in acoustic exemplars (identity is therefore attained in the perceptual domain). Tokens will vary according to context, but for any given context, we would expect the tokens for homophones to be very similar. For full homophones such as 湾 (*wān*; 'bay') and 豌 (*wān*; 'pea') there is presumed to be extensive overlap in tokens and a close phonetic identity, resulting in a *de facto* shared phonological form, but different lemmas. Under this assumption of close phonetic identity between the exemplars of full homophones, there is scant (if any) phonetic indication to lemma identity, and word disambiguation must depend

[2] The processes at play may be more general, and not specific to tone, since also with phonological neighbours, an impoverished stimulus as in abbreviated text messaging could create apparent homophones, e.g. both *wish* and *wash* could be written as 'wsh'.

on other factors, such as syntactic and semantic context. Theoretically, on this analysis, incidental, transient homophony can arise, when a given a lemma or sequence of lemmas is articulated in such a way that it falls within the range of possible traces for another lemma with which is not normally considered to be in a homophonic relationship. For example, 'all full' may overlap with 'awful', if certain conditions are met (vocalisation of the final lateral in 'all', rhythmic stress on 'all' and reduction of the final syllable). This transient experience of homophony is more satisfactorily modelled within an exemplar framework because the phonological representation is at once gradient in construction and phonetically transparent. It remains to be investigated whether a transient homophone benefits from the frequency effects of the other lemma (in terms of speech of word recognition), as has been shown for pairs of true homophones (e.g. Gahl, 2008): in this example, this would mean that 'all full' would 'benefit' from the higher frequency of 'awful'. If it did, this would provide quite strong evidence for an exemplar approach.

Whether we adhere to a categorical and abstract representation or to a phonetically-rich, gradient representation has implications for how we view full homophony, and how strict we are on the requirement of identity. Research by Gahl (2008) on English homophone durations indicates that phonetic overlap of homophones may not be as extensive as in classic phonological models, even for words with the 'same' presumed phonological composition. Using a corpus of spontaneous spoken English, Gahl examined the durations of homophones in high/low frequency pairs (e.g. *time* and *thyme*), and found that higher frequency homophones were significantly shorter in duration than their low frequency mates. In this case, the distribution of exemplars for two supposedly homophonous lemmas is not the same - the higher frequency lemma is associated with more numerous, shorter tokens. At least for English, high/low frequency homophone pairs are not in fact strictly homophonous.

What the implications of this finding are for Mandarin is an open question. Gahl (2008) only looked at production, and though she found that higher frequency homophones

were consistently shorter than lower frequency mates, it is unknown whether this information is utilised in perception - it may be the case such 'homophonous' pairs are not processed as homophones, but rather as near homophones. Mandarin, of course, has much more extensive homophony than English, both in terms of cohort size and the extent of homophony in the lexicon. Sherr-Ziarko (2015), in a study on a Mandarin corpus of broadcast news, found significant differences in duration between the highest frequency and lowest frequency members of homophone cohorts, with the low frequency members having longer durations than their high frequency counterparts. This replicates for Mandarin, at least on the extreme ends of the frequency spectrum, Gahl's (2008) results. Sherr-Ziarko did not report on the durations of characters of intermediate frequency, and so we can only speculate as to their behaviour. It is exceedingly unlikely that all of the 25 full homophones of *yì* have unique, perceivable differences in duration to individuate them. Based on Gahl's (2008) and Sherr-Ziarko's (2015) research, it is reasonable to expect some effect of lemma frequency on duration, and possibly along other phonetic parameters, though precisely what form this would take is unknown - perhaps as basic as a division between higher and lower frequency homophones, entailing a phonetically distinguishable division between higher and lower frequency lemmas for all syllables in Mandarin, resulting in (at least potentially) half as much 'true' homophony as appears to exist in Mandarin from a phonological perspective. Perhaps very high frequency lexemes are distinguished phonetically/phonologically from lower frequency full homophones. As a first step, it is necessary to establish whether there are systematic differences in production which can serve as a basis for perceptual differentiation (see **Chapter VII** for an experiment investigating this).

2 Distributional Properties of Mandarin Homophony

Having discussed possible definitions and representations of homophony, we now turn to its distributional properties – both structural and in terms of usage – in modern standard

Mandarin. In the second part of this chapter I present an analysis of homophony distribution based on a corpus of written Mandarin (Da, 2007). This analysis looks at the frequencies of different types of homophone (full and segmental) and the varying density of homophone cohorts in order, in the first place, to understand the following:

i) how homophony is structurally distributed (degree of evenness) and the nature of any patterns of clustering; how are characters/morphemes distributed over unique syllables? Given the small syllable inventory of Mandarin, we might expect to find that characters, and homophony, are spread fairly evenly across all available syllables; is this the case?

ii) How homophony is distributed in terms of frequency of usage. What is the relationship, if any, between homophone cohort size and frequency?

iii) How tone is distributed over segmentally unique syllables, and how does homophony maps onto the four lexical tones. What is the distribution of tones over characters, and does it reduce or exacerbate homophony? Is one tone more common than another, or are they relatively even? Is there any relationship between the tone density of a homophone cohort, and the size of that cohort?

A further issue to examine, given the use of the written medium, is the relationship between morphemes/characters and phonological form. Some characters have more than one phonological form, while being associated with different lemmas. This issue is rarely discussed in the literature, but multiple pronunciations of characters (with multiple semantic content) is a potential confounding factor in all kinds of phonetic and psycholinguistic research on Mandarin. To assess the extent of this phenomenon, we examine how many characters have multiple phonological forms, and how frequent they are.

Our findings will then be used to assess the extent to which distributional properties shape the lexical processing and production of homophones in native speakers, and the extent to which the processing of homophony incorporates the tonal tier. Density statistics can tell us something about the extent of the potential processing 'burden' resulting from homophony, and how this 'burden' is distributed across the lexicon. Frequency statistics can tell us about the relative exposure to an item in a homophone cohort, and whether and how the strength of that exposure influences disambiguation and the fine details of production.

2.1 Selection and Preparation of the Corpus

A large written corpus was used in this study (Da, 2007), containing number-of-occurrence information for each individual character. This corpus contains approximately 200 million characters in total, taken from a wide range of texts in diverse subject fields. All texts are 'formal', in that the corpus does not include emails, text messages, or postings to online forums - though it does include texts published only online, such as 'newsletters and magazines and online course learning materials, etc.' (Da, 2007). The corpus contains 9,933 unique characters - this is a fairly high number, and indicates that the corpus is reasonably extensive in lexical scope, as estimates of how many characters are required to be an accomplished reader are in the range of from 3-4,000 (Norman, 1988) up to 4,500 (Zhang and Peng, 1992). The corpus provides only character counts. Where a character has more than one phonological form, the corpus does not allow the separation of the occurrence of individual forms. For example, the character 撒 is associated in the corpus with two phonological forms, *sā* and *sǎ*, and it is impossible to differentiate the two, without context.

Admittedly, written Mandarin will differ from any given variety of spoken Mandarin in a number of ways - certainly including frequency of use, which is an important statistic in this work. It is also difficult to create a written corpus based uniquely on only one specific Chinese language. With the exception of communities using traditional characters, such as in

Taiwan and Hong Kong, the Chinese languages are written with the same script (i.e simplified characters). As characters represent semantic units and do not reliably reflect phonetic information, it is difficult to discern the language background of a given text. A written corpus is effectively a conglomeration of language varieties - and we can presume that the relationship with the spoken form, and hence also the specific processing parameters - will differ between varieties. Given the difficulties inherent in collecting and processing comprehensive spoken corpora, it was decided that the advantages in terms of the size and scope of Da's (2007) corpus of written material over available spoken corpora outweighed the potential confound of language medium.

As the focus is on the nature and distribution of homophony as it relates to representation and access in a putative language user, some of the figures will differ from other published analyses of Mandarin homophony, even analyses based on the same corpus. The intention here is to describe the distribution of morphemes over a limited syllable inventory for an idealised literate individual, rather than the maximum homophony of Mandarin in the abstract. This means the calculated homophone densities in this work will often be smaller than might be expected when considering the Mandarin syllable inventory and an exhaustive dictionary, as the goal is to prune a comprehensive dictionary of rare words until it roughly resembles a speaker's lexicon.

One additional possible difficulty is that it is not the case that a speaker will know every character for every morpheme he or she knows and uses in the spoken medium. In a more phonologically transparent orthography, for example in English, the number of words a literate individual is capable of reading is more closely equivalent to the size of that person's vocabulary. In Mandarin, vocabulary and the visual representation of that vocabulary are two distinct sets of knowledge. Rather than a simple binary literate-illiterate distinction, which may be roughly appropriate for English, there is a considerably broader spectrum of literacy categories for the Chinese languages. This is potentially problematic in a study which is

concerned with investigating the effects of homophone density - though not insurmountable, as educated individuals will generally be capable of writing most of the morphemes they would typically comprehend or utter.

As the current work is primarily interested in the morphemes/characters which average educated individuals can reasonably be aware of, many characters in the corpus need to be excluded from consideration. A very large number of the characters in the corpus are so infrequent as to be unrecognisable by even highly educated Mandarin speakers - for example, more than 2,000 unique characters occur fewer than three times in the corpus - or are highly technical and therefore only known and used by those in the appropriate fields.

A frequency-cutoff was therefore employed, such that all characters with fewer than 1000 occurrences in the corpus were assumed to be unknown for most speakers, and therefore excluded. This resulted in a list of 3,524 unique characters. This list was then scrutinised with the assistance of an educated native speaker. Any characters which were unknown to this native speaker were also excluded, resulting in a final number of 3,483 characters. The general intention was to produce a fairly conservative list, with a bias toward excluding characters that some speakers would know in order to avoid including characters many would not know, because the distribution effects of characters being unknown were deemed more problematic than those resulting from excluding some known characters. The primary driver of this conservatism was utilitarian, as a wide range of experimental materials were to be drawn from this corpus.

The effect of excising some 65% of the characters represented in the corpus resulted in only a minor drop in the percentage of the corpus utilised, however - 99.5% of the roughly 200 million character tokens in the corpus are covered by just these 3,483 unique characters. While it is possible that speakers will know some of the excluded characters, difficulties will always be encountered when trying to calculate what in effect amounts to an estimate of character-vocabulary which is intended to be widely applicable to different persons. This is

not expected to be a significant source of error in homophone counts however, as the characters in question are of such low frequency that their presence/absence should have only minimal impact on the overall profile of a syllable, though it would potentially be interesting to see how those speakers who do know these characters produce them.

For the purposes of the experiments which are designed to utilise this profiling information, it was important to ensure that no marginal characters were included that some participants would not know (and therefore induce guessing). The resulting corpus represents a fairly conservative estimate of the number of characters/morphemes an individual possesses knowledge of, but it results in a set of characters which all literate Mandarin speakers should know, and it is assumed that though the precise extent of homophony for an individual is beyond easy calculation, the homophone densities used in this work are at least within the range of the average person.

2.2 *Summary of Terminology*

Before beginning an analysis of the distributional attributes of homophony in modern standard Mandarin, there are a number of additional terms that will be employed which require some explanation.

Character Frequency

This is the number of times a particular character occurs in the corpus, and is expressed as occurrences per million characters. The corpus contains approximately 200 million characters in total (see below for more details of the corpus selected) and so ***character frequency*** is the number of occurrences divided by 200. For example, the character 子 occurs 640,640 times in the corpus, and therefore has a frequency of 3327.23.

Full Homophone Frequency

This is the combined frequency of all characters sharing the same segmental string and tone, expressed as occurrences per million characters. For example, the syllable *wǎn* has six associated characters in the (reduced) corpus, with a combined frequency of 430/million characters.

Segmental Homophone Frequency

This is the combined frequency of all characters sharing the same segmental string regardless of tone, again, expressed as occurrences per million characters. For example, the syllable *chai* has only two associated characters in the reduced corpus - 拆 *chāi* 'to tear', and 柴 *chái* 'firewood', with counts of 6,412 and 7,276. The segmental homophone frequency of *chai*, i.e. the combined frequency of all full homophone frequencies (in this case *chāi* and *chái*), is therefore roughly 71/million.

Tone Density

This is the number of unique tones occurring in conjunction with a particular segmental string, regardless of the number of homophones in each tone (i.e. for a cohort of segmental homophones). The syllable used in the previous example, *chai*, has a tone density of two, as *chāi and chái* are the only two syllables in the reduced corpus.

Segmental Homophone Density

The number of characters which share the same segmental structure, regardless of tone. The syllable *chai* has a segmental homophone density of two.

Full Homophone Density

The number of characters which share both segmental structure and tone (i.e. full homophones). The syllables *chāi* and *chái* both have full homophone densities of one.

3 Results

3.1 Characters and Phonological Form

The excised corpus contains 3,483 characters, around 88% of which have only one phonological form *in modern Standard Mandarin*. The character 已 'already', for example, has only one phonological form, *yǐ*. The remaining characters have multiple phonological forms corresponding to different lemmas. Two independent lemmas may come, historically, to be written with a single character in a variety of ways (e.g. writing errors). The character simplification programme of the 1950s produced a set of these 'multiple pronunciation' characters by eliminating certain orthographic distinctions. For example, two separate lemmas *fā* 'send out' and *fà* 'hair', which were historically written with independent characters (發 and 髮 respectively, which are still used in communities using 'traditional characters' such as Hong Kong and Taiwan), are now both written as 发 in mainland China.

In some cases a character can represent multiple lemmas which differ from each other in both segments and in tone. For example, the character 乐 can be either *lè* 'joyful' or *yuè* 'music'. In other cases, a character has multiple forms that differ only in tone; there are 212 examples of these in the corpus (i.e. one character represents multiple segmental homophones), as with *fā*/*fà* 发 above. Of these, 202 have only two forms (i.e. two different tones), but 9 have three forms (tones), and 1 character has all four tones (as well as the neutral tone) as possible forms. Figure 3.1 illustrates the relationship between characters and phonological form for all 3,483 characters in the corpus.

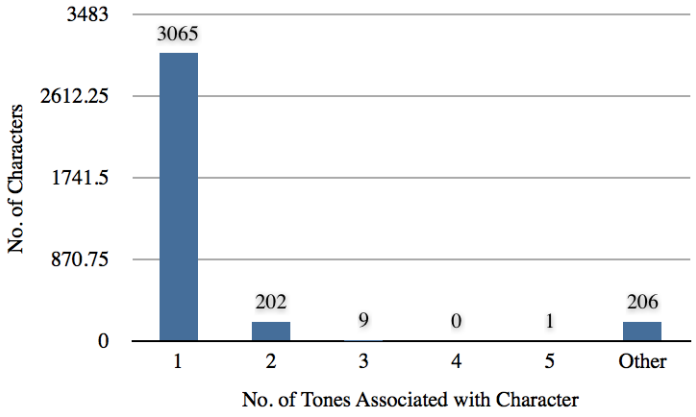


Figure 3.1. Number of tones associated with a single character, i.e. '2' indicates that the character represents a unique segmental string that has two possible tones associated with it (the character represents two segmental homophones). 'Other' indicates that a character has more than one associated unique segmental string.

While it can be seen that homographs (same character, different phonological form and lemma) constitute a minority, it is nevertheless not an insignificant minority. What is more, a high proportion of these denote a sub-set of segmental homophones (i.e. the multiple forms are segmentally the same and differ only in tone). In this thesis we do not further investigate the possible effects of homography on lexical processing, but acknowledge more generally that the processing of reading written characters involves another layer of associations and possible inheritance effects that deserve investigation.

3.2 Distribution of Tone

The distribution of the four lexical tones (and the neutral tone) over characters, as well as the total number of occurrences in the corpus by tonal identity, is given in Figure 3.2. A roughly comparable number of characters are associated with phonological forms with Tones 1 and 2, while rather fewer are associated with Tone 3, and rather more with Tone 4. There are only 12

characters in the corpus that are associated with Tone 5, and these are exclusively grammatical particles, interjections, and onomatopoeia.

The finding that Tone 4 characters are the most numerous, and Tone 3 characters the least numerous, is directly mirrored in child acquisition of Mandarin tone, where Tone 4 has been found to be acquired earliest in development (in terms of production), and Tone 3 latest (Wong, 2013). The reasons for the early acquisition of Tone 4 and the late acquisition of Tone 3 are perhaps phonetically transparent. Tone 4, which has a falling pattern, involves a simple movement of pitch and, being a fall, is also physiologically most 'natural' (requires least effort), in that, all other things being equal, F_0 will naturally decline as the supraglottal air pressure equals and then exceeds that below the glottis. As such, a falling tone can be said to be phonetically 'unmarked' in some way. Tone 3, which has a falling then rising pattern, is the most articulatorily and acoustically complex Mandarin tone (Wong, 2013), and thus can be said to be phonetically 'marked' in some way. Unlike other Mandarin tones, it also has two distinct forms - the 'full' form just described, and a 'short' form consisting of only the fall. Also uncommon for Mandarin tones is the fact that tone 3 is also involved in a tone sandhi process, where two tone 3 syllables in succession will result in the tone of the first syllable becoming tone 2. Tone 2, which is a rising tone, is commonly confused with tone 3 (Wong, 2013), and this particular tone sandhi may be related to this perceptual confusion.

The number of occurrences of characters by tone is roughly in proportion to the number of characters associated with a particular tone, with the clear exception of tone 5. Most of the tone 5 characters are grammatical particles, and have extremely high frequencies.

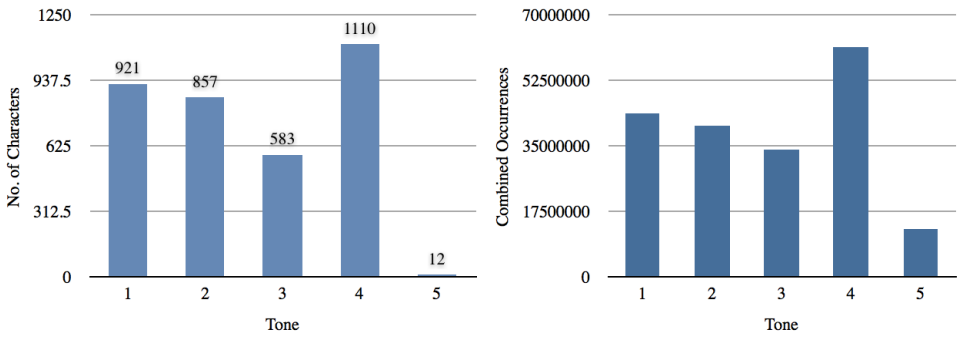


Figure 3.2. *Left* - Number of characters associated with the four lexical tones and the neutral tone. *Right* - Combined number of corpus occurrences for characters by tone.

3.3 Tone Density

There are approximately 400 unique base syllables in modern Mandarin (i.e. syllables disregarding tone), the precise figure depending upon whether certain interjections are counted as linguistic syllables. The excised corpus contains 399 unique segmental strings. As the number of unique segmental strings is small (and particularly as nearly all morphemes are monosyllabic), we might expect tone to be maximally exploited by having each syllable paradigm fully utilised with respect to syllable and tone combinations, yielding around 1,600 unique tonally specified syllables. As there are certainly more than 1,600 morphemes in Mandarin, it is inevitable that a large number of these syllables will have a number of morphemes/characters associated with them (homophony).

It is perhaps surprising then that, even though there are only a small number of theoretically possible unique Mandarin syllables, many of them do not appear. We would of course expect to find a number of accidental gaps, however in Mandarin there are (many) gaps in the take-up of the possible syllable inventory. Of a potential 1600, there are only 1089 unique tonally specified syllables in the corpus. Also interesting is the under-employment of the tonal range with each unique syllable. On average, fewer than 3 tones are associated with

each segmentally unique syllable. While it is true that a number of tonally specified syllables are excised out from the abridged corpus, Duanmu (2007) gives the absolute maximum at only 1255. In a language with a fairly limited syllable inventory and very few polysyllabic morphemes, the fact that around one fifth of apparently phonotactically viable syllable+tone combinations are unused merits investigation. Gaps in the paradigm will have arisen through historic processes, but it is curious that effectively 'squashing' the lexicon into a even more reduced set of forms did not constitute a counter-reductive force, historically. It also adds evidence that 'super-homophony' is not, in functional terms, an unviable burden.

The *tone density*, or the number of tones associated with a given segmentally unique syllable, of the 399 unique segmental syllables in the excised corpus, is presented in Figure 3.3. It is clear from the figure that the full tonal paradigm of four tones (five, if the 'neutral' tone is included) is not consistently utilised - only around a third of all syllables have four or five associated tones. For example, the unique segmental syllable *ma* has characters with all four lexical tones - *mā*, *má*, *mǎ*, *mà*. Around a third of unique segmental syllables have characters with *some subset* of three of the lexical tones. For example, *xia* has characters for the tonally specified syllables *xiā*, *xiá*, and *xià* (but no characters for *xiǎ*). We might expect a few gaps like this, but what is more surprising is the large number of segmentally unique syllables which have only one or two associated tones (137 of 399 syllables, or over 34%). For example, only one tone (tone 4) occurs with the syllable *miu*, and only one character occurs with the syllable *miù:* 谬 'absurd'. For four potential syllables, only one is utilised, and that only has a single character. This is in contrast to a syllable like *ji*, which occurs with all four tones, each of which has a number of characters - *jī* has 20 characters; *jí* 14; *jǐ* 3; and *jì* 17. The difference between these two syllables illustrates the vast unevenness in the relationship between characters/morphemes, segmental strings, and tones in the lexicon - and results in even denser concentrations of homophony than are structurally necessary. It is a truism that linguistic paradigms are very seldom complete or perfect, being as they are the

patchwork result of many interacting and complex historical process, and not the product of any form of 'intelligent design'. Nevertheless, if we think of homophony as a source of potential ambiguity, and ambiguity as placing a burden on lexical processing, any model of word recognition valid for Mandarin must be capable of explaining how this degree of homophone density is accommodated. At the same time, concentrating homophony into areas of the syllable inventory allows for a given part of the lexicon to be homophony-free. Furthermore, a differentiation of the lexicon (into homophonous and non-homophonous areas, and within homophonous areas a very uneven distribution of homophony) may actually aid lexical processing by carving out statistically sketched 'landmarks'.

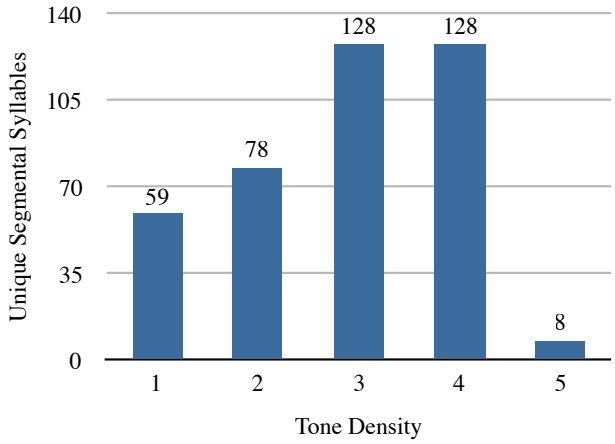


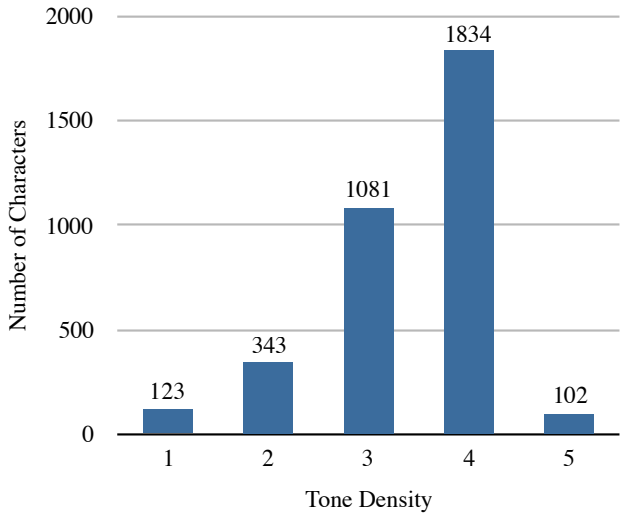
Figure 3.3. Tone Density of Unique Segmental Syllables (i.e. number of tones associated with each segmental syllable).

It is worth repeating here that these figures differ from a dictionary-based 'theoretical maximum', as a great many low frequency characters have been excluded. If a particular excised character was the only occurrence of *chài*, for example, this changes the tone density of *chai*. Duanmu (2007) gives as a maximum 402 unique segmental syllables - 35 having only one tone, 59 having two tones, 130 having 3 tones, and 178 having four associated tones. Though the current excision has increased the number of unfilled tonal paradigms (from 56%

to 66%), it is interesting that even at the theoretical maximum less than half of all Mandarin syllables carry all four tones.

3.4 Homophony and Tone Density

There are 3,483 unique morphemes/characters in the corpus, with a mean full homophone density of around 3.2. Over half of these characters occur in syllables with all four tones (e.g. Tone Density of 4) (see Figure 3.4).



Figures 3.4. Number of Characters by Tone Density.

There is an average of 3.58 characters associated with each tonally specified syllable that occurs in a fully utilised tone paradigm - more than the 3.2 mean characters for the entire corpus, and *all* of the other possibilities have fewer associated characters than the average (see Figure 3.5).

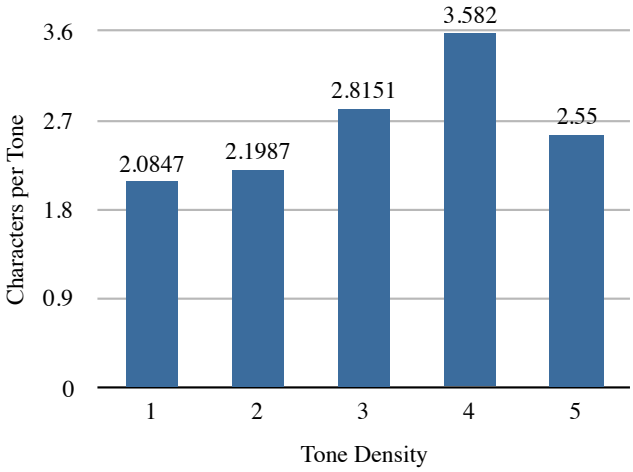


Figure 3.5. Average Full Homophone Density by Tone Density.

These figures show there is a tendency for character clustering in Mandarin, and the more tones associated with a given segmental string, the *more* characters are associated with each tonally specified variant of that string. This is a general tendency for Mandarin. The average full homophone density for syllables with only one tone is 2.08. The average full homophone density for syllables with two tones is 2.2, for syllables with three tones it is 2.82, and for syllables with four associated tones it is 3.58 (Figure 3.5). The exception is when tone 5 is involved. Tone 5 (or the neutral tone) is not a full lexical tone, and there are only 8 syllables in the corpus in which all four lexical tones and tone 5 are represented. Tone 5 is generally only associated with grammatical particles, interjections, and so forth, and these very rarely have any homophones at all, possibly explaining why the average full homophone density for syllables containing all five tones is lower than those with the four lexical tones.

In a general sense, we find a few highly dense syllables along with many that have only one character. Syllables with higher full homophone densities are not randomly distributed, but systematically occur in segmentally-specified syllables that have higher tone densities.

Having more tones associated with a given segmental syllable in theory also allows for a greater number of characters to be associated with that segmental syllable (increases the number of segmental homophones), without increasing full homophone density. An idealised syllable with three tones, with the mean full homophone density, would have 9.6 segmental homophones. A syllable with four tones and the same full homophone density would have 12.8 segmental homophones. In reality, syllables with three tones have on average a segmental homophone density of 8.46, while syllables with four tones have a segmental homophone density of 14.32. Base syllables of higher tonal density appear to be more capable of sustaining higher full homophone density.

If we were to take a traditional view, any concentration of homophony beyond the mean places additional burdens on lexical processing, as it increases competition, and we might expect languages to avoid this in general. In Mandarin there is a regular increase in full homophone density as tone density increases. It is of interest how this uneven homophony is accommodated, and whether homophone-clustering may actually aid processing through a statistically mapped out lexical topography.

3.5 Homophone Density

Segmental Homophone Density

The distribution of homophones is extremely uneven. An approximately equal distribution of the 3,483 characters across the 399 unique segmental syllables would yield an average of about 9 segmental homophones per syllable. In reality, we find that high levels of homophony are heavily concentrated in a relatively small number of syllables. A number of unique syllables have segmental homophone densities of more than fifty, with full homophone densities exceeding twenty-five. At the same time many syllables are not fully exploited, and have few (and sometimes only one) associated morphemes, thereby precluding in part, and sometimes in full, the potential disambiguating benefit of tonal contrasts.

Figure 3.6 shows the distribution of characters to unique segmental syllables (i.e. segmental homophone density). If all characters were spread evenly across all syllables, we would expect to find around 13 characters per segmental syllable, i.e. a mean segmental homophone density of 13. Any syllable exceeding this uniform 'ideal' of 13 is over-burdening homophone pressure, while any syllable with fewer segmental homophones is under exploited.

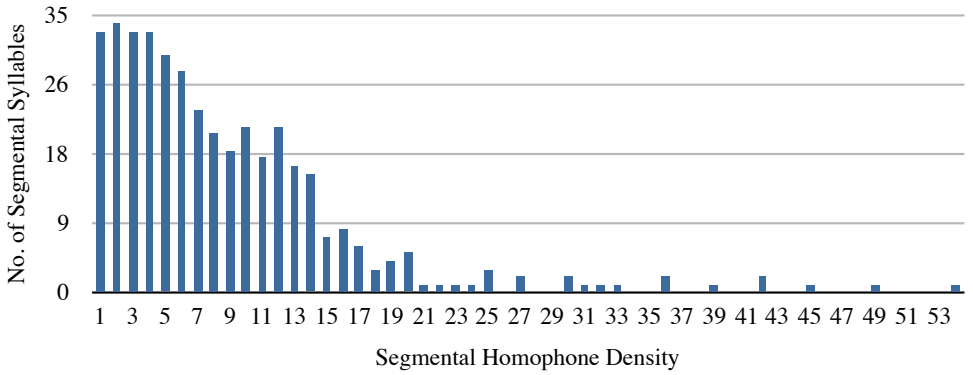


Figure 3.6. Number of unique segmental syllables by segmental homophone density. For example, there are 33 unique segmental syllables with a segmental homophone density of 1 (e.g. *miu*, which has only one character 谬 *miù* 'absurd').

We find that the majority of segmental syllables have fewer than 14 characters (segmental homophone density of less than 14). Many syllables in Mandarin have an even lower segmental homophone density, with a particular concentration of around 4 or fewer. A segmental syllable with a density of 4 could, theoretically, completely avoid have any full homophones, through each character being associated with a different tone. Syllables with densities lower than 4 are unable to fully exploit the four-tone system.

As segmental homophone density increases beyond around 14, the number of syllables drops dramatically. There are only a few segmental syllables which have very high densities - for example *ji*, on the far right of Figure 3.6, has a segmental homophone density of 54. Such homophone-dense syllables, though there are few of them, are very visible, as

they represent a substantial portion of Mandarin lemmas. There are 33 segmentally unique syllables which have segmental homophone densities of 1. Such syllables, like *miu*, have only one character associated with only one of the four possible tones (in this case *miù*). The syllable *ji* has more characters than all 33 of these syllables together.

Full Homophone Density

The situation with tonally specified syllables is similarly, if less spectacularly, uneven in distribution. There are a large number of tonally specified syllables which have only one associated character, meaning these characters have no full homophones. As with the segmentally unique syllables, the number of tonally specified syllables decreases sharply as the number of full homophones increases (see Figure 3.7).

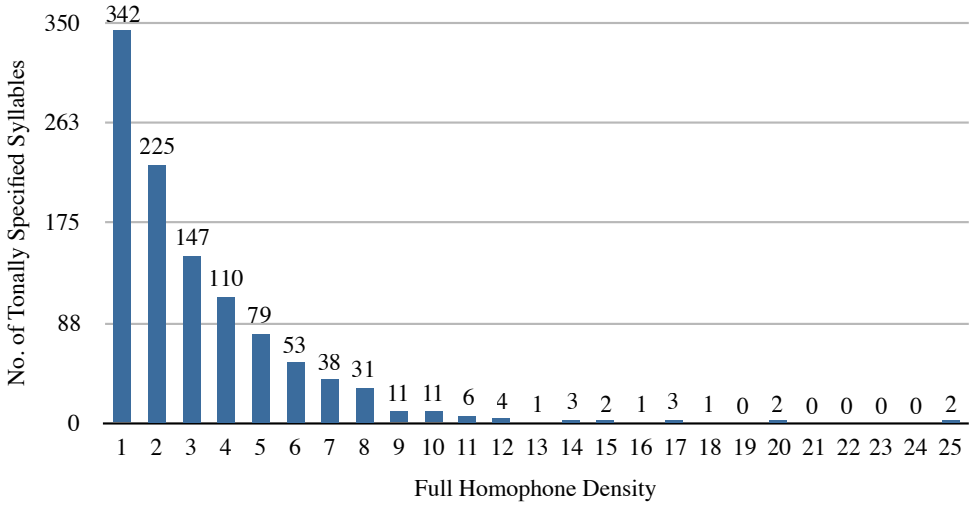


Figure 3.7. Number of tonally specified syllables by full homophone density. The uniform 'ideal' distribution is 3.2 full homophones per tonally specified syllable.

Around 13% of tonally specified syllables have a full homophone density in the proximity of the average of 3.2. In contrast there are 342 syllables in the corpus, or around 31%, with only one associated character (i.e. no homophones at all). A total of 36 syllables have 10 or more full homophones, and 2 have 25 full homophones.

The historical processes which have led to this concentration of characters/morphemes into a relatively small pool of select syllables within the already small pool of total syllables (while many syllables have no homophones at all) may be related to ongoing processes in Chinese languages in general - for example, the loss of coda plosives in Mandarin (c.f. Cantonese, which maintains them), the loss of diphthongs in Shanghainese, and the loss of tonal contrasts in Mandarin and Shanghainese (Cantonese and Taiwanese have 7, Mandarin has 4, and Shanghainese only 2) (Norman, 1988). From a historical perspective, the syllable pool has consistently decreased (while homophony balloons). Duanmu (2007) posits a syllable inventory, including tones, of 3,000 for Middle Chinese in the 6th century (compared to 1255 for modern Mandarin).

Rather than increasing the number of potential syllables in the face of hyper-homophony, or even maintaining that number, from a historical perspective Mandarin appears to be continuing to lose syllables and increase homophony. Sound change mechanisms for the ongoing loss of syllables in Mandarin are complex. Duanmu et al. (1998, reported in Duanmu, 2007) found in a corpus of natural speech that up to one third of all syllables in Mandarin are produced unstressed and reduced, and suggested that as higher frequency syllables exhibit a greater tendency toward phonetic reduction (c.f. Bybee, 2001), and as Mandarin has relatively few syllables which are therefore used relatively more often, reduction and loss of contrasts becomes progressively more likely.

It is unlikely these increases in homophony would have occurred had they placed insuperable demands on language users. It should be noted that concentrating homophony into one part of the lexicon effectively leaves a significant portion of the lexicon homophony-free, thereby creating a privileged sub-part (around 31% of characters/lemmas). In addition, where it exists, homophony is unevenly distributed, with high concentrations occurring where tone density is also high, and thus marking out yet further distinctions that may even aid speakers. How speakers have adapted to this unevenly structured lexicon (with respect to

homophony) will be explored in the following chapters, through a series of psycholinguistic and production studies.

4 Summary

Analysis of the corpus revealed that the four lexical tones are not distributed equally. There is a comparable number of characters bearing tones 1 and 2, but substantially fewer tone 3 characters, and more tone 4 characters. This distribution may be related to acoustic and articulatory factors: Tone 3 is the most complex of the Mandarin tones, and is acquired the latest, while tone 4 is physiologically less marked, and is acquired the earliest (Wong, 2013). The analysis also revealed variable utilisation of the Mandarin tonal paradigm. Given the 399 segmentally unique syllables in Mandarin, we might expect to find around 1600 tonally specified syllables all 4 tones were used for each syllable), but we only find 1089, with around a third of all segmentally unique syllables associated only with one or two tones. In addition to, and indeed as a result of the under-utilisation of tonal distinctions, it was also found that homophony is clustered. We find that syllables with the full tonal paradigm (e.g. all four lexical tones associated with a particular segmental string) have high concentrations of homophony, while syllables with fewer associated tones have less. Many syllables have no segmental or full homophones at all, also precluding in part or in full the potential disambiguation potential of tone across the lexicon.

The distributional topography of homophony in Mandarin would strongly suggest that, despite the potential for ambiguity in homophony, super-homophony is anything but avoided in the Mandarin lexicon. The uneven distribution of homophony creates a privileged sub-part of the lexicon (just over 30%) with no homophony and no danger of ambiguity. Nevertheless, that leaves almost 70% of the lexicon as homophonous, and even here the distribution of homophones is markedly uneven. Such distributional disparities may even aid processing. A corpus analysis of this type only hints at a number of questions related to actual

lexical organisation and processing. For example, what is the precise role of tone in Mandarin homophony? Previous research has indicated that tone is far less critical in homophone judgements than segmental material, and we find a strong tendency for tone to be under-utilised in distinguishing syllables in the corpus. Are segmental homophones organised or processed differently than phonological (segmental) neighbours? Are there any effects of segmental homophone density on word recognition or phonetic production? What about full homophone density?

In addition to structural disparities, there are of course also large differences in word frequency. Is there any benefit to being a low frequency character in a highly dense syllable, containing high frequency competitors? For example, is there an effect of frequency at the 'phonological' level, such that a high frequency homophone mate is beneficial for low frequency characters? Conversely, are there actual differences in the production of characters, based on frequency of use, which mean such characters are 'less-than-homophonous'? Are high frequency characters more prone to having low frequency full homophone mates, because the two will actually differ in production, whereas two low frequency characters cause difficulties?

This chapter has provided an overview of the distribution of Mandarin homophony, based on an analysis of a written corpus. The extent of homophony in modern Mandarin is unparalleled in the European languages upon which most models of lexical representation and access were developed, and it presents a number of unique challenges to those models. The next chapter will look at some of these models, and discuss their appropriateness and adequacy in dealing with a super-homophonous language such as Mandarin.

Chapter IV. Theories of Lexical Representation and Access

The previous chapter examined the structural distribution of homophony in Mandarin, and highlighted the varied ways that distribution was found to be irregular. The extensive but variegated nature of Mandarin homophony raises a number of questions regarding lexical representation and organisation, which have implications for both word recognition and perception, as well as word production.

For example, how is tone represented in the lexicon - what is the nature of the relationship between segmental homophones and full homophones, and in what ways does homophone density (variously defined) affect access? Do full homophones share a single phonological representation, and if so, how is the representation structured? Or does each lexeme have an independent phonological representation, that happens to share many or all aspects with other lexemes? What is the role in processing, and locus of, lexical frequency? What is the nature of the representation - is it abstract, or phonetically rich (e.g. exemplars)?

The way that tone is represented in the lexicon has implications for homophone disambiguation, and more generally for our definition of homophony in Mandarin. How 'embedded' tonal information is in the lexical representation has ramifications for homophone density and lexical processing of homophones. It also brings into question the role of 'identity' in homophony, and at what level such identity lies. For example, if homophony is fundamentally segmental in nature, what is the role of tonal information?

The variegated nature of homophony in Mandarin and its irregular distribution poses a number challenges for current models of homophony. The present chapter will provide an overview of lexical representation and lexical access models, and discuss if and where such models fail to adequately account for Mandarin. The first section consists of a brief review of abstract representational theory, followed by a discussion of usage-based models. In section

three the various models are discussed in light of how each structures homophony, and what expectations for linguistic behaviour such structures impose.

1 Abstract Representational Models of Lexical Representation

Word recognition is the process of matching the pattern of the input to a pattern in memory (Tyler and Frauenfelder, 1987; Luce and Pisoni, 1987). The point at which recognition occurs depends on physical properties of the input such as duration, and intrinsic properties of the word such as lexical frequency. Lexical access on the other hand is the process by which the properties of a lexical entry are accessed, and includes phonological, syntactic, semantic, and pragmatic information (Tyler and Frauenfelder, 1987).

Four influential models of spoken word recognition are discussed in this section: the Cohort model (Marslen-Wilson, 1987), TRACE (McClelland and Elman, 1986), Logogen (Morton, 1969), and the Neighbourhood Activation Model (Luce and Pisoni, 1998).

1.1 Cohort

The word recognition process begins with the input (i.e. sound waves, or the written word). In the Cohort model (Marslen-Wilson, 1987), a *contact representation* is created from the early portion of the input, which generates the set of words in the language with matching contact representations. This is the eponymous *cohort*. The assumed nature of the contact representation varies somewhat by the era of the Cohort model, but in Marslen-Wilson (1987) consists of a feature matrix, and the resulting cohort contains every word that begins with a near-match to this particular feature-matrix set of feature values. The 'near-match' aspect of the model is to accommodate larger cohorts, so as not to exclude the possibility of mispronounced or misperceived (for example, due to noise interference) stimuli (Marslen-Wilson, 1987). The composition of the contact representation in a particular model determines two important things: 1) the time-course of 'initial lexical contact', as such

contact cannot be made until enough speech has been heard to compute the contact representation (and larger contact representations such as syllables will take longer to hear than phonemes, for example); and 2) the size of the initial cohort (less-specified contact representations casting larger nets, so to speak) (Tyler and Frauenfelder, 1987). Words in the cohort are evaluated against continuing input in parallel, and eliminated by failing to match the bottom-up input or top-down context. A word is said to be recognised when it is the only remaining word in the cohort.

Cohort was initially conceived as a 'discrete initial contact' theory, meaning that there is a discrete phase in which the members of a cohort are determined. This reduces the number of lexical entries that need to be considered in later stages, though it does require the system to 'know' when a word begins in order to compute an accurate contact representation - and any error in perception results in a complete failure of the perceptual system. Later iterations of the model attempt to get around these difficulties in two ways. First is the concept of 'activation' - units are not stricken from a discrete list when they fail to perfectly match the input. Rather, they cease to increase in activation, and begin to decay. The second is the introduction of feature matrices in place of phonemes. Marslen-Wilson (1987) attributes this change at least in part to a desire to limit the disturbance caused by a phoneme misidentification - mistaking a /t/ for a /d/ may have drastic implications for the cohort of word candidates, while mistaking the feature [voice] is, at least potentially, less detrimental. The 1987 version of Cohort also introduced lexical frequency effects to the model, which were not accounted for previously. Frequency is assumed to affect the activation level of a particular word, with higher frequency words requiring more contrary evidence to eliminate than lower frequency words.

These modifications serve to bring phonetic variability and usage-effects into an abstract representational model - for example, well-known problems with converting continuous phonetic variability into discrete phonological categories (e.g. mistaking /t/ for /d/

due to noise interference, speaker variation, etc.), and frequency of use effects. They introduce a degree of uncertainty into a discrete, abstract phonological system which demands classificatory certainty. Models of spoken word recognition purposefully designed to account for usage will be discussed in Section 2.

In Cohort, words are activated in parallel and at no cost, meaning the size of the cohort has no independent effect on the outcome of the recognition process (Luce and Pisoni, 1998). However, a cohort with more members is presumably more likely to have words which fail to be eliminated early in the recognition process (because they match the intended target word for longer, for example). The recognition point for larger cohorts is, therefore, likely to be later than for smaller cohorts. This issue is of particular concern in Mandarin, where the result of a bottom-up phonetic or phonologically-based winnowing process is likely to result in a cohort still heavily populated with homophonic word-candidates. Models specifically designed to account for homophony are discussed in Section 3 of this chapter.

1.2 Logogen Theory

In Logogen theory (1969), units called logogens monitor both bottom-up sensory input and top-down context relevant to a particular word. Input which corresponds to a particular logogen has an excitatory effect. When a threshold of input for a logogen is crossed, the word becomes available as a response, and the word is recognised.

Context raises the activation level of a logogen, so that it requires less input to reach threshold, while increasing frequency serves to lower the threshold level itself. Context and frequency therefore have similar effects in the logogen model, as both serve to reduce the amount of sensory input required to reach threshold. Logogens are independent of each other, and we therefore would not expect to find any effect of either neighbourhood or homophone density on the point at which a logogen reaches threshold. Logogens, or their near

equivalents, are utilised in both TRACE and the Neighbourhood Activation model. These two models are discussed in Sections 1.3 and 1.4.

1.3 TRACE

In a competing model contemporary to Cohort, known as TRACE (McClelland and Elman, 1986), a discrete initial contact representation is not assumed. Like the later formulations of Cohort, information processing occurs through the 'excitatory' and 'inhibitory' interactions among 'units' - units which are mutually consistent with each other are mutually excitatory, while units which are not mutually consistent are mutually inhibitory.

In TRACE units are of three types, each on a separate level - features, phonemes, and words. Units on the feature level detect features, and excite units on the level above that match that feature (i.e. phonemes which exhibit the feature), and inhibit phonemes which do not share the detected feature. The connections between units on the same level are also inhibitory, such that an activated feature will also suppress the activation of other features, further preventing excitation on other levels inconsistent with the input (Luce and McLennan, 2005). Phoneme units function in a similar fashion, receiving excitation from features, spreading it upward to words which contain those phonemes, and receiving excitation from the word level. For example, the lexical effect - where an ambiguous phoneme is interpreted such that it contributes to forming a word - is explained as the phoneme which creates a word receiving excitation from the word level, causing that phoneme to dominate competitors which do not form words. Word level units receive excitation from phonemes, but also from higher 'contextual' levels such as semantics. Throughout the duration of the word, in both Cohort (Marslen-Wilson, 1987) and TRACE (McClelland and Elman, 1986), successive portions of information provide activation to lexical entries which continue to match the input, while those which do not continue to match stop receiving activation and begin to decay. Lexical frequency effects are accommodated in TRACE and Cohort through altered

levels of 'resting activation'. Higher lexical frequency is represented as a higher level of resting activation, meaning such words require less external activation to reach the activation threshold.

One potential drawback of the two models, and other structuralist-type models such as Shortlist (Norris, 1994), is the reliance on abstract phonological constructs such as features (for TRACE and some versions of Cohort) and phonemes (Cohort). This removes the possibility for a great deal of potential information in the input, such as speaker variability and phonetic context, to play a role in word recognition (Luce and McLennan, 2005), as well as imposing difficulties on the recognition process due to misidentifications caused by that variability.

1.4 Neighbourhood Activation Model

In the Neighbourhood Activation model proposed by Luce and Pisoni (1998), stimulus input activates acoustic-phonetic patterns that resemble the input (a similarity neighbourhood), which in turn activate 'word decision units' (similar to logogens) sensitive to those patterns. The word decision units monitor the activation levels of the acoustic-phonetic patterns to which they correspond, as well as relevant higher level lexical information, until one decision unit reaches 'criterion' and the word is recognised. Word frequency is assumed to be a part of higher level lexical information, and serves to bias the word decision units by altering their activation levels.

A similarity neighbourhood is the collection of words that are phonetically similar to the stimulus word, and the activation of the neighbourhood is identical to the activation of the associated word decision units. Luce and Pisoni (1998) acknowledge that phonetic similarity is difficult to define, but settle on the working definition that two words are similar if they differ by one phoneme - substituted, added, or subtracted.

Thus, the stimulus input generates a similarity neighbourhood of words, of which the target is (typically) a member. The number of words in a similarity neighbourhood (neighbourhood density), and the average frequency of those words (neighbourhood frequency) differ according to the target word, and the model assumes that these differences, in conjunction with the frequency of the target word itself, affect word identification.

Frequency effects are assumed to be relative in the Neighbourhood Activation model. That is, the model predicts that it is not the frequency of the target word on its own that is critical in determining the time-course of word identification, but its frequency in relation to the words in its similarity neighbourhood. The model therefore predicts that high frequency words in high frequency neighbourhoods will be identified on a similar timescale to low frequency words in low frequency neighbourhoods. High frequency words in low frequency neighbourhoods are expected to be identified relatively quickly, while low frequency words in high frequency neighbourhoods relatively slowly. Higher neighbourhood density is expected to lead to worse performance generally.

Luce and Pisoni (1998) found evidence supporting some of these predictions, though not all. In their Experiment 2, monosyllabic words and non-words were auditorily presented to subjects, who were asked to decide whether the presented syllable was an actual word by button press. Variables of interest were stimulus word frequency, neighbourhood density, and mean neighbourhood frequency. The median frequencies and densities were determined, and anything above the median was classified ‘high’, while anything below was classified as ‘low’.

High frequency words had a mean frequency of 254 occurrences per million, and low frequency words 5 occurrences per million. High density neighbourhoods had 22 neighbours, while low density neighbourhoods had 11. The mean frequency of high frequency neighbourhoods was 370 occurrences per million, and low frequency neighbourhoods 46 per million.

In general, responses to high frequency words were both faster (55ms) and more accurate than responses to low frequency words. Responses to words in high density neighbourhoods were generally slower (13.5ms) than those in low density neighbourhoods, but they were also more accurate. Words in low frequency neighbourhoods were responded to faster and more accurately than those in high frequency neighbourhoods.

An interaction was found between word frequency and neighbourhood density, with neighbourhood density having a different effect on responses based on word frequency. In low density neighbourhoods, high frequency words were responded to 10.3% more accurately than low frequency words (and 62ms faster). In high density neighbourhoods, high frequency words were only 4.5% more accurate (and 47.5ms faster). High frequency words were 21ms faster on average in low density neighbourhoods than they were in high density neighbourhoods. For low frequency words the opposite was the case. Low frequency words were responded to more accurately in high density neighbourhoods than in low density neighbourhoods. Thus, though high frequency words are both faster and more accurate than low frequency words generally, this advantage is less marked in high density neighbourhoods.

To summarise the interaction: low frequency words were found to elicit the same response times regardless of density, though they were more accurate in high density neighbourhoods; while high frequency words were equally accurate regardless of density, but faster in low density neighbourhoods.

2 Usage-Based Models of Lexical Representation

The following section provides a basic overview of lexical representation in a usage-based system, paying particular attention to the exemplar model as outlined in Bybee (2001). The key assumption of usage-based models generally is that the precise form and content of the lexicon is governed by an individual's language experience. Usage and frequency are encoded in the representation/lexical organisation. Linguistic forms enter the lexicon by dint of

frequency, and not because of membership in an abstract class (i.e. because the form is a morpheme, or a word, etc.). Repeated exposure to a certain sound pattern (or form) which is associated with a certain meaning results in that form being committed to memory.

The representation of a form expresses a range of phonetic variation, with each token of use adding an image to the representation. Higher frequency forms are weighted, and can make those variants more 'characteristic' than other less frequent forms. Identical or very similar exemplars map to the same representation, increasing its strength. Increased strength makes the form easier to access (resulting in faster response times), and less likely to undergo certain kinds of change, such as analogical levelling (e.g. the relatively high frequency alternation *keep*/*kept* is preserved, while *burn*/*burnt*/*burned* is regularising).

Analogical levelling is a case of irregular forms becoming regularised, for example the more productive pattern of past tense being applied to a verb to form the past instead of the irregular form being accessed (e.g. 'weeped' is produced instead of 'wept'). Bybee (2001) contends that this happens only in lower frequency words, when an irregular form is so infrequent that it has a very weak lexical representation. The lexical representation of the irregular form is weak, making the form both hard to access and insubstantial in making 'correctness' judgements. Compare the case of 'learn' to that of 'sleep', where the form 'sleeped' is fairly uncommon. This is due to the relatively high frequency of the form 'slept'. This form has high lexical strength, and the form is therefore easy to access (making a regular inflection unlikely), and highly influential in correctness judgements (making it effective at blocking the few regularly inflected forms that are encountered).

The productivity of paradigms is related to type frequency - patterns with higher type frequency apply more easily to new items (i.e. irregular past tense in English has high token frequency, but low type frequency, and new words tend to get regular past tense). This may be because higher type frequency contributes to greater analysability. Words (complex) are first learnt as unanalysable wholes. Repetition of the affix across forms leads to the discovery of

the affix (see Marchman and Bates, 1994). Once identified and stored, the affix and associated schema gain in strength from use.

With increasing frequency, the representation of the form grows stronger, and with repeated use the articulation process is streamlined. This makes high frequency words both easier and faster in terms of access, as well as causing them to have a greater tendency to phonetic reduction and semantic bleaching. For example, 'would have' is used frequently, and reduces to 'would've'. The connections between 'have' and '-ve' also weaken, such that it is sometimes interpreted as 'of' (Bybee, 2001). High lexical frequency also weakens the connections with other related forms, allowing them to become associated with non-related forms (for example 'go' and 'went'). This is of potential interest with regard to Mandarin - is it the case that, due to this disassociation effect, words of exceptionally high frequency are treated qualitatively different than the rest of the (lower frequency) homophones? By this interpretation, words of very high lexical frequency are likely to differ the most phonetically in a given homophone cohort, with lower frequency forms phonetically closer to each other.

In a strong version of the exemplar model, the phonetic range of a form is exhaustive of encountered tokens. Each token, along with its contextual information, is committed to memory. Contextual factors which may influence the phonetic form of a word can include - phrase position; prosody; the register or situation in which a particular token occurs (e.g. 'I dunno' occurring at an informal family gathering, and 'I do not know' in a public speech); the immediate phonetic context (such as the word 'is' with the form [rɪz] immediately after a word ending in an orthographic 'r' like 'car' [kaː]); and the frequency of surrounding words, among others (Bybee, 2001). In a 'prototype model' each token and its context is summarised, and contributes to a prototype of the form.

Words, both simple and complex, are the most common unit having lexical representations. This is due to their high combinability, and the fact that they can occur in isolation. High frequency phrases (such as 'don't') can also have 'lexical' representations,

however (Bybee, 2001). Bound morphemes (i.e. '-ness') likely do not have independent lexical representations, despite their potentially high productivity, as they cannot occur in isolation. They exist in the lexicon as part of the representations of the words they occur in. If the separate occurrences of a particular morpheme are identified as having sufficient phonetic and semantic similarity (i.e. they are identified as being 'the same' in these senses), lexical connections form between the words containing the same morphemes.

A similar relationship exists between the form of a free morpheme that occurs in isolation (e.g. 'opaque'), and other forms which occur in conjunction with other morphemes (e.g. 'opacity'). If the various forms are, or are a component of, words of sufficient frequency to maintain independent lexical representations, and the forms (in this case [op^{h}eɪk] and [op^{h}æs]) are identified as 'the same', lexical connections will form between the isolation form and the words in which the other forms occur. Low frequency forms (or novel formations) are created by using high frequency schemata.

In some sense, this constitutes a redefinition of what it means to be a word. A form does not achieve word status through belonging to an abstract pre-existing or universal linguistic category. Words are, rather, just those forms which are of sufficient frequency and 'combinability' to have lexical representations. In this sense words are an 'emergent' category, determined not by pre-existing dispositions and structures, but arising out of patterns of use. Other categories of linguistic structure such as syllables, segments, and so forth, can also be seen as being emergent in a similar way to words. The lexicon is not reminiscent of a dictionary, with words composed of discrete segments arranged in a list. It is rather a network of words and high frequency phrases, with associations between related words, morphemes, segments, and so forth - the result of the categorisation of perceived and produced tokens of words.

It has been noted that in Mandarin, a large number of 'words' consist of two-morpheme compounds, and there has been disagreement over what sized units are used in

lexical recognition and access - morphemes, or words consisting of one or more morphemes. In general, the morphemes that combine to form two-character compounds are not bound, and occur in isolation under certain conditions. Following Bybee (2001), it may be that such morphemes have independent representations, and that those high frequency compounds in which the morpheme occurs also have separate representations, with links between them. It may even be the case that the fairly extensive list of four-character idioms (*chengyu*) have networked lexical representations.

Whereas Bybee's (2001) exemplar model attributes the existence and form of a mental representation for a word to its frequency of use, more abstract theories assume separate storage and processing mechanisms for different categories of words independently of usage. For words with irregular inflection (e.g. 'sing' and 'sang'), the assumption is similar in outline if not in detail, in that all of the forms of the irregular verbs are stored in memory. For the regular verbs however, the assumption is that some type of base form is all that is stored in memory (e.g. 'jump'), and all other forms are generated by symbolic rule as needed ('jump' + '-ed' = 'jumped'). The implication of this kind of model is that regularly inflected verbs should have slower response times than irregulars as they require processing, when controlled for frequency.

There is however evidence that lexical storage is a function of lexical frequency, and not of word class. Bybee (2001) notes that the irregular verbs as a class are generally of high frequency, while the regular verbs are of variable frequency, which may explain why some studies have found an effect of inflection-type on response times - it is not inflection type that is affecting response times, but frequency. Usage-based theories also provide an explanation for why irregular verbs would tend to be of relatively high frequency, as high frequency words seem to be resistant to analogical levelling, and therefore do not regularise.

There have been a number of studies that have noted that, when controlled for frequency, the apparent effect of inflection-type disappears. For example, Sereno and

Jongman (1997) found that the response time in a lexical decision task was related only to the independent lexical frequency of a given form. The low frequency verb 'opened' was found to have a relatively slow response time, despite the fact that the verb 'open' is of relatively high frequency and had fast response times in the task. Verbs with high frequency past tense forms were found to have quick response times, regardless of the frequency of the base form. If the regular past tense form was derived by rule, and had no independent representation in memory, we would expect the frequency of the base, potentially in addition to some lag caused by applying a rule, to determine response time. Instead, it is only the frequency of a particular form that appears to impact this metric, leading us to suspect that the regular forms are stored in memory, just as the irregulars are.

3 The Bayesian Reader

In more recent years, other models have been proposed to explain the complex behaviours found in lexical access and recognition, and incorporate advances in knowledge about the effects of lexical frequency and how speakers interact with this knowledge into a more fully worked-out model. One approach has been to model human readers/listeners along the lines of an idealised Bayesian decision maker (Norris, 2006).

The core assumption of Norris (2006) is that human readers/listeners behave as idealised Bayesian decision makers. Essentially, this approach asks, 'How would an ideal observer respond to ambiguous language input?,' and assumes that the answer to this question will correspond, at least in the main, with the behaviour of humans. If the input were unambiguous, an ideal observer would just match the input to the identical lexical representation, and factors such as lexical frequency and neighbourhood density would not be expected to affect the behaviour of the system. If the input is at all ambiguous, however, this simple matching process is no longer reliable, and prior probabilities of words (i.e. lexical frequency) become potentially useful in the word recognition process (Norris, 2006).

Bayesian inference, as the optimal combination of perceptual information with knowledge of prior probability, is a strong candidate for modelling how knowledge of lexical frequency is combined with perception to produce lexical recognition (Norris, 2006). In effect, Bayes' theorem combines knowledge of word frequency (i.e. prior probability) with the probability that the input was generated by a particular word, and gives a posterior probability for each word given the evidence. Specific inputs could be produced by different words, and the likelihood of generating a specific word comes from summing the probabilities that each word generated the particular input.

If the reader/listener is confronted with a noisy representation of a word, either because there is noise in the signal, or because the input is noisy, there is a possibility for a mismatch between the intended word and the signal. If the word most closely matching the input is of very low frequency, it is possible that a mistake in recognition has occurred, and that a word that matches less well (but is of higher frequency) is the actual intended target. Other factors will affect the weighting of the input, such as syntactic and semantic fit with the rest of the utterance at hand (Norris, 2006).

The nature of the frequency effect in a Bayesian model is of critical importance. Like other models it acts as a bias, but unlike other models this bias is mitigated by input. Norris (2006) notes that in Logogen (Morton, 1969), the response bias is independent of the quantity and quality of available perceptual information - higher frequency words require fewer features to exceed threshold than lower frequency words. This makes higher frequency words easier to recognise, but at the expense of making lower frequency words *more difficult* to recognise. As there are more low frequency words than higher frequency words (Zipf's law), this results in an overall inefficient system. Though a blatantly inefficient language recognition system is possible, it should raise skepticism: why would an evolved system preserve gross inefficiency at no benefit?. We should rather expect that we have made a false assumption somewhere, and that lexical frequency (which we know affects recognition) is

utilised in the perceptual system because it provides some net *benefit*, rather than a cost. A Bayesian decision maker is biased toward words with higher prior probabilities (e.g. higher lexical frequency) but combines this with the present probability that the given input was generated by these words. As the probability that a particular word was generated by the input increases, the impact of frequency on decision making decreases.

The previous outline gives a sense of how an optimal recogniser utilises frequency, but Norris (2006) goes on to give a demonstration of a Bayesian model of visual word recognition, which he calls the Bayesian reader. The model assumes that words can be represented as points in a 'multidimensional perceptual space'. Perceptual evidence is accumulated over time through sampling from a distribution around the perceptual coordinates of the input, with the mean location and standard error of all samples calculated as the process progresses. Over time the likelihood that the input was generated by a particular word increases, and an identification can be made. Words with higher prior probability take less time to reach this point, resulting in the well-attested frequency effect.

As a result of the assumption that human readers approximate Bayesian decision makers, the model also predicts a number of other patterns of results for different perceptual tasks related to neighbourhood density that match numerous findings in lexical access and recognition research. For neighbourhood density, it has been found that increasing density makes word identification more difficult, while making lexical decision tasks easier (Norris, 2006). The Bayesian reader predicts that high-density words will be more difficult to identify, as the given input is more likely to have been generated by closer neighbours (and particularly high frequency neighbours) than more distant neighbours. This means that the perceptual system will require more time (more 'samples') to centre on the intended word.

The Bayesian reader also predicts that words in dense neighbourhoods will be easier to classify. This is because an input distribution which corresponds with a region of perceptual space dense with words is *more likely to be a word* than one which corresponds

with a sparsely populated region. The denser a region is, the less likely it is for a non-word to fit into the remaining space. The model also predicts an interaction between frequency and neighbourhood density, with low frequency words receiving greater facilitation due to residing in high density neighbourhoods than high frequency words (Norris, 2006). This is because the increased probability of a given word existing, which derives from the presence of many words in close proximity in perceptual space (i.e. the 'dense cloud' effect), will have a relatively greater effect on words with independently lower probability (low frequency).

To test this prediction, Norris (2006) simulated two lexical decision experiments originally carried out by Andrews (1989, 1992). The Bayesian reader, in agreement with Andrews (1989, 1992), replicated the interaction between frequency and neighbourhood density. Norris (2006) also simulated an experiment by Snodgrass and Mintzer (1993) which used the Andrews (1989) items in a perceptual identification task. Like Snodgrass and Mintzer (1993), the Bayesian reader produced an inhibitory effect of neighbourhood density in identification, for the same items which produced facilitation in lexical decision.

Applications for the study of Mandarin Homophony

A simulation of Bayesian decision making in Mandarin is beyond the scope of this work, but we can speculate as to what a Bayesian model would predict for a language with lexical tone and a high degree of homophony. A language like English has a relatively large pool of syllables (more than eight times as many as Mandarin), and considerably greater freedom in morpheme composition - polysyllabic morphemes abound in English, and are quite rare in Mandarin.

One of the more critical differences between Mandarin and a language such as English, then, is the distribution of words in perceptual space. The perceptual space of English is larger, with more distinct regions of density (i.e. has more internal differentiation). Our assumptions about the nature of the perceptual space in Mandarin will depend on our

assumptions regarding the type of 'unit' that exists in that perceptual space. Do all morphemes have coordinates in perceptual space, or only 'words', many of which are disyllabic? Or, is the composition of perceptual space due not necessarily to linguistic category, but an emergent class of 'combinable units' (as in exemplar theory)?[3] Regardless of the answer to this question, the perceptual space of Mandarin is inevitably smaller and denser than for a language like English. If we take morphemes as a starting point, then all of the many morphemes in Mandarin are concentrated in only some 400 syllables, disregarding tone. That is to say that there are only around 400 'clouds' in the perceptual space. Within these clouds are sub-clouds centred on the four lexical tones (and presumably these contain tonal variations, such as tone sandhi). Finally, within each tonally-specified sub-cloud (e.g. *wan3*), there are anywhere from one to dozens of homophones, potentially subdivided even further by frequency (with high frequency homophones have shorter durations than low frequency homophones, for example).

Applying the principles of the Bayesian reader to this dense perceptual space, then, we should expect to find a spectrum of effects relating to the interaction between frequency and 'neighbourhood' density. Morphemes (or words) which have more neighbours closer in perceptual space will experience greater facilitation in tasks such as lexical decision, and inhibition in identification tasks. For example, we expect morphemes with many full homophones and many 'segmental' homophones to experience greater facilitation in lexical decision tasks than those with many full homophones and few segmental homophones, and both of these will experience greater facilitation than those with few full homophones and many segmental homophones. The Bayesian approach, then, is complementary with our

[3]Which linguistic items receive coordinates in perceptual space is a question of critical importance for our expectations in a Bayesian model, and in models of speech perception generally. This thesis uses morphemes as the basis for investigation throughout, partly as a means to simplify a number of complex issues in Chinese wordhood generally, though given the high salience of morphemes in Mandarin and their relatively high degree of combinability it is likely that many Mandarin morphemes are independently represented in the lexicon. Still, a more accurate account of what is, and what is not represented in perceptual space is required for a thorough understanding of Mandarin speech perception.

previous suppositions that were based on exemplar theory, and provides a computational framework in which to model how overlapping exemplar 'clouds' work in practice.

4 Models of Processing Homophony

In a Cohort-type model, a stimulus is encountered, and the cohort of words is whittled down such that only one phonemic representation remains (or alternatively, only one representation is highly activated, while competitors are heavily damped). For a language such as English, this is presumed to be the end of the word recognition process in most cases. In the minority of cases where there are two or more words associated with a single phonemic representation, some further winnowing is required from top-down contextual information. For Mandarin, however, we would expect this 'homophone decision' issue to occur for nearly every syllable - and for some syllables the degree of homophony grossly exceeds anything encountered in European languages (if the word identification system successfully recognises *yì*, for example, we are still left with around 25 words to choose from). The way in which homophones are processed is of considerably more pressing concern for a language such as Mandarin.

A number of proposals have been made as to the nature of the 'homophone decision' process. These models are primarily based on European languages, and rarely contain specific reference to two Mandarin features, tone and extensive homophony, which we might reasonably expect to act as factors in homophone decision. Several models are outlined in this section, and evaluated with reference to Mandarin. First, it is necessary to clarify what specific questions are being asked of a model of lexical structure and recognition as it applies to Mandarin homophony. The questions fit into two basic categories: the nature of lexical representation; and the role of neighbourhoods.

1. On what level are two words homophonous? Is it in identity of abstract phonological representation? Does this representation include the encoding of usage/frequency? What about phonetic detail? Research by Gahl (2008) has indicated that in English, words conventionally considered to be homophonous but distinguished by lexical frequency (e.g. *time*/*thyme*) are distinct in production, particularly word duration. A usage-based model would assume these two words have independent lexical representations, and are in some sense not homophones. The Neighbourhood Activation model explicitly assumes that frequency is not encoded in the phonetic/phonological representation, but acts as a top-down 'bias' on word decision units (e.g. homophones have the same representation, but differing biases). What about tone? Is it integrated with the phonetic/phonological representation in the same way a phoneme is? A number of studies have found that tone is not equivalent to segmental information in lexical processing tasks such as naming, same-different judgements, etc., in either speed or accuracy (Taft and Chen, 1992; Ye and Connine, 1999; Cutler and Chen, 1997; Lee, 2007). Ye and Connine (1999) posit a separate toneme tier in a TRACE architecture, in addition to the feature, phoneme, and lexical tiers. Is this acceptable? How ought tone be modelled in a usage-based framework? In short, where is homophony, and what is homophony?

2. Does homophone density matter? The Neighbourhood Activation model predicts that the number and frequency of members in the similarity neighbourhood of a word will affect both speed and accuracy in word identification. If this is the case, then we might expect that the number and frequency of homophones will also affect word identification. Does the size and composition of the homophone neighbourhood have greater influence over word identification than a similarity neighbourhood? What about the 'segmental homophone' neighbourhood? Mandarin, having far fewer unique words and many more homophones than English, should be expected to have correspondingly larger homophone and similarity

neighbourhoods than English. Should we expect the role of neighbourhood density to be similar to English?

4.1 Context-Guided Single-Reading Model

The 'context-guided single-reading' model (Simpson, 1981) assumes that the context preceding an ambiguous word determines what meaning is accessed, and that only one meaning is ever accessed. There are a number of problems with such a theory, but perhaps the most critical problems occur when an ambiguous word occurs as the first word in a sentence. According to the model only one meaning is ever accessed and no more than one reading of a word is allowed, thus forcing an immediate choice with 1) no guiding information, and 2) no possibility of changing the interpretation once further context has been established.

The theory clearly predicts that we will be unable to correctly interpret ambiguous words occurring prior to any contextual disambiguation, except perhaps by fortunate accident. This prediction is correct up to a point - by definition, an interpretation of a word devoid of context will be a guess. The theory fails, however, in that it does not allow for reinterpretation. Even in a language such as English, which has only minimal homophony, this is not an ideal result. In Mandarin, in which every morpheme has an average of four full homophones (and some morphemes have considerably more than this), the single-reading stipulation is egregious. The first word of nearly every utterance in Mandarin would be interpreted incorrectly, but this is by no means the only failure. Without the first word, there is no reliable context upon which to base a reading of the second word, which is also likely to have a number of homophones. By the 'context-guided single-reading' model, communication in Mandarin is predicted to be theoretically impossible. On the assumption that this is incorrect, the model must be abandoned.

4.2 Ordered-Access Model

A second type of model, 'ordered-access', assumes that when an ambiguous word is encountered, the meanings of that word are automatically accessed in order of lexical frequency. Essentially, the most frequent meaning is checked against the context and is either accepted (so the composition of the homophone neighbourhood should be irrelevant), in which case the process is finished, or it is rejected, in which case the next most frequent meaning is checked against the context. This process is continued until the correct meaning is judged to have been reached.

A version of this model, called the 'multiple-access' model (Swinney, 1979), has the feature that all of the meanings of an ambiguous word are initially activated simultaneously (i.e. access is 'exhaustive' and autonomous). Context is then utilised to pick the contextually relevant meaning, leaving it as the only active interpretation. According to Swinney (1979), it is the speed of this 'post-access' decision process which is affected by the strength of the biasing context and the lexical frequency of the meanings, not that such differences in response times are due to the number of meanings which are being accessed.

We would expect to find priming effects for lower-frequency meanings of ambiguous words, even in contexts in which a higher frequency meaning is strongly indicated (and selected). This is because all of the meanings are activated in all circumstances. More importantly, this model makes strong predictions about the effect of homophone density on lexical access. Because the time-course of the 'post-access' decision process is only affected by context and the lexical frequency of the words, the model predicts that homophone density will not affect lexical access - having more homophones will not speed up (through some sort of frequency inheritance effect) or slow down (through competing for activation) the process, as this is determined solely by context and lexical frequency.

This prediction is possibly counter-indicated by the finding that neighbourhood density can affect the time-course and accuracy of lexical identification (Luce and Pisoni,

1998). A parallel processing model such as the Neighbourhood Activation model should predict an effect of homophone density. An ordered-access model should also predict an effect of homophone density for lower frequency words, as the more words in the list preceding the target, which must be assessed and rejected, presumably the longer the process.

There is some evidence that ambiguous words take longer to process than unambiguous words in English, even when a particular meaning was strongly indicated by context, potentially providing support for 'exhaustive' access of all of the meanings of an ambiguous word (Foss, 1970). Such a conclusion is disputed however, particularly by Hogaboam and Perfetti (1975). Hogaboam and Perfetti (1975) presented ambiguous words (each with a high and low frequency meaning) at the end of strongly biasing context sentences. Participants were asked to respond if this word was ambiguous. Participants *detected ambiguity* faster for low-frequency meanings than high-frequency meanings. This was interpreted as a form of ordered access - meanings are accessed sequentially in order of lexical frequency, until a meaning is reached that fits the context. The reason responses are faster to low frequency meanings is because higher frequency meanings have already been accessed, making ambiguity detection a rapid process. For high frequency meanings the process terminates once the first meaning is accessed, and under normal circumstances lower frequency meanings would never be accessed, making ambiguity detection significantly slower.

The low-frequency ambiguity detection effect found by Hogaboam and Perfetti (1975) is not necessarily inconsistent with an exhaustive access account, however. Take for example two ambiguous meanings of a phonological form *yì*, one high frequency and one low frequency (see Figure 4.1). Both are activated upon recognition of the string *yì*, occurring at the end of a sentence strongly biased to the meaning of the low frequency character. By the time the lower frequency word is recognised, the higher frequency word has also been recognised. This accounts for the low frequency bias in ambiguity detection. When the higher

frequency word is the intended target, the process is very similar. Both are activated simultaneously, and the higher frequency word is recognised relatively quickly. As this word fits the context, however, the process stops, and the lower frequency word is never recognised.

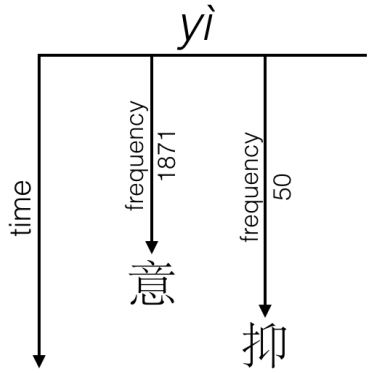


Figure 4.1. Timing of lexical access for two full homophones of *yì*, as affected by lexical frequency.

It appears difficult to ascertain whether access is 'exhaustive' or 'ordered' in languages with homophony comparable to English, where a particular phonological form is associated with at most only a few meanings. When there are only two or three words associated with a single phonological form, there may simply not be a large enough difference in density for a meaningful density effect. It may be possible to distinguish between exhaustive and ordered access for Mandarin however, where *yi4*, for example, actually has twenty-five homophones and not two (see Figure 4.2).

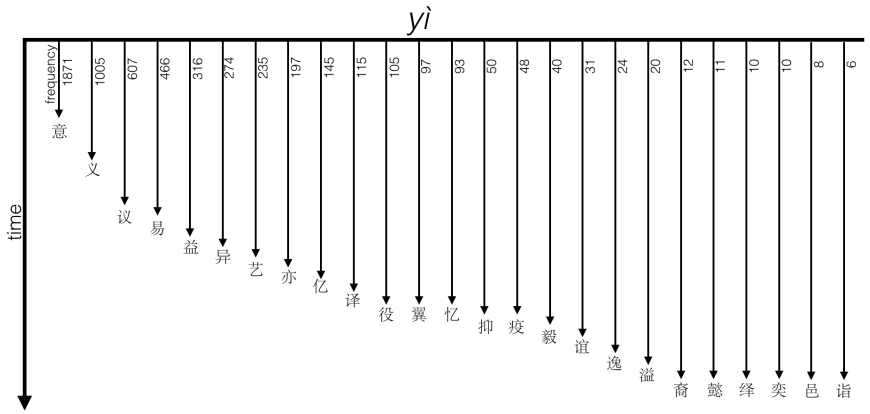


Figure 4.2. Artificially-smoothed time required for lexical identification for all full homophones of *yì*, as affected by lexical frequency.

Taking again a sentence biased toward the highest frequency *yì* character 意, in both the exhaustive and ordered-access models, we expect the time-course for lexical access to be the same for situations like those exhibited in Figures 4.1 and 4.2, whether the full homophone density is 2 or 25.

For a context biased to the lowest frequency character 诣, the predictions ought to be radically different between the two models. For an exhaustive access model, we might expect the time it takes for this character to be accessed to simply be the time it takes for any character of this frequency in this sort of biasing context. For an ordered-access model, this is decidedly not the case. There are 24 characters of higher frequency than the target, each of which must be activated and rejected in turn. This presumably takes some time, which we may be able to measure experimentally.

If, as the size of a homophone cohort increases processing time also increases, there are potentially serious implications for the processing speed of Mandarin, in which every morpheme/syllable in every sentence is potentially ambiguous. Furthermore, depending on

the way lexical tone impacts lexical access, the number of activated words may actually be even higher (around 49 meanings for *yi* alone).

The difference in the degree of ambiguity between a language like English and Mandarin, in terms of the number of morphemes potentially requiring access over a short sentence, is well served by a fairly typical example. In an English sentence of five words, with an ambiguous word, such as "This is a river bank," we posit the activation of a minimum of six meanings (bank having two). For the Mandarin equivalent of this sentence, "*zhè shì yī gè hé àn,"* there are an astounding //*zhè* [2(5)] + *shì* [25(45)] + *yī* [5(49)] + *gè* [2(16)] + *hé* [6(11)] + *àn* [5(10)] = 45(136)// meanings (number of segmental homophone meanings denoted in brackets). Precisely how homophone density affects processing will dramatically affect the amount of time a typical sentence like this will require - from what we might call a 'normal' amount of time based solely on the lexical frequencies of the individual characters, to potentially a very long time in an ordered-access model, depending upon the size and frequency ranking of the characters.

A third option is the 'context-dependency' hypothesis, which posits that it is possible for context to constrain which meanings of an ambiguous word are activated, i.e. there is not mandatory blind access when the context is clearly biased towards a particular meaning. Li and Yip (1996) have found such a context effect in Cantonese. When a particular meaning was biased, context effects were found to begin immediately, before the entire word had even been pronounced. When there was a high level of contextual bias, frequency effects were weakened, and high and low frequency words were accessed at similar speeds. In low levels of context, frequency was found to be the main factor, and higher frequency words were accessed much faster than lower frequency words. Interestingly, no effect was found of homophone density, though this was possibly due to admitted flaws in the experimental design (Li and Yip, 1996). Such findings support the context-dependency hypothesis, in that

context affects lexical activation immediately, constraining activation from a very early point (i.e. even before the offset of the syllable).

The context-sensitive model predicts that both the relative frequencies of the available meanings, as well as the directionality and strength of the biasing context interact, influencing lexical processing, such that the most likely meaning for a word in a given context is selected (Vu et al, 1998). The model predicts that context will be the dominant factor in disambiguation, overriding frequency when context is strongly biased to a particular meaning. When context is 'neutral', or only weakly biasing, the higher frequency meanings will be chosen.

4.3 Usage-Based Models and Homophony

Relatively less work has been done on the way in which a usage-based model might treat homophone processing. As discussed in **Chapter III**, an investigation by Gahl (2008) into the production of high and low frequency homophone pairs in English found that higher frequency homophones were consistently shorter than lower frequency mates. From a usage-based perspective, we might assume that these two words do not have the same representation, and are not strictly homophonous. It is unknown whether a distinction in duration such as this is utilised in perception, however, and whether such pairs are processed as homophones, or merely as near homophones. And while it is at least possible that, for a language such as English, conventionally-defined homophone pairs such as *thyme/time* are distinguishable in perception, the extent of homophony in Mandarin makes it extremely unlikely that each word in a homophone cohort of two dozen words is distinct in terms of duration, or any other phonetic variable.

Given Gahl's (2008) findings, it is at least possible that *some* of the conventionally-defined homophones in Mandarin are phonetically distinct from each other in production -

perhaps high frequency homophones have distinctly shorter durations than the general mass of lower frequency homophones, which share a generalised representation.

5 Summary

Two categories of questions were identified at the start of this chapter as being of particular interest for homophone processing in Mandarin: 1) what is homophony, as applied to Mandarin; and 2) does density matter?

Abstract phonological treatments of tone tend to classify tone distinctions in Mandarin as equivalent to phoneme distinctions in importance (see Ye and Connine, 1999). There is evidence, however, that tone is not as important to lexical identity as segmental information (Taft and Chen, 1992; Ye and Connine, 1999; Cutler and Chen, 1997; Lee, 2007). Further, there is evidence that neighbourhood density and frequency affect the speed and accuracy of lexical identification (Luce and Pisoni, 1998). Homophone density is not the same as neighbourhood density, but we may reasonably suspect that similar effects may emerge from the competition amongst a homophone neighbourhood as have been found with similarity neighbourhoods. Finally, there is evidence that words conventionally considered to be homophones can be distinct in production, particularly in terms of duration, as a function of frequency (Gahl, 2008).

Bringing these three strands together, we predict that the lexical identification process will be affected by full homophone and segmental homophone density in similar ways as it has been found to be affected by similarity neighbourhoods. Tone is not assumed to be unimportant, but is predicted to be less critical in determining phonetic distance than segmental-type information. There is therefore expected to be a continuum of similarity, from full homophones of similar frequencies (near phonetic identity), to full homophones distinguished by frequency, homophones distinguished prosodically, and through to phonemic neighbours. Along this spectrum, the closer to phonetic identity a set of words is, the greater

amount of evidence will be required to choose the correct candidate, i.e. the greater amount of time (and potentially less accurate) lexical identification will require.

The second category of questions has to do with density - is having higher full homophone density beneficial, harmful, or neutral (or some combination, depending on frequency and density interaction)? There are two aspects to this - 1) locus of frequency effects; and 2) type of homophone processing model.

For any model assuming an abstract phonological form is the locus of frequency (i.e. all homophones share one phonological representation), increasing homophone density should have a positive effect on access (frequency inheritance). This should be most noticeable in low frequency words, which benefit most from the cumulative frequency of neighbours (high frequency words are already high in frequency, and increase in cumulative frequency only marginally by adding low frequency neighbours). We should find clear differences in response times for cumulative frequency of all full homophones. Precisely how this position interfaces with choosing a homophone requires thought - all homophones have effectively the same frequency, so presumably context is completely determinative. Exhaustive-access would imply all meanings are activated simultaneously, as frequency is not a factor.

If frequency is associated with each meaning independently (compatible with usage-based models, though also compatible with abstract representational models where frequency is associated with the lemma), we expect specific-character frequency to be the primary determinant of response times in lexical access tasks, and not cumulative homophone frequency. We expect responses to high frequency characters to be quick, and responses to low frequency characters to be slow.

The effect of homophone density is expected to relate to the structural tendencies of homophony found in **Chapter III**, where it was noted that there is a marked pattern of homophone clustering as the number of tones associated with a base syllable increases (i.e.

the more tonal neighbours, the more full homophones associated with each tone). With this in mind, we expect to find some benefit for characters occurring in highly dense syllables (those with many full and segmental homophones).

There is controversy in the literature over whether we should expect response times to homophones to reflect either specific-character frequency or cumulative homophone frequency, and indeed conflicting results for very similar experiments (see Jescheniak and Levelt, 1994; Caramazza et al, 2001; Jescheniak et al, 2003).

In their Experiment 6, Jescheniak and Levelt (1994) looked at production timing for Dutch-English bilinguals, who were asked to give Dutch equivalents for English words, which were either: low frequency words (with a high frequency homophone); or words without homophones. Non-homophones were chosen so that their frequency matched either the specific-character or the cumulative frequency of words with homophones. They found that low frequency words with high frequency homophones patterned like words of the cumulative frequency of the two homophones. High frequency words with low frequency homophones were not tested, but it is assumed that they would pattern like words of slightly higher frequency (given the small contribution of the low frequency homophone to cumulative frequency). In their Experiment 1, Jescheniak et al (2003) exactly replicated Experiment 6 of Jescheniak and Levelt (1994); in their Experiment 2, they replicated Experiment 6 of Jescheniak and Levelt (1994), but substituted German for Dutch. In both experiments, they found that low frequency words with high frequency homophones patterned like words of the cumulative frequency of the two homophones.

In their Experiments 1A and 2A, Caramazza et al (2001) presented English and Mandarin subjects with a picture naming task. The name of the picture either had homophones or not. Pictures whose names did not have homophones were matched for frequency with the word-specific frequency and cumulative homophone frequency of the homophonic picture names. Pictures were seen by subjects three times. Caramazza et al

(2001) found, for both English and Mandarin, that pictures were named according to their word-specific frequency, and not their cumulative homophone frequency (i.e. no frequency inheritance effect was found). Repetition was not found to affect speed of responses.

It is difficult to adjudicate between these two positions. It is important to note, however, that for the foregone experiments, latency in *production* was being measured. As Jescheniak and Levelt (1994) point out, the frequency inheritance effect found in production is not found when subjects are asked to recognise objects, but not name them. We should, therefore, expect different effects in production tasks and perception tasks.

In order to address some of these issues, a series of experiments was designed to test the predictions put forward in this chapter. These experiments are outlined in **Chapters V**, **VI**, and **VII**.

Chapter V. Experiment 1

Two categories of questions were identified in **Chapter IV** as being of particular interest for homophone processing in Mandarin: 1) what is homophony as applied to Mandarin; and 2) what is the role of homophone density? The purpose of Experiment 1 is to survey response patterns to ambiguous Mandarin syllables, with particular attention on any possible effects of full homophone density on responses. A better understanding of the role of tone and full homophone density in lexical access is vital for understanding the structure of the Mandarin lexicon, and ultimately a more complete understanding of homophony in Mandarin.

Experiment 1 was designed to function as a way to gather preliminary data on the hypothesised effect of homophone density on lexical access in Mandarin. Should density appear to be influencing the response behaviour of native speakers, the experiment will provide valuable information for constructing more extensive experiments, to explore any potential effects in greater detail. As such, a simple experiment was designed, in which toneless *pinyin* monosyllables (e.g. *miao*) were presented to experimental participants, and single character responses were elicited. Experiment 1 was conducted as one part of a series of eight experiments conducted during a single session for the MPhil. Only the first section is reported in this chapter.

It is important to mention here why we have decided to use monosyllables as the focus of study, rather than another unit such as the word. There are inherent difficulties in using terminology like 'lemma' and 'word' devised primarily with respect to European languages for Mandarin (and Chinese languages generally). These difficulties are of both a theoretical and a historical/practical nature.

In terms of how speakers of Mandarin themselves refer to meaningful units, historically the term 词 *cí* 'word' has not been in common use, either colloquially or in scholarship, and the nearest equivalent is the character/morpheme 字 *zì*, which is

monosyllabic (Chao, 1968a). Literate speakers are very much aware of characters/morphemes - morphemes are learned individually, with an associated meaning, character, and pronunciation, unlike in English.

While morphemes are clearly signalled in Chinese writing through characters, what might be thought of as words are not themselves delineated (as they are by spaces in European and other writing systems). Word boundaries are also not reliably signalled by phonetic markers in the speech signal (Duanmu, 2007). Judgements of word delineation have been found to be variable in speakers of Chinese languages, such as Cantonese (Hoosain, 1992) and Mandarin (Sproat et al, 1996). Duanmu (1998) rules out native speaker intuition as an effective method of determining wordhood in Mandarin, as the results are unacceptably variable.

That (monosyllabic) characters, and not words, are the unit of common discussion (and even of modern scholarship) may be historical accident, but is in itself potentially interesting. At the least, the character/morpheme is of high salience for literate Mandarin speakers - certainly more than the word, which is relatively vague to native speakers, and far more than the salience of morphemes to native speakers of English, for example.

This is not to say that words do not exist in Mandarin, and there are many compound words which are largely unambiguous. Indeed, in the same work in which Duanmu (1998) rules out native speaker intuition as an unreliable guide, compelling arguments are made for other criteria (e.g. semantic reduction, freedom of parts, etc.) determining word status. We would simply contend that the category 'word' has not been used in historical scholarship, is not utilised in colloquial discussions of word-like units (characters/morphemes being preferred), is not reliably accessible as a concept even to native speakers, and has eluded successful academic delineation, at least until quite recently. Many compounds will be categorised as words by most scholars, perhaps, but there remains theoretical debate.

Avoiding using compounds in our experimental materials avoids this controversy, at the risk of using morphemes in isolation which do not have independent lexical status.

Practically speaking, as has been noted, the number of disyllabic words varies with register (among other potential factors, such as metrical considerations). Further, it is often difficult to distinguish between varieties of Chinese in writing (as they use mostly the same characters), and the varieties differ in terms of which morphemes occur as monosyllabic words. Word (or character) choice also differs by variety - though in excluding low frequency characters of the corpus from consideration, 'dialectal' characters should be largely excluded, and in using only participants with very similar language backgrounds (e.g. Mandarin speakers from Beijing with Beijing parents), at least the responses to characters are expected to be consistent. In a large written corpus such as the one used in this thesis (Da, 2007), word counts may be more unreliable than character counts.

Theoretically, it is perhaps not clear which particular characters/morphemes have independent lexical representations. It has been noted that in Mandarin, a large number of 'words' consist of two-morpheme compounds, and there has been disagreement over what sized units are used in lexical recognition and access - morphemes, or 'words' consisting of one or more morphemes. In general, the morphemes that combine to form two-character compounds in Mandarin are not bound, and occur in isolation under certain conditions. Following Bybee (2001), it may be that such morphemes have independent representations, and that those high frequency compounds in which the morpheme occurs also have separate representations, with links between them. It may even be the case that some types of longer strings, like the fairly extensive list of four-character idioms (*chengyu*), have networked lexical representations. If we assume that a disyllabic (bimorphemic) word has an independent representation, and its constituents have independent representations as well, *and* that the constituents are linked to the compound in some manner (Bybee, 2001), then the

frequencies of all entities may matter. This is in addition to whatever effect the homophone density and phonological neighbourhood density have.

As a further practical consideration, this thesis is interested in homophony. Disyllabic words have far fewer 'homophones' (an interesting point in itself), making it difficult to examine the effects of homophone density for morphological compounds. It was very difficult to create stimuli for Experiment 3, in which the two target (homophonic) characters were surrounded by the same syllables - in many cases, part of speech had to be altered for this purpose. It would be very difficult indeed to create such sentences with disyllabic homophones, of the same part of speech, surrounded by the same syllables, in the same intonational position.

A thorough account of the lexical status of compounded characters is undoubtedly needed for a full understanding of Mandarin lexical organisation and homophony. There are, as outlined above, a number of complications involved in giving this account. A better understanding of characters/morphemes in isolation will provide a necessary foundational step, as it is likely that frequency effects of any compounded units are to some degree influenced by frequency effects of meaningful parts.

1 Method

1.1 Stimuli

Five Mandarin syllables were selected as stimuli - *miao*, *pang*, *wan*, *shi*, and *yi*. These syllables were chosen as they represent a range of Mandarin syllable types, both in terms of frequency and homophone density. The five syllables are briefly described below.

miao

The syllable *miao* is of low general frequency, with a median frequency for all *miao* characters at around 14 occurrences per million, and a sum phonological frequency (i.e. all

miao characters regardless of tone) of 367 occurrences per million. Despite having characters for all four lexical tones, it is also of fairly low density, with only 12 total characters in the modified corpus. The two highest frequency *miao* characters are 描 (*miáo*) and 妙 (*miào*), both at around 105 occurrences per million. No other *miao* character exceeds a frequency of 42. These two characters are also close in terms of full homophone density, with three *miáo* characters and two *miào* characters in the corpus. If increasing full homophone density has a positive effect on homophone disambiguation, then there is a possibility for a marginal advantage for 描.

pang

The median frequency of *pang* characters is very low at only 5.4 occurrences per million, with a sum phonological frequency (i.e. all *pang* characters regardless of tone) of 271 occurrences per million. It is also a low density syllable, though all four lexical tones are represented, there being only 8 *pang* characters in the modified corpus. The highest frequency character is 旁 *páng*, at 168. Unlike *miao*, where the two highest frequency characters are almost identical, 旁 stands out from its segmental homophones in terms of frequency, with no *pang* character exceeding a frequency of 45. It is also notable that the syllable (*páng*) is also the highest density, at 4, with the remaining *pang* syllables have densities of 2, 1, and 1. Given that 旁 stands out from other *pang* syllables both in terms of frequency and in full homophone density, we expect it to be strongly favoured in responses.

wan

The syllable *wan* is of middle-to-low density, with 17 characters, and middle-to-low frequency, with a median frequency for all *wan* characters of 38 occurrences per million and a sum phonological frequency (i.e. all *wan* characters regardless of tone) of 2359. Two *wan* characters stand out in terms of frequency: 完 *wán* 743.31, and 万 *wàn* 697.09. These

characters differ in full homophone density, with the higher frequency character at 2, and the lower frequency character at 4. No other *wan* character exceeds a frequency of 325. Given that the two highest frequency characters for *wan* are fairly similar in terms of frequency, but differ in homophone density, we may gain insight into the relative weighting of frequency and density in disambiguation.

shi

The syllable *shi* is of high frequency, with a median frequency for all *shi* characters of 288 occurrences per million, and a sum phonological frequency (i.e. all *shi* characters regardless of tone) of 36695. It is also of high density - there are 50 *shi* characters in the modified corpus). The highest frequency character is 是 *shì*, at 13584. No other *shi* character exceeds a frequency of 4331. This character also has a high full homophone density, at 27. As 是 is of such standout high frequency, it is expected to dominate in terms of responses.

yi

The syllable *yi* is of high frequency, with a median frequency of 93 and a sum phonological frequency (i.e. all *yi* characters regardless of tone) of 30493. It is also of very high segmental homophone density, at 56. This syllable was selected as a counterpart to *shi*. The highest frequency *yi* character is 一 yi1, with a frequency of 15844. Unlike the very high frequency *shi* character 是, which has a full homophone density of 27, 一 has a relatively small full homophone density of 8. The syllable *yi4* has a full homophone density of 27, but has no character exceeding a frequency of 2000. The next most frequent *yi* character is 以 *yi3*, with a frequency of 4729 and a full homophone density of 7.

1.2 Participants

22 native speakers of Mandarin, 11 females and 11 males ranging in age from 18 to 36 years of age, were recruited from the University of Oxford, and paid to participate in the experiment. Though coming from a number of regions in China (and some speaking a number of other Chinese 'dialects', in addition to English), all self-identified as native speakers of Mandarin. All were familiar with the *pinyin* romanisation, having learned it in elementary school as part of the national curriculum.

1.3 Experimental Procedure

Experiment 1 was conducted using Google Forms - a free platform for creating and disseminating surveys. Participants were given a time limit within which to complete the experiment, which was based on a slightly sped-up average time per question from pilot runs, in order to put them under moderate time pressure. This timing averaged five seconds per slide. Some participants finished slightly before time, though generally the time limits were marginally inadequate (as intended).

Free Choice

The *free-choice* elicitation method was used as it arguably mimics what happens in the processing of ambiguous natural speech, in that there are no artificial restrictions on responses. In this method, participants are asked to write down characters in response to *pinyin* stimuli - in particular, the first character occurring to them. The expectation is that participants will write down the character which 'best' matches the stimuli, i.e. that character which is activated the fastest or strongest. For stimuli which are fully specified, e.g. *yī*, the expected response is the highest frequency character available, in this case 一.

Impoverished stimuli, such as those which are not specified for tone, offer the possibility of discovering effects of full homophone density on lexical activation. Should two

characters of approximately equal frequency match the stimulus, e.g. *wan*, but differ in tone and have differing full homophone densities, we expect to find an effect of homophone density on responses. As discussed in **Chapter IV**, there is reason to suspect that homophone density will have differing effects on lexical activation by frequency. High frequency words are not expected to be greatly affected by density, while low frequency words are expected to benefit from having higher homophone density. Thus, we expect a bias towards low frequency words with higher homophone densities, when these words are in competition with other low frequency words having fewer homophones.

One potential downside of free-choice is that it is somewhat difficult to implement. It was found to be impractical for participants to type their responses, as typing in Mandarin introduces biases of its own. One common method of character-input is accomplished by typing *pinyin* (without tonal diacritics). The input system then prompts the user with a list of potential characters determined by the probability of occurrence with neighbouring syllables. This mimics some of the probabilistic effects under investigation, and it was decided that such promptings should be avoided. Participants were therefore asked to write their responses in characters on a provided sheet of paper.

As an initial step, it was decided to test for the effects of lexical frequency, homophone density, and the presence vs. absence of tone, in a context-free environment, in order to establish the extent to which these guide lexical choice in isolation, and how they interact with each other. The target syllables were displayed one at a time on an otherwise blank screen. Three pre-test stimuli were presented to each subject, to allow subjects the opportunity to familiarise themselves with the experiment and ask questions.

Each syllable was presented on the computer screen in *pinyin*, without tonal diacritics (see Figure 5.1). The square box beneath the target of each slide is an unfortunate and unavoidable result of using Google Forms - the software was not designed to run surveys to which no (computer-based) responses were expected. Participants were asked to write down

on a numbered sheet of paper the first character which occurred to them for the displayed syllable. There were only five test screens.

Figure 5.1. Screen capture of a question in Experiment 1.

As high frequency words have stronger lexical representations, or higher levels of resting activation (Bybee, 2001; Luce and Pisoni, 1998; Marslen-Wilson, 1987; McClelland and Elman, 1986), these should reach activation threshold the fastest, and are therefore expected to constitute the bulk of responses. Given the open-ended and slow nature of the task, requiring participants to write on paper, some variation in responses was also expected.

As the stimuli lack tonal information, there is a possibility for two characters to be of approximately the same frequency, but to differ in tone, full homophone density, and the frequency of full homophones. Such cases potentially allow for the exploration of the effects of frequency and density on lexical access. Given related findings by Luce and Pisoni (1998) for neighbourhood density in English, we tentatively expect to find that lower frequency characters experience a degree of facilitation when having high full homophone density, but not for higher frequency characters.

Should such a density effect exist, it is not clear what level constitutes 'high' as it pertains to frequency or density in Mandarin. In Luce and Pisoni's (1998) study on English neighbourhood density, high frequency words had an average frequency of 254 occurrences per million, and low frequency words an average of 5 occurrences per million; high neighbourhood density averaged 22 neighbours, and low density 11; high frequency

neighbours a mean of 370 occurrences per million, and low frequency neighbours 46 per million. An additional aim for Experiment 1 is therefore to observe subjects' response behaviour, with an eye to tentatively establishing 'high' and 'low' frequency and density regions for Mandarin.

2 Results

2.1 miao

For *miao*, three characters were given as responses by subjects, out of 12 characters in the corpus. Surprisingly, the most frequent character 描 *miáo* (at 105.15 occurrences per million in the corpus), was not given as a response by any subject. This character has 3 full homophones, and the syllable *miáo* a median frequency of 41. The marginally lower frequency *miao* character, 妙 *miào* (frequency: 104.73) was given as a response by three subjects. This character has a lower full homophone density of 2, but a higher median frequency of full homophones for *miào* at 74. The fourth and fifth most frequent characters, 苗 *miáo* and 秒 *miǎo* (frequency: 41 and 35, respectively) received 9 and 8 responses. The syllable *miǎo* has a median frequency of 23.

2.2 pang

For *pang*, the three highest frequency characters were given as responses. The highest frequency character by a considerable margin, 旁 *páng* (frequency: 168), was given as a response by 11 subjects. The second most frequent character, 庞 *páng* (frequency: 44), was given as a response by 4 subjects, and the third most frequent, 胖 *pàng* (frequency: 43) by 7 subjects. All other *pang* characters are of low frequency (5 occurrences per million or fewer), and received no responses.

2.3 wan

The highest frequency character, 完 *wán* (frequency: 743; full homophone density: 2; full homophone median frequency: 88) was not given as a response by any subjects. Surprisingly, the second highest frequency character, 万 *wàn* (frequency: 697; full homophone density: 4; full homophone median frequency: 357) was given by 14 subjects. Similar to *miao*, the four highest frequency *wan* characters all have different tones.

2.4 shi

The syllable *shi* is of high frequency (median frequency of all *shi* characters is 288 occurrences per million) and density (there are 50 *shi* characters in the modified corpus). The highest frequency character, 是 *shì*, is of very high frequency (13584 occurrences per million), and has a very high full homophone density of 27. It was given as a response by 10 subjects. The second highest frequency character, 时 *shí* (frequency: 4329) is also of very high frequency, though it is dwarfed by 是. It was not given as a response. The remaining 12 responses were for 8 characters across the frequency spectrum. None of these characters were in the same tonally-specified syllable (*shì*) as the highest frequency *shi* character 是.

2.5 yi

The highest frequency *yi* character, 一 *yī* (frequency: 15844) was given as a response by 17 subjects. The second highest frequency character, 以 *yǐ*, is of considerably less frequency (4729 occurrences per million), and was not given as a response. The remaining 5 responses were given for four characters across a range of frequencies, one of which, 衣 *yī* (frequency: 266) is a full homophone of 一.

As noted for other syllables in the experiment, the first three most frequent *yi* characters have different lexical tones. Unlike *shi* however, the most frequent *yi* character

does not have the most full homophones. The syllable *yì* has 27 full homophones, and its most frequent character is the high frequency (though low relative to 一) character 意 (frequency: 1871).

3 Discussion

A number of points stand out from the results of Experiment 1. Interestingly, all three responses for *miao* were from different tonally-specified syllables, and apart from the anomalous highest frequency character 描 (which was not given as a response), each of the three characters given is the highest frequency character in its respective full homophone cohort.

On the surface, this appears to indicate that tonal identity is critical in governing responses (and potentially lexical access), as only the highest frequency character from each tonally-specified syllable was given as a response. In the case of *miao* however, the highest frequency character from each tonally-specified syllable also corresponds to the few highest frequency characters in *miao* generally, i.e. the highest frequency syllables in *miao* happen to have different tones. This is still interesting in a distributional sense, as it results in each *miao* syllable having a character which is distinct in frequency from its full homophone mates.

If tone is in fact a weaker cue to lexical identity than phonemic/segmental information, as we might expect given previous research (e.g. Ye and Connine, 1999; Taft and Chen, 1992), this distributional oddity makes unexpected sense. Low frequency characters are generally indistinct, and tone alone may not provide a sufficient distinguishing characteristic. Higher frequency characters combine both tone and frequency as distinguishing characteristics, and the system is better able to cope with more homophony when this is the case. This may also partially explain the concentrations of homophony in high frequency syllables, as discussed in **Chapter III**.

For *wan*, as with *miao*, the highest frequency character was not given as a response. The two highest frequency characters, being of around 700 occurrences per million in frequency, qualify as fairly high frequency by Luce and Pisoni's (1998) standard, where the average high frequency word was around 250 occurrences per million. At this level, we would not expect to find an effect of density on responses. Recall that we expect to find a facilitation effect of full homophone density and full homophone frequency for lower frequency characters, but not for higher frequency characters (which are unaffected). Yet for *wan* we found an apparent facilitation of *lower* full homophone density but *higher* full homophone frequency, as the lower frequency (but lower full homophone density) character was the primary response.

The somewhat lower frequency 万 has only one full homophone, the low frequency character 腕 (frequency: 17). The higher frequency character 完 has three full homophones, one of which, 玩, is also of moderately high frequency (140 occurrences per million), and was itself given as a response by three subjects. This result may be consistent with our expectations. We expected to find a facilitation effect of higher full homophone density for lower frequency characters. In this case, there are two characters of comparable high frequency, and we find that subjects opt for the lower frequency, but lower density (and perhaps crucially, lacking a high frequency full homophone) character. At higher frequencies, the presence of relatively high frequency full homophones may have a dampening effect significant enough to overcome a relatively small frequency advantage.

For the syllable *shi*, the existence of an extraordinarily high frequency character in *shì* appears to have blocked all other *shì* characters, and any non 是 response (i.e. the highest frequency in *shi*) was for the other syllables. We would expect the low frequency *shì* characters to benefit in terms of lexical access from having many, high frequency full homophones. It is not possible to see whether this is the case due to the design of this section of Experiment 1. The fact that such a high frequency character has such a large number of full homophones, many of which are of very low frequency, does provide circumstantial evidence

that the homophone processing system is more capable of handling high homophone density when high frequency characters are present.

Finally, for the syllable *yi*, apart from the responses given to the very high frequency 一, all but one response went to *yì* characters. Subjects either gave as a response the highest frequency character, or a character from the very high density syllable *yì*. For words of lower frequency, having higher density is apparently of some benefit.

4 Summary

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. In tonally-deficient stimuli with relatively large frequency disparities, for example in the syllable *shi*, frequency was found to strongly dominate homophone density. The character most often chosen in *shi* was also in the highest density full-homophone cohort, but was still chosen due to being of high frequency *compared to its competitors*.

Experiment 1 produced a number of pieces of evidence for our hypotheses, namely that higher homophone density and frequency appear to be of benefit to lower frequency words. Further, as for *wan*, it may be the case that for high frequency words, the existence of many full homophones (and crucially, homophones that are also of high frequency) has a dampening effect.

However, it was generally the case that at least one character of fairly high frequency was to be found in each syllable. Such characters tended to obscure lower frequency characters due to the experimental design, by attracting the majority of responses. The results of Experiment 1 provide a small amount of evidence supporting the hypothesis that full homophone density interacts with frequency in affecting lexical access, in a roughly

analogous way that neighbourhood density has been found to affect the speed and accuracy of lexical identification in English (e.g. Luce and Pisoni, 1998).

Experiment 1 was, however, extremely limited in scope. The experimental design allowed for hints of a full homophone density effect to be observed, but provided no insight into, for example, the time-course of character decisions. In order to build on these results, and circumvent the limitations of Experiment 1, a greatly expanded experiment was designed. This is discussed in the following chapter.

Chapter VI. Experiment 2

1 Motivation for Experiment 2

Experiment 1 presented lexically ambiguous Mandarin monosyllables in the *pinyin* romanisation to native speakers, and asked participants to write a character corresponding to the given pinyin syllable. As discussed in detail in **Chapter V**, the purpose of the experiment was to investigate the influence of certain lexical variables (namely character frequency, full and segmental homophone density, and mean full and segmental homophone frequency) on the interpretation of lexically ambiguous monosyllables presented in isolation. It was found that for a given syllable (segmental string with no tone) associated with a range of possible characters, character frequency guided selection, with the character of the highest frequency triggering the highest response rate. When the disparity in frequency between the highest frequency character and the next most frequent character was large, relative full homophone density (the number of characters with the same segmental string and tone) was not observed to have any impact on selection (i.e. the given response). However, when the frequency disparity between possible options was reduced, characters with higher full homophone density were preferred (at least in the lower frequency bands) over characters with fewer homophones.

The results from Experiment 1 suggest that, in the absence of contextual information, character frequency, while the dominant factor, is not the *only* factor influencing the process of lexical selection among homophones. Relative density of the full homophone cohort also appears to play a role, and in an unexpected way. It might be expected that a particular homophone option would 'stand out' more in a smaller cohort, due to having less competition. This, at least, is what has been established for lexical recognition among phonological *neighbours*, with words having many neighbours recognised more slowly and less accurately than those with fewer neighbours, due to increased competition (cf Luce &

Pisoni, 1998; McClelland and Elman, 1986; Vitevitch and Luce, 1998). As modelled in TRACE (McClelland and Elman, 1986), the Shortlist model (Norris, 1994) and the Neighbourhood Activation Model (Luce & Pisoni, 1998), the activation of phonological neighbours causes a delay or even failure in accurate lexical access. However, the diverging results for homophones indicate that the activation of a particular representation benefits from a *larger* cohort of full homophones. This divergence in recognition patterns between phonological neighbours and homophones is puzzling. If we scrutinise the evidence from the phonological neighbourhood models more closely, however, we find an interesting qualification of the 'high density = greater competition and inhibitory affect' conclusion. In particular, there is a reported interaction with frequency. Luce and Pisoni (1998) find that *low* frequency words do derive some benefit from being in higher phonological density neighbourhoods. High frequency words are always more accurate and faster in a word decision task, but in low density neighbourhoods they were 10.3% more accurate (62ms faster), while in high density neighbourhoods they were only 4.5% more accurate (47.5ms faster). In other words, for phonological neighbours, low frequency words do *worse* when they have fewer neighbours than when they have more. Low frequency words would appear to gain some benefit from overall greater levels of activation of the general phonological neighbourhood. Nevertheless, a boost of high homophone density to high frequency words would be harder to explain.

Mandarin homophones may be processed in a similar fashion to words in a similarity neighbourhood, possibly in conjunction with words differing only in tone, which are phonetically quite similar. Experiment 1 was exploratory in scope, with each of the five syllables differing a great deal from the others in terms of segmental homophone density, and the frequency and full homophone density of constituent characters. The experiment was thus wide, but shallow, successful in identifying a potential interaction of frequency and full

homophone density in character selection, but insufficiently controlled for the key variables to provide a clear picture of this interaction.

A second experiment was therefore conducted, designed to more thoroughly explore the effects of density, and of any interaction between frequency and density, on lexical access.

2 Experimental Materials

2.1 Stimuli

One limitation of Experiment 1 is the restricted scope of the data it examined, giving only a snapshot of the various possible combinations of frequency and density. The experiment used only five unique tonally unspecified syllables, thus severely restricting the range of combinations possible. The design of Experiment 1 also forces any effects of the independent variables to be interpreted as categorical, and does not allow for the possibility of observing gradation. The five syllables were selected so as to be maximally distinct from each other (in terms of distributional statistics), and the data provided by the experiment therefore generally only come from the extreme poles of the variables' ranges, favouring an interpretation of them as belonging to distinct categories rather than different points on a scale, and no insight into what may be happening in between. Experiment 1 was designed in this way primarily to allow for the identification of the *potential* effects of frequency and density on selection, but it does not allow for the examination of the impact of gradience of those variables on subject responses.

Experiment 2 was thus designed to provide a fuller, more finely grained picture of the effects of frequency and density on the process of lexical activation in Mandarin. The experiment examined 153 segmentally unique syllables (approximately 38% of all segmentally unique syllables in Mandarin), spanning the full range of the statistical variables

(i.e. frequency, full and segmental homophone density). No full syllable (segment+tone) was used more than once in the experiment, in order to avoid the potential effects of repetition on responses.

2.2 Elicitation

A further limitation of Experiment 1 centres on the elicitation method. The 'free choice' method presents a number of potential problems. First, when using free choice it was necessary to present somewhat unusual tonally unspecified *pinyin*, in order to prevent biasing subjects. Syllables were presented (e.g. *wan*), in which the two highest frequency characters have different full homophone densities (as they have different tones), in order to observe any potential biasing of selection based on full homophone density. Using tonally unspecified pinyin is not ideal, however, as it is rarely encountered by the average person. Secondly, the method also required participants to write on paper, as in this context typing would not be a practical option. Typing is frequently accomplished by typing tonally unspecified pinyin, which generates a list of matching characters (from all tones) from which to choose. This would potentially skew results, as the computer-generated suggestions are themselves based on statistical probabilities. Writing a character is not ideal either, as it is open-ended and difficult to limit or measure for reaction time, allowing participants time to contemplate, and perhaps write characters other than those achieving lexical access most rapidly.

A further and more serious limitation of the elicitation method adopted in Experiment 1 is that we are presented only with the results of the selection process. This allows little insight into the process of character selection, in particular the degree to which density is maybe inhibiting or promoting a particular response even when ultimate selection is determined by frequency. This difficulty is illustrated by way of two examples in Table 6.1.

Example	Character	*pinyin*	Responses	Frequency	Full Homophone Density
1	一	*yī*	22	15844.24	8
	劓	*yì*	0	0.10	27
2	时	*shí*	22	4329.04	9
	舐	*shì*	0	1.21	27

Table 6.1. Hypothetical examples of the results-oriented nature of Experiment 1.

In both examples, all participants selected the higher frequency character, confirming that character frequency influences the probability of character selection, but providing no information on the relative certainty of selection (as shown by response times), whether a higher relative frequency speeds up selection, or whether the speed of selection is affected at all by homophone density. Does the difference in relative frequency between the two 'winners', the selected characters, affect the selection process? Does the higher density in the non-selected character slow down the response, despite (apparently) not influencing the end result?[4] Collecting and comparing response times would allow inferences to be drawn about the impact of relative frequency and homophone density on the process of selection.

In order to explore these questions, a different method was used for Experiment 2, which involved participants' reaction times and accuracy in matching tonally-specified pinyin syllables with characters (see *Method* below for full details). Of the 208 characters used as stimuli in Experiment 2, 119 were in the 'match' condition (meaning the pinyin perfectly matched the standard pronunciation of the character), while the remaining 89 were in the 'mismatch' condition (the pinyin did not perfectly match the standard pronunciation of the character). By recording a scalar characteristic of the response, i.e. reaction time, and allowing mistakes to be made, the design of Experiment 2 allows insight into potential effects

[4] It is also of note that the 'winner' in Example 1 is considerably less visually complex than the competitor. The potential impact of this additional variable is acknowledged, but will not be explored in detail in this thesis.

of variable strength, and how particular variables condition responses and interact with each other. Thus, this more complex experimental design allows potentially greater insight into the organisational structure of the lexicon with regard to homophones, and of the nature and representation of homophony in Mandarin itself.

2.3 Participants

One potential confound in Experiment 1 concerns the recruitment of participants. All participants in Experiment 1 self-identified as native speakers of Mandarin, though coming from a variety of regions in China. This is potentially problematic, as the term 'Mandarin' can cover a number of languages/dialects, which can vary from relatively small details of tone realisation and the tonal identity of certain characters, to relatively large scale variance in inter-intelligibility, though they are often partially mutually intelligible with each other (Tang and van Heuven, 2009). Compounding the complexity of this issue is the fact that most speakers are taught Standard Mandarin in school, or use it as a *lingua franca*, and may self-identify as a native speaker of Standard Mandarin regardless of language background[5]. As a large number of people from the PRC speak a language in the Mandarin branch, and not Standard Mandarin *per se,* recruiting 'Native Mandarin Speakers' is inadequate in ensuring a broadly similar language background in all participants.

To avoid the potential difficulties caused by recruiting participants with differing language backgrounds, all participants in Experiment 2 were recruited from the same city. To ensure this, all participants were required to have been born in, and (primarily) raised in Beijing, and to have parents from Beijing. Beijing was chosen primarily because the phonology of Beijing Mandarin most closely resembles that of the idealised Standard Mandarin, upon which the corpus *pinyin* transcriptions of characters are based. Careful subject recruitment is of particular concern when examining homophone density effects in

[5] In addition, some varieties of Mandarin are of low prestige in China, and many university students are not enthusiastic about describing their language as one of these 'regional accents' (see Li, 2004).

Chinese languages. Small phonological differences between varieties, for example the coalescing of Standard Mandarin coda alveolar and velar nasals into just a velar nasal in Hangzhou Mandarin, results in dramatically different homophone densities - *jīn* and *jīng,* which have homophone densities of 8 and 11 in Standard Mandarin, collapse into *jīng* with a full homophone density of 19 for such speakers.

It should be noted, however, that the actual corpus data comes from all over China. It is difficult to differentiate dialects, and even languages, in Chinese script, as with some exceptions (e.g. specific dialect words, character variations such as the 'traditional' characters used in Hong Kong and Taiwan) the same characters are used regardless. Frequency counts may not be as accurate as we would like, as the corpus likely contains not only multiple dialects but multiple languages (e.g. Cantonese).

3 Experiment 2 Method

In this section Experiment 2 will be described in detail. This description will include a discussion of the experimental design, selection of materials, and experimental procedure.

In Experiment 2, using the software package E-Prime (Psychology Software Tools, Pittsburgh, PA), participants were shown a pinyin monosyllable on a computer screen for 250ms followed immediately by a character, and asked to respond whether the pinyin matched the character by pressing either 'yes' or 'no' on a response pad, from which actions E-Prime collected responses and response times. Once a response was registered, participants were presented with a pause screen, which required a separate button press ('continue') to proceed to the next stimulus. This pause screen was included to allow participants to self-pace the experiment, and prevent the collection of erroneous responses and response time data due to fatigue and inattention. The stimuli were in two primary conditions - a *match* condition, and a *mismatch* condition (see Table 6.2).

Condition		Presented Pinyin	Character	Character Pinyin
Match		*bān*	班	*bān*
Mismatch	**Complete**	*qì*	乌	*wū*
	Segmental	*shuō*	他	*tā*
	Tonal	*bàn*	般	*bān*

Table 6.2. The conditions (and sub-conditions) of pinyin-character pairs used in Experiment 2. The 'Presented Pinyin' column shows the *pinyin* as presented in the experiment, while the 'Character Pinyin' column gives the actual phonological composition of the character.

In the *match* condition the pinyin syllable matches the pronunciation of the character with which it is paired, and participants were expected to respond 'yes'. In the *mismatch* condition the pinyin differs from the standard pronunciation of the character in one of three ways - 1) Complete Mismatch (different segments and tone); 2) Segmental Mismatch (different segments, same tone); and 3) Tonal Mismatch (same segments, different tone).

3.1 Selection of Materials

Three primary variables for each character have been identified as potentially influencing the outcome and time-course of disambiguation - Character Frequency, Full Homophone Density, and Segmental Homophone Density (which is the number of segmentally identical words, regardless of tone, and is a potential intermediate cohort between the full homophones of the target and the much larger conventional phonological neighbourhood). Other variables such as character complexity and the size and nature of traditional phonological neighbourhoods were not controlled for in stimuli selection. As discussed in detail in **Chapter IV**, characters of very low frequency (less than 1,000 occurrences in the corpus) had been removed from the

corpus. This was done in order to avoid using as stimuli characters which participants are not likely to know, and which do not have any character representation in the lexicon[6].

The remaining characters were then categorised along three variables - Frequency, Full Homophone Density, and Segmental Homophone Density. The category divisions are ultimately arbitrary, in that no particular claim is made as to their functional or psychological reality. The categories were devised for practical purposes, to allow for the selection of stimuli characters across a defined, controlled range for each variable (Table 6.3). This work resulted in a list of 3,524 characters which were categorisable by the three variables, providing a pool from which a large number of carefully controlled stimuli could be selected for the experiment.

Dividing the corpus according to these criteria results in 40 unique 'Character Profile Types' - each profile type being a unique combination of the three variables. This makes it possible to select representative stimuli across a typology of profile types constructed from all combinations of a full range of Frequencies, Full Homophone Densities, and Segmental Homophone Densities, by selecting at least one character from each profile type. The expectation is that the possible effects of frequency, density, and the interaction of the two, will be represented in the data across the full range of each variable.

[6] There were several characters used Experiment 1 which a number of participants were unfamiliar with. The corpus was winnowed to avoid presenting characters which participants were unlikely to recognise.

Corpus Character Count	Segmental Homophone Density	Full Homophone Density	Character Profile Type
Very High (HH) - 500,000+	High (H) - 21+	High (H) - 11+	1
		Medium (M) - 4-10	2
		Low (L) - 1-3	3
	Medium (M) - 8-20	High (H) - 11+	4
		Medium (M) - 4-10	5
		Low (L) - 1-3	6
	Low (L) - 1-7	Medium (M) - 4-10	7
		Low (L) - 1-3	8
High (H) - 100,000-499,999	High (H) - 21+	High (H) - 11+	9
		Medium (M) - 4-10	10
		Low (L) - 1-3	11
	Medium (M) - 8-20	High (H) - 11+	12
		Medium (M) - 4-10	13
		Low (L) - 1-3	14
	Low (L) - 1-7	Medium (M) - 4-10	15
		Low (L) - 1-3	16
Medium (M) - 50,000-99,999	High (H) - 21+	High (H) - 11+	17
		Medium (M) - 4-10	18
		Low (L) - 1-3	19
	Medium (M) - 8-20	High (H) - 11+	20
		Medium (M) - 4-10	21
		Low (L) - 1-3	22
	Low (L) - 1-7	Medium (M) - 4-10	23
		Low (L) - 1-3	24
Low (L) - 10,000-49,999	High (H) - 21+	High (H) - 11+	25
		Medium (M) - 4-10	26
		Low (L) - 1-3	27
	Medium (M) - 8-20	High (H) - 11+	28
		Medium (M) - 4-10	29
		Low (L) - 1-3	30
	Low (L) - 1-7	Medium (M) - 4-10	31
		Low (L) - 1-3	32
Very Low (LL) - 1,000-9,999	High (H) - 21+	High (H) - 11+	33
		Medium (M) - 4-10	34
		Low (L) - 1-3	35
	Medium (M) - 8-20	High (H) - 11+	36
		Medium (M) - 4-10	37
		Low (L) - 1-3	38
	Low (L) - 1-7	Medium (M) - 4-10	39
		Low (L) - 1-3	40

Table 6.3. Categorisation of variables used in Experiment 2.

It might first appear that there should be 45 profile types - Frequency(5) x Segmental Homophone Density(3) x Full Homophone Density(3) = 45. However, High Full Homophone Density, which requires 11+ full homophones, cannot be coterminous with Low Segmental

Homophone Density, which requires a segmental cohort of fewer than 8, resulting in five fewer possible types (thus the missing types are an artefact of how the variables were categorised).

As discussed in the previous section, the experimental stimuli were presented in two primary conditions - a match condition, and a mismatched condition (sub-branching into three kinds of mismatch). The criteria for selecting the characters and accompanying pinyin for these two conditions will be discussed in the following section.

Matching Condition

For the matching condition, where the presented pinyin correctly represents the standard pronunciation of the paired character, an attempt was made to select four characters for each of the forty profile types in the corpus. The decision to test four characters from each profile type was arbitrary, but was constrained by two primary factors - the number of characters in each profile type available for selection, and practical constraints on the overall length of the experiment. The number of characters in the corpus falling into each profile type is highly variable (and reflects the uneven distributional nature of the Mandarin lexicon as profiled in **Chapter III**, e.g. extremely high frequency words tend to occur in high density neighbourhoods - making it difficult to locate four very high frequency words with low segmental and full homophone density, such as Character Profile Type 8). For example, the corpus contains 733 characters which fit Character Profile 37 (LL Frequency, M Segmental Homophone Density, M Full Homophone Density), but contains only 1 character which fits Character Profile 4 (HH Frequency, M Segmental Homophone Density, H Full Homophone Density). When it comes to stimuli selection, this non-uniformity is of practical concern. There is little point in setting a target of 20 stimuli per profile type when many of the profile types do not have that many characters.

The second constraint on the number of stimuli selected for each profile type is the ultimate length of the experiment. There is a limit to the amount of time participants can be expected to concentrate on what is a relatively mundane task - reading pinyin and characters, and pressing buttons, in a sound-proof booth. The aim was to keep the actual experimental section of the experiment (minus introductions, instructions, etc.) to around 20 minutes. With these constraints in mind, a target of four characters per profile type was decided upon for the matching condition.

In choosing which characters to include in the experiment, an effort was made to select as many unique tonally specified syllables as possible - both to provide a wide range of test syllables in order to avoid potential anomalous behaviour triggered by individual characters or syllables, as well as to minimise possible priming effects from the repetition of particular *pinyin* syllables as stimuli. Though there are a few characters which share the same standard pronunciation in the experiment, such as 般 *bān* and 班 *bān*, no two pinyin stimulus syllables are the same - in this example, the former character is paired with *bàn* (i.e. tonal mismatch) and the latter *bān* (i.e. match). Characters with multiple associated pronunciations (e.g. 乐: *lé*, 'happy'; or *yué*, 'music') were excluded from the experiment. For a variety of reasons it was not always possible to select four characters for each character profile type - in most cases this was due to a particular profile type having fewer than four representative characters. A total of 118 characters were selected for use in the matching condition in the experiment, out of an ideal total of 160. (See *Appendix 1* for the full list of *match* stimuli).

Mismatch Condition

In addition to the matching condition, 3 additional characters per profile type were paired with imperfectly matched or completely mismatched pinyin per Character Profile Type - one character per mismatch sub-condition per profile type. The pinyin syllable chosen to go with each character was selected from the same 'Character Profile Type' as the character itself,

where possible, to control for possibly distorting effects of different distributional properties. This means that the given mismatched pinyin syllable has a character of the same 'character profile type' as the target character, i.e. a character with the same frequency, full homophone density, and segmental homophone density as the target character. This was done in order to provide a mismatched pinyin syllable with the closest possible profile to the target character, and prevent possible confounds due to pairing, for example, an extremely low frequency and density syllable with an extremely high frequency and density character.

There are a few cases where this ideal matching was simply not possible, particularly for the Tonal Mismatch condition. It was not usually difficult to find in the corpus two characters of the same profile type which differ in segmental composition and tone (i.e. 'complete' mismatches such as 以 *yǐ* and 时 *shí*, both of which are of character profile type 2 - HH Frequency, H Segmental Homophone Density, M Full Homophone Density), or which differ in segmental composition only (i.e. 'segmental' mismatches such as 以 *yǐ* and 里 *lǐ*, also both of character profile type 2). It is considerably more difficult to locate two characters of the same profile type which differ only in tone (i.e. 'tonal' mismatches such as 以 *yǐ* and 一 *yī*, again both of character profile type 2), as it is unlikely that a particular syllable, even one with as high a homophone density as *yi*, will have two characters of different tones that are of the same character profile type. This is particularly problematic as homophone density decreases, and the odds of finding two characters from the same segmental cohort in the same profile type diminish (in some cases to the point of impossibility, e.g. 擦 *cā*, which has a segmental homophone density of 1).

In cases where a precisely matched syllable could not be found, a syllable was selected which was the closest in frequency (cumulative frequency of all full homophones) to the target syllable, and which had the necessary tonal and segmental properties. This process was often something of a balancing act, particularly for high density syllables. For example, in the corpus there are only 5 characters that fit profile type 2 (HH Frequency, H Segmental

Homophone Density, M Full Homophone Density). Of these 5 characters, 4 were used in the matching condition. The only remaining character for the mismatch condition is 一 *yī*, which was designated for use as a 'tonal' mismatch. Ideally there would be another *yi* character with a different tone in this character profile type. As it so happens, in this case there is such a character - 以 *yǐ*. We would then pair the pinyin of this character, *yǐ*, with the character 一. In this case however, *yǐ* is already in use in the matching condition, and the experiment does not allow for the repetition of syllables as stimuli. There are no other *yi* characters of a different tone in character profile type 2 - there are no other characters in this profile type of any description. We therefore look at the phonological frequencies (meaning the cumulative frequency, for a given tonally specified syllable, of all constituent characters) of each of the *yi* syllables, to find the one that is closest in frequency to *yī* (see Table 6.4).

Syllable	Corpus Occurrences
yī	3253593
yí	186067
yǐ	1262437
yì	1166958

Table 6.4. Phonological frequencies of the four tonally specified *yi* syllables.

In this case, the target character comes from a very high frequency syllable with no particularly close neighbours. The syllable *yǐ* is the closest, but is unavailable, so we would initially choose the slightly less ideal *yì* as the tonal mismatch for 一 *yī*. As it turns out, this syllable is needed for the matching condition of character profile type 33 (LL Frequency, H Segmental Homophone Density, H Full Homophone Density), which is only partially filled (having only 2 out of the 4 ideal characters). We are therefore left with only *yí*, the least bad among non-ideal options. This process of balancing the requirements of one profile type for

stimuli against the requirements of the other profile types was repeated throughout the experiment - though generally the resolution of such conflicts was easier than in the case of *yi* just described.

For each 'character profile type', an attempt was made to find one character per mismatch type, resulting in a potential total of 120. A suitable candidate for each of these types was not always found, and a total of 90 mismatched characters were selected for use in Experiment 2 (see *Appendix 2* for a full list of the *mismatch* stimuli).

3.2 Experimental Procedure

The experiment was conducted in a sound-dampened booth in the Department of Chinese Language and Literature at Peking University in Beijing, on a laptop using the software package E-Prime 2.0 (Psychology Software Tools, 2002). Participants sat at the computer, and interacted with the experiment via a Cedrus response pad (Figure 6.2) - the keyboard and mouse were not used at any point. Three buttons were labelled (in English) - left ('No'), right ('Yes'), and down ('Continue'). English was used on the response pad to avoid priming certain syllables and characters for participants. There are a number of options for labelling the buttons in Mandarin, but perhaps the most straightforward choice is 是 *shì* 'is' for 'Yes', and 不是 *bù shì* 'is not' for 'No'. The character 是 is used in the matching condition, and it was thought desirable to avoid participants seeing the character throughout the experiment (also, the word for 'no' is derived from the word for 'yes'). Other possibilities, such as 匹配 *pǐ pèi* 'match' and 不匹配 *bù pǐ pèi* 'not match', encounter similar difficulties (*pèi* is also used in the mismatch condition). English labels were selected to avoid these problems, and participants were asked to verify that they understood the labels. For around half of participants 'Yes' and 'No' were reversed, to allow for the potential of a handedness bias in response times.

The experiment began with an instruction screen and a short tutorial, in which the experiment and operation of the response pad were described. It was explained that pinyin syllables would be presented on the screen, followed by characters which may or may not match the pinyin. Participants were asked to respond by pressing 'Yes' for a match, and 'No' for a mismatch, and were put under moderate time pressure by being reminded to go as quickly as possible while still avoiding mistakes.

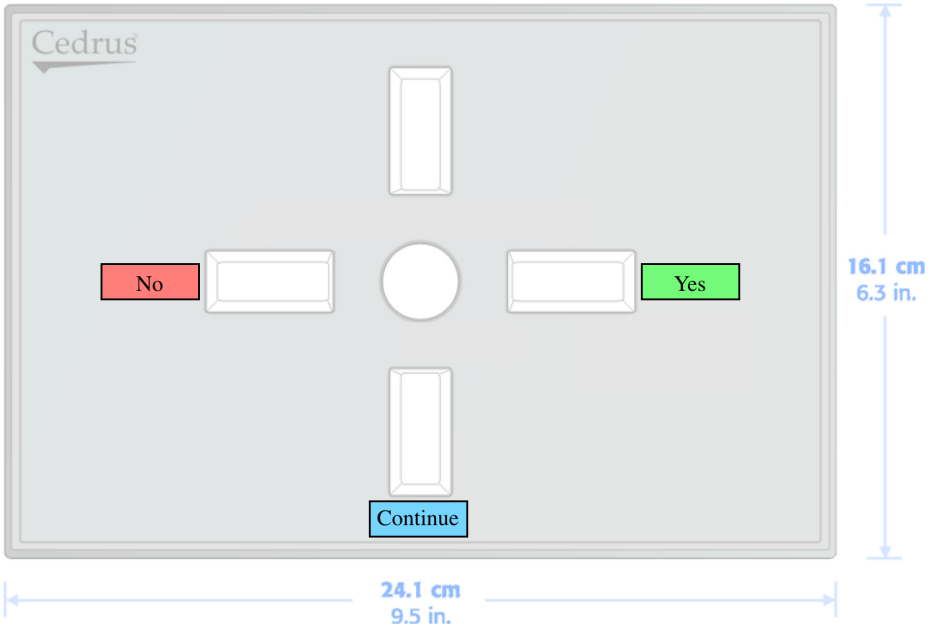


Figure 6.1. Cedrus RB-530 response pad used in Experiment 2, in the 'Right/Yes' configuration.

This instruction was followed by a short practice session. A total of 15 characters were selected for the practice - 8 in the matching condition and 7 in the three mismatched conditions. The practice stimuli were independently randomised for every participant, and none of the practice material overlapped with any of the experimental material. Upon completion of the practice section, participants were asked if they had any questions about the experimental procedure. In the absence of any questions, or when all questions had been satisfactorily answered, the participants then proceeded onto the experiment proper.

The experiment itself contained 208 stimuli (see *Appendices 1* and *2* for the complete list). Like the practice stimuli, the experimental stimuli were independently randomised for each participant. This was done to allow for the monitoring of potential learning curves in response times and ordering-of-stimuli biases.

The process of a single question was as follows (refer to Figure 6.2, which shows example slides). First, a slide stating 'Press the down key to continue' appeared on the screen, and remained until a response was given (i.e. the 'Continue' key was pressed). This was followed by a fixation star in the centre of the screen, which remained for 1000ms. The purpose of this fixation star was to act as a focus, and prompt participants to maintain eye contact with the position on the screen where the stimuli would appear. This serves to prevent the collection of misleading response time data, such as long response times due to wayward eye focus or attention, and not the experimental variables. For future experiments, it may be of benefit to modify the duration of the fixation star to give it a variable duration, which may encourage closer attention and prevent subjects from getting into an answering rhythm.

The fixation star was then replaced by a pinyin syllable, which remained on the screen for 250ms. This duration was arrived at after informal piloting with native speakers. The intention was to create moderate time pressure, and encourage participants to pay close attention - the pinyin syllable was to be visible for long enough to be read, but not so long as to allow for rumination (and would be missed entirely if the participant was looking away). As has been discussed, it was decided that in Experiment 1, in which there was no time pressure, participants had too long to think about their answers before responding. This possibly resulted in a small proportion of responses resulting not from the strength and speed of character activation, but from conscious thought and selection.

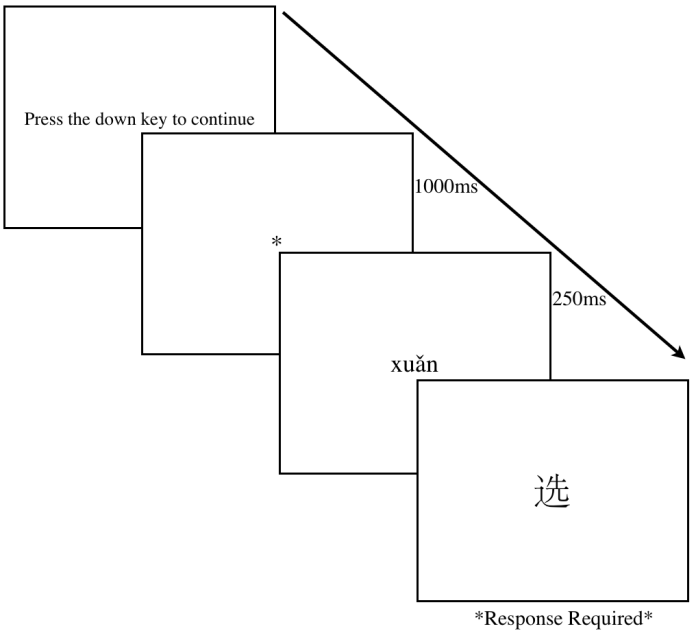


Figure 6.2. The time-course of Experiment 2.

After 250ms, the pinyin syllable was replaced by the target character, with no inter-stimulus event. This part of the timing component of Experiment 2 was informed by a study of homophone priming in English (Lukatela et al, 2002), which found more homophone priming as the inter-stimulus interval decreased, with the most occurring at 0 (i.e. nothing intervening between the prime and the stimulus, in this case the pinyin syllable and the character, respectively). As Experiment 2 was designed to look at the effects of certain stimuli (i.e. pinyin syllables) on the processing of the targets (i.e. characters), no inter-stimulus object was included in the experimental design.

Participant responses were recorded from the onset of the appearance of the character, and the character remained on the screen until a response was given. Once a response was registered by the software, the character was replaced with the instruction 'Press the down key to continue'. This allowed participants to relax their attention periodically throughout,

and pace the experiment as they felt comfortable. As the relatively short duration of the pinyin stimulus required significant concentration, this was viewed as a necessary feature.

3.3 Participants

40 native speakers of Mandarin, 21 females and 19 males ranging from 18 to 30 years of age, were recruited at Peking University in Beijing, China, and paid to participate in the experiment. All participants were natives of Beijing, and all self-identified as native speakers of Mandarin. All were familiar with the pinyin romanisation, having learnt it in elementary school as part of the national curriculum. No speech or hearing difficulties were reported.

No participant self-identified as left-handed in the questionnaire, though 3 identified as ambidextrous (all in the right/yes group). It is highly likely that these three individuals are biologically left-handed, given that in China (as has formerly been the case in 'Western' countries) left-handedness is strongly discouraged and 'corrected' in children, particularly for eating and writing (Teng et al, 1979). In this case, participants who self-identified as ambidextrous may indeed use their right hands to perform certain tasks, such as writing and eating, while maintaining some degree of left-hand dominance in the experimental task.

3.4 Hypotheses

Experiment 2 was designed to examine the broad stroke frequency findings of Experiment 1 with greater precision and for a wider set of data and, through a greatly expanded stimuli pool and the collection of response times, also enable the exploration of a number of other frequency-related questions. By recording response times, we also investigate whether different homophone densities have an effect at higher frequencies (an effect that would, if it exists, have been obscured by the methodology of Experiment 1).

Given robust findings of the effect of frequency on lexical access generally in previous studies, both for Mandarin and more generally (e.g. Rumelhart and Siple, 1974; Gordon,

1983; Whaley, 1978; Peng, Liu, and Wang 1999), and the bias in character selection from Experiment 1 suggesting stronger representations for higher frequency characters, it is hypothesised that higher frequency characters will have faster average response times, and that response times will decrease as frequency increases.

Of more specific interest is the locus of frequency effects in the model. As discussed in **Chapter IV**, within any model that assumes that an abstract phonological form is the locus of frequency effects rather than, say, the lemma, increasing homophone density should have a positive effect on access, through frequency inheritance. However, this effect should really only be significantly noticeable in low frequency words, which benefit most clearly from the cumulative frequency of neighbours, especially if they have high frequency neighbours. High frequency words are already high in frequency, and increase in cumulative frequency only marginally through the addition of lower frequency neighbours. If all homophones share a single representational form, however, then we would also expect all homophones in the cohort to take the same amount of time to reach threshold activation, and have the same response times in Experiment 2 (c.f. Levelt et al, 1999). If frequency inheritance is assumed, it is still possible that when presented *in context*, the context provides top-down activation, speeding selection of the most appropriate candidate; devoid of context, as in these experiments, and on this interpretation, there is nothing to distinguish high from low frequency homophones if the phonological representation is taken to be unique and shared, and frequency effects to reside at this level.

Given the findings of Experiment 1, in which characters of high frequency were preferred, even over characters with extremely high frequency homophones, frequency inheritance of this type is not expected. Rather, the locus of frequency effects is deemed to be the lemma. Smaller effects related to homophone density and cumulative homophone frequency are also predicted, in line with observed effects of neighbourhood density (e.g.

Luce and Pisoni, 1998)[7]. This is also assuming, of course, some form of exhaustive (parallel) access, where all homophones participate in the activation process simultaneously. In serial access, high frequency characters are unaffected by density, as they are at the top of the 'list'. For low frequency characters with very high density, activation should be slowed by dint of a low position in the 'list'. It will be recalled that Experiment 1 found that low frequency words with high homophone density garnered more selections, but no response time data was available. Luce and Pisoni (1998) found that low frequency words did *better* in high density phonological neighbourhoods than in low density neighbourhoods, and this effect is expected to be mirrored for homophony.

Given previous studies on the effect of increasing similarity neighbourhood size for low frequency words (e.g. Hong and Yelland, 1996; Luce and Pisoni, 1998; Pisoni et al, 1985), and on the findings of Experiment 1, which indicated that high homophone density could be beneficial in terms of selection when character frequency was relatively even, it is expected homophones are processed in similar fashion to a similarity neighbourhood in the Neighbourhood Activation Model (Luce and Pisoni, 1998). It is possible that a similarity neighbourhood is activated in parallel that includes non-homophones (e.g. phonological neighbours), but does not include full homophones, i.e. the competition is only between phonological strings. Once this process has completed and the segmental and tonal string of the target has been ascertained, then all words matching the given string (e.g. the full homophone cohort) would be evaluated. This seems unlikely, as the system would only have access to bottom-up information for words with homophones - if top-down information were available to the system, then presumably it would be used to distinguish between homophones prior to certain phonetic identification, just as it is assumed to do for words without homophones in a phonological neighbourhood in the Neighbourhood Activation Model.

[7] Of course any neighbourhood density effects in this particular context would be in addition to homophone density effects, the implications of which lie beyond the remit of this thesis but are potentially substantial.

Our expectation is that homophones will participate in the word recognition process in parallel from the beginning, and not through some kind of post phonological identification decision process. This would indicate that the phonological string receives positive identification without making reference to lemma identity (i.e. semantic fit to context in an utterance). Given the deeply homophonous nature of Mandarin, full homophones (and potentially segmental homophones, as being quite phonetically similar) are expected to constitute the core of similarity neighbourhoods, with higher full and segmental homophone densities expected to result in faster and more accurate responses in recognition, particularly for lower frequency characters, in part based on the results of Experiment 1. On average each word in Mandarin has around 13 segmental homophones, though some words have zero (no full or segmental homophones), and some have well over one hundred. It is expected that these words (full and segmental homophones), being phonetically closest to the target word, and potentially very numerous, will have a greater impact on the time-course and accuracy of recognition than more conventionally defined segmental phonological neighbours.

Mandarin has very few unique syllables, of limited phonotactic construction, meaning the (conventionally defined) similarity neighbourhood for any given syllable is proportionally large. To take a random example, the similarity neighbourhood for the character 坝 *bà* 'dam' consists of 27 syllables (*a*, *ban*, *ca*, etc.) - around 7% of all Mandarin syllables. Each of these syllables, of course, has a number of segmental homophones (on average around 13), resulting in a similarity neighbourhood of around 350 words. By way of comparison, Marian et al (2012) estimate a considerably lower average phonological similarity neighbourhood size of 5.49 for English. It is possible then that, given the deeply homophonous nature of Mandarin, and its phonotactic structure and distribution, 'segmental' phonological neighbourhoods play a diminished role in Mandarin, this role already being filled primarily by full and segmental homophones.

3.5 Mismatch Conditions Hypotheses

The preceding hypotheses apply most straightforwardly to the *match* condition (in which the presented pinyin syllable correctly represents the pronunciation of the character with which it is paired). Participants were expected to respond 'yes' to characters in the *match* condition, and the speed at which 'yes' responses are given is expected to be sensitive to the frequency and homophone density of characters as described above. The number of inaccurate 'no' responses is expected to be small.

The *mismatch* conditions are slightly more complicated. One of the reasons for including mismatches was as a means of avoiding a response time bias towards 'yes' answers - as participants were asked to judge whether a pinyin syllable and a character match, there must be some incompatible pairs in order to make the task meaningful. However, mismatches are also expected to provide a testing ground for the hypotheses outlined above. Complete mismatches are expected to be identified as mismatches quickly and accurately, but it is the partial mismatches which are potentially more interesting.

For example, a critical question in defining Mandarin homophony involves the relative weighting of segments and tones, and how this affects speech recognition. There is evidence that characters which differ only in tone are more difficult to reject as matches than characters which differ by a single segment (Taft and Chen, 1992; Ye and Connine, 1999; Cutler and Chen, 1997; Lee, 2007).

The cause of this observed difference in response patterns (i.e. two characters differing only in tone taking longer to identify as being *different* than two characters differing only in vowel quality) is of some interest. Is it due to a categorical (i.e. abstract representational) difference between tones and vowels, regardless of content, or to a difference in perceptual similarity?

This topic is not directly addressed in this thesis. An experiment utilising auditory stimuli would be capable of testing for 'homophone nearness' for phonetic features, and

might reveal whether tonal differences are at the end of a scale of perceived difference (i.e. different, but difficult or slow in perception), or are categorically different from segmental differences. It would also potentially reveal that certain tones are perceptually 'closer' to some tones than others - for instance a full tone 3, with its falling portion, may be more difficult to distinguish from a tone 4 (also falling) than a tone 2 (rising).

Mandarin may not be the ideal language for testing this idea, as it has three contour tones and one level tone, which have different isolation points (Lai and Zhang, 2008). In fact, the observed difference in vowel versus tone mismatches in Mandarin may be due solely to a difference in the time required to identify a monophthongal vowel and a contour tone. Indeed, Lai and Zhang (2008) found that tone 1 (level tone) has the earliest isolation point, followed by tone 4, then tones 2 and 3. Investigating a language with multiple level tones (e.g. Cantonese or Min; Norman, 1988) might better reveal potential scalar differences within tones (e.g. level high and level mid are more difficult to distinguish, take longer to distinguish, or both, than level high and level low).

This thesis does not assume an abstract or categorical difference between segments and tones. Rather, we predict based on previous research (e.g. Ye and Connine, 1999; Lee, 2007) that characters that differ only in tone will be more difficult/take longer to identify as different than those differing in vowel quality. We expect to find, therefore, a continuum of similarity (identity) from full homophones, to segmental homophones (i.e. words distinguished prosodically, through tone), through to phonemic neighbours. Along this spectrum, the closer to phonetic identity a set of words is, the greater amount of evidence will be required to choose the correct candidate, i.e. the greater amount of time, and potentially less accurate, lexical identification will be.

With this in mind, two types of partial mismatch were devised - 1) Segmental Mismatch (different segments, same tone, i.e. 他 *tā* paired with *shuō*); and 2) Tonal Mismatch (same segments, different tone, i.e. 般 *bān* paired with the pinyin syllable *bàn*). For

Segmental Mismatches, it may be that sharing the same tone as the target character results in a degree of activation of the character representation, making it somewhat more difficult to reject as a match. Responses to Segmental Mismatches are expected to be nearly as quick and accurate as for Complete Mismatches, as the character substantially mismatches the pinyin prompt, but does match for tone. In other words, our hypothesis is that segmental information is weighted more heavily than tone.

In a Tonal Mismatch (e.g *bān-bàn*), the character and the pinyin match in terms of segments, but have different tones. There is evidence that tonal information is processed differently than segmental information. For example, tone identification is significantly slower than vowel identification in Mandarin, despite the obvious temporal overlap (Ye and Connine, 1999). There is also evidence that syllables which match in segmental composition but not in tone do activate each other (Lee, 2007; Taft and Chen, 1992). Lee (2007) found direct priming effects in Mandarin only when the prime and target had the same segmental structure and tone. However, in mediated priming at short inter-stimulus intervals (50ms), priming effects were found when the target and prime matched in segmental structure, regardless of tone. This indicates that activation of segmental homophones was taking place, even when the tones were mismatched.

Taft and Chen (1992) found a similar effect of segmental homophone activation. They presented two characters in sequence to native speakers of Mandarin, and asked them to judge whether the two were homophones. When the two characters matched in segmental composition but differed in tone, responses were significantly slower and more inaccurate (with around five times more errors) than when the two characters had differing segments. This indicates that tone was poorly utilised in making homophone judgments, at least potentially due to temporal considerations. Mandarin tones utilise pitch contours spanning the voiced portion of a syllable, and a sufficient amount of the contour must be heard before a successful tonal identification can be made (c.f. Ye and Connine, 1999).

It is therefore expected that for Tonal Mismatches, where the pinyin syllable and the character have identical segmental structure but differ in tone, the character representation will undergo strong activation, making it more difficult to reject quickly as a match. Responses to Tonal Mismatches are expected to be both error-prone and slow - significantly slower than either Complete or Segmental Mismatches, as well as Matches. Ideally, for this experimental paradigm, the Segmental Mismatch would involve the pairing of a word in pinyin with a character that differed only by a single segment. The primary interest in examining Segmental and Tonal Mismatches revolves around the potentially differing effects of segments and tones on lexical organisation in Mandarin, and whether a Tonal Mismatch is similar to a canonical phonological neighbour. A comparison of this sort would give insight into whether tones are as critical in lexical organisation and speech processing as segments.

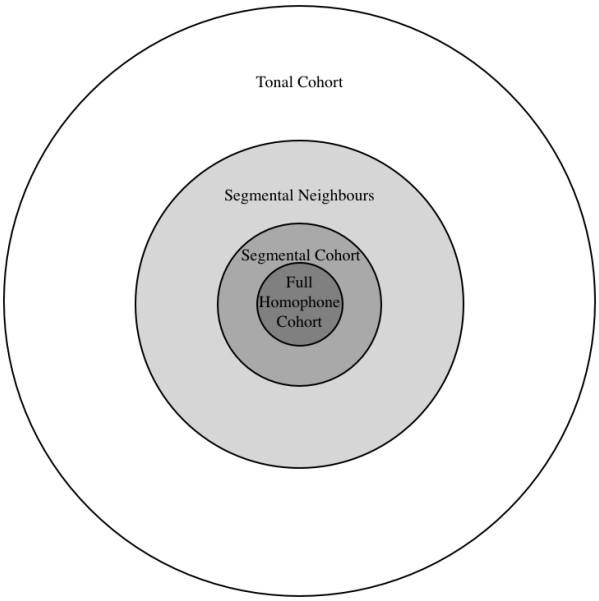


Figure 6.3. Visualisation of 'neighbourhoods' by hypothesised strength of mutual activation. (Note: Tonal Cohort - all characters share the same tone only. There are only four subcategories, i.e. the four lexical tones; Segmental Neighbours - all same segments except one, e.g. phonological neighbours; Segmental Cohort - all characters which share the same segments, but have different tones; Full Homophone Cohort - all characters which share the same segments and tones).

Given the evidence discussed above, the expectation is that tone and segments *are not* equivalent, and that characters in the same segmental homophone cohort (i.e. those characters identical in segments, regardless of tone) experience much stronger mutual activation, and are closer to being true homophones, than those characters falling into the more traditional definition of a phonological neighbourhood (i.e. 'Segmental Neighbours'), in which 'words' differ by a single segment. This *could* be due to an abstract, categorical difference between segments and tones, though we expect that it is due to perceived phonetic similarity. Changing *all* of the segments is predicted to have a larger impact than changing just one segment, making the Segmental Mismatch condition more an examination of a *tonal* cohort (meaning characters share the same tone regardless of segments) than a traditional phonological neighbourhood. This serves as an examination of how much mutual activation derives from solely sharing a tone, and no segments. The expectation is that, for any two characters, the farther away they get from being full homophones, the less mutual activation they will experience. This is represented by the schematic diagram in Figure 6.3, with posited degree of mutual activation decreasing with distance from the origin of the circle. This diagram represents an interpretation of homophony as being at the end point of a scale of identity – a concept to which we will return further on.

4 Results: Matching Condition

Two characters were removed from the analysis of the matching condition. Character 137 was initially intended to be used in the matching condition. It was later modified for use in the mismatch condition, and will be included in those results (see Section 3). The other character removed from the analysis is Character 188. Only around 50% of participants responded to this character correctly (i.e. that the character matched the given *pinyin*). It is not entirely clear why this was the case - the character is in the LL frequency category, but is by no means the lowest frequency character used in the experiment. It is possible that the character is

utilised primarily in a certain context, leading to an inflated frequency for the character given its apparent obscurity. Regardless of what precisely was responsible, there was clearly some confusion with this question, and it was removed.

Only two subjects had more than 5% incorrect responses to the matching condition. The data for both of these subjects have been retained in the analysis. In terms of response times: five subjects had more than 10% of response times over two standard deviations away from the mean; three subjects had around 15% or over; and one subject had 26% of responses over two standard deviations from the mean (and, in addition, was frequently talking during the experiment). This last subject has been removed from the analysis.

All remaining responses of more than 3 seconds were excised. This resulted in the removal of 11 responses, or 0.27% of the total. Given the nature of the task, it was decided that any response time exceeding 3 seconds was likely due to non-experimental factors, e.g. loss of concentration or unsure knowledge of the character under consideration, and not due to the distributional characteristics of the character. The fact that the frequency category of the characters to which these +3 second responses were associated was evenly distributed supports this assumption: 3 of the excised responses were to HH characters; 3 to H characters; 2 to M characters; 1 to an L character; and 3 to LL characters. Response time cutoffs of less than 3 seconds skewed towards lower frequency characters, and could have obscured the effect of frequency on responses.

No further action was taken in terms of removing responses from the data set. It was decided that any more stringent 'data clean-up' had the potential for obscuring genuine results. For example, removing all responses that were more than two standard deviations above the mean response time results in this distribution of cut answers: 17 from HH; 27 from H; 36 from M; 38 from L; and 57 from LL. In other words, in the distribution of responses there is a relationship between frequency category and exceeding the standard deviation, with lower frequency characters having in general a greater tendency toward a wider spread of

response times, in particular longer response times. Cutting these responses potentially obscures the fact that subjects systematically respond to characters in different frequency bands in different ways (i.e. this is already showing a frequency effect).

The results were first examined as groups by the five frequency categories which were defined for the purpose of creating the experimental materials (see Table 6.3). The delineation of these categories was made primarily on the grounds of practical considerations, and is not based on any prior assumptions about how the mind processes characters of different frequencies - different categorisations were tested (both in the number of categories and their boundaries), with very similar results. Each category, however, contains roughly the same number of characters in the experiment (it should be stressed that this does not mirror the corpus generally, where lower frequency characters greatly outnumber higher frequency characters), and the categories cover the full range of lexical frequencies in our Mandarin corpus, with as wide a range of full homophone densities within each frequency category as possible. In the absence of previous research on potential regions for frequency effects in Mandarin, and having no *a priori* reasons to settle on different categories, these original categories were selected as a first approximation.

In the absence of a robust frequency inheritance effect - where the cumulative frequency of all full homophones, and not individual character frequency, would be expected to affect response times - character frequency should influence response times. The sensitivity of response times to frequency may mask the effect of full homophone density, should it exist, as the density effect is expected to be smaller in magnitude. This masking, in terms of analysis, is largely due to the limitations of Mandarin structure - these two variables do not perfectly interact. Ideally, Mandarin would present us with sets of characters of exactly the same frequency, with regularly increasing full homophone densities (Table 6.5).

	Character 1	Character 2	Character 3	Character 4	Character 5	Character 6	Character 7
Frequency	342.96	342.96	342.96	342.96	342.96	342.96	342.96
Full Homophone Density	1	2	4	6	8	10	12

Table 6.5. Idealised character distribution by Frequency and Full Homophone Density.

If such sets existed, across a broad spectrum of frequency ranges, it would be relatively easy to observe the effect of full homophone density on response times in an experiment such as this one. Such 'ideal' sets are not common in Mandarin, however. Table 6.6 gives an example of one of the sets closest to this ideal from Experiment 2.

	错	独	刚	限	帝	育	举
Frequency	324.09	333.43	335.62	342.87	342.96	345.64	357.96
Full Homophone Density	3	3	8	9	7	17	4

Table 6.6. Actual character distribution by Frequency and Full Homophone Density in Experiment 2.

For a series such as this, it is more difficult (in, say, a linear regression) to tease apart differences in response patterns as being due to frequency or homophone density, as both variables are changing simultaneously (and there is bound to be noise in the results). Linear regression on the entirety of the data from Experiment 2 returns negligible effects for any variable. However, by grouping series of characters into frequency bands, it may be possible to draw out the effect of full homophone density on responses, and how density interacts with frequency. This type of analysis comes at the considerable cost of blurring the frequency effect, as all characters within a certain band are treated as having the same frequency. Each character in the experiment was only seen once, so the variation is high. We have a small snapshot, and the trend is only clear in aggregate. A large experiment on a small set of characters would, perhaps, provide greater clarity.

If there is a consistent effect of full homophone density across the full spectrum of character frequency, the bandwidth of the frequency categories will be the primary concern. Presumably categories with more limited frequency ranges will exhibit more clearly the effects of full homophone density, as the effect of differing character frequencies on responses will be minimised. If homophone density effects interact with frequency, with lower frequency words experiencing greater facilitation from higher full homophone densities than high frequency words, as predicted above, then the location of frequency category boundaries will also be important. The inclusion of characters of sufficient frequency to override homophone density effects in a frequency category in which the lower frequency characters *are* affected by full homophone density, would blur the observed density effect in the analysis.

The results for the original frequency categories devised for the experiment are given in this section. Frequency has been Log10 transformed, as it has been found that frequency affects behaviour along a logarithmic scale (Gahl, 2008). There are few very high frequency characters, and a great mass of fairly low frequency characters in Mandarin. The categories were designed to allow a sampling of segmental and full homophone densities within a given frequency range, resulting in decidedly non-linear frequency categories. The lowest Log10 frequency score for any character in the experiment is -5.28. This is equivalent to 1000 occurrences in the corpus (or a frequency of 5.19 per million), and represents the low frequency cutoff. A summary of the categories, or 'bins', is given in Table 6.7.

Original Bins	Log10 Frequency Score	No. of Characters
HH	-2.59 or greater	20
H	-3.28 to -2.59	25
M	-3.59 to -3.28	25
L	-4.28 to -3.59	22
LL	-5.28 to -4.28	23

Table 6.7. Bins based on the 'original' frequency categories of Experiment 2.

The results of Experiment 2, organised by frequency category (i.e. 'bin'), and arranged to reveal the influence of Full Homophone Density on response times, are given in Figure 6.4. The figure gives the mean response time to all characters of a given full homophone density, within a frequency category (i.e. the responses to two characters in the same frequency category, with the same full homophone density, are counted as referring to the same character). This blurs the frequency statistic, while providing an 'idealised set' for comparing full homophone densities. Each column in the figure represents a particular full homophone density, increasing from left to right. Each symbol in the figure represents an individual participant response (different frequency categories are represented by different symbols), with response time on the y-axis. There are approximately 4,500 responses represented in Figure 6.4.

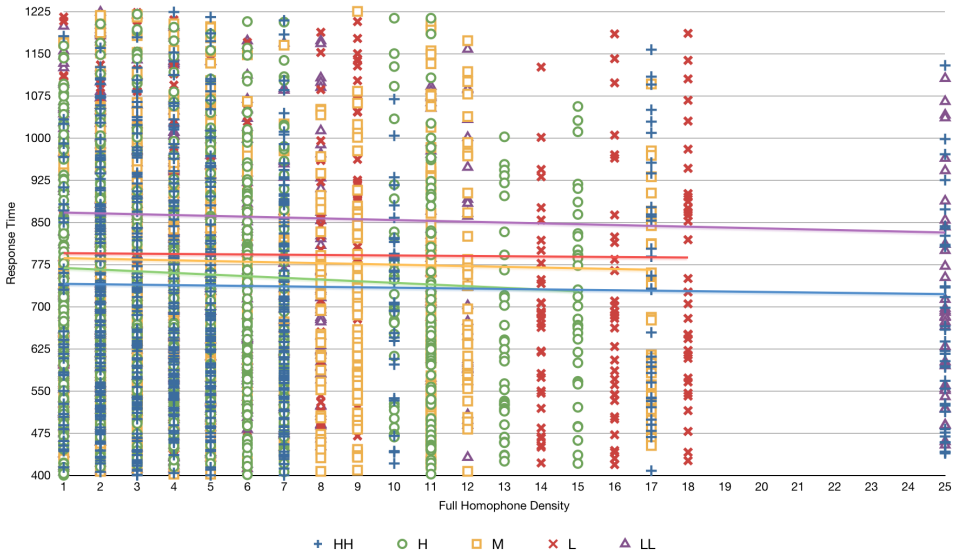


Figure 6.4. Response times (RT) by Full Homophone Density, utilising the Original Frequency Bins. Bin 1 corresponds to the HH Category (highest frequency), and bin 5 the lowest frequency category.

For full homophone densities with many representative characters in the experiment (e.g. there are 20 characters with a full homophone density of 2 in the experiment, with around 800

responses), the visual result is somewhat chaotic. It appears from the figure that most of the characters with a full homophone density of 2 are from bin 1 (highest frequency). In fact, bin 1 only has four characters with a full homophone density of 2 - exactly average. Some full homophone densities (e.g. 14) are only represented by a single character in the experiment, while the density ranges for the frequency categories are not always equivalent. Categories 2, 3, and 4 have no characters with densities over 18. It was not always possible to find characters in each frequency range with identical density ranges. As has been stated a number of times, there are simply far more characters with low homophone densities. This translates into difficulties creating perfectly even bins, particularly in the higher density ranges. A trendline, giving the slope of each line, has also been calculated for each frequency category. Individual columns with this many data points are impossible to divine graphically, but are used to generate the slopes.

In terms of a general frequency effect, the results are as expected. The lowest frequency group (LL, or Bin 5) generally has the longest response times for any given Full Homophone Density. This can be observed in Figure 6.4, where the trendline for bin 5 (the top line, in purple), is consistently higher than the other trendlines. This is followed in the expected order, in decreasing response time, by bin 4 (L frequency category; red line), bin 3 (M frequency category, yellow line), bin 2 (H frequency category, green line), and bin 1 (HH frequency category, blue line). The figure also shows that, as full homophone density increases, response times decrease (the lines fall as they proceed left to right). This is true for all five frequency categories. A summary of the trendlines is given in Table 6.8.

Bin	Log10 Frequency Score	Trendline
HH (1)	-2.59 or greater	y = -0.76x + 741
H (2)	-3.28 to -2.59	y = -2.95x + 772
M (3)	-3.59 to -3.28	y = -1.29x + 788
L (4)	-4.28 to -3.59	y = -0.46x + 796
LL (5)	-5.28 to -4.28	y = -1.48 + 869

Table 6.8. Trendline results of ‘bins’ analysis.

From the trendline summary, the frequency effect is quite clear. The y-intercept (response time) increases as frequency decreases, with the largest jump (73ms) between the lowest and second lowest frequency characters in the experiment. For all frequency categories, response times decrease as full homophone density increases (all trendlines have negative slope). It was predicted that low frequency characters would experience facilitation from occurring in higher density neighbourhoods, while higher frequency characters would experience little or no facilitation. From these results, those expectations are only partially borne out. In Experiment 2, characters appear to have benefited from increased homophone density regardless of frequency category - having more full homophones appears to be facilitating lexical activation and access (all trendlines in Figure 6.4 drop as they progress from left to right, i.e. response times shorten as full homophone density increases within a given frequency category). Somewhat surprisingly, the category experiencing the least amount of facilitation is bin 4 (the second lowest frequency category, L), while that experiencing the most facilitation is bin 2 (the second highest frequency category, H). However, in partial support of our expectations, though all frequency categories benefited to some degree from increased density, all categories apart from bin 4 (L) exhibited greater facilitation from increased full homophone density than bin 1 (the highest frequency characters, HH). This can be observed from the slope of the lines in Figure 6.4.

On this analysis, there is little evidence of a robust frequency inheritance effect, as conventionally understood, which assumes homophones share the same phonological representation. Under a strong frequency inheritance effect, character frequency should not correlate strongly with response times (as found in Experiment 2), as the interpretation would be that it is *not* lemma frequency, but cumulative phonological frequency of all homophones (i.e. the combined frequency of all *lìn* characters, not just the frequency of 蔺) that matters. Given zero context, as in Experiment 2, homophones of low character frequency should have the same time-course for lexical activation and response as high frequency homophones.

Homophone pairs were not used in Experiment 2, disallowing any direct comparisons (full homophones were not used, as both would require the same *pinyin* stimulus; if subjects had already seen *lìn* followed by 蔺 in the experiment, it was thought likely they would remember, and respond differently to 吝 than if they had seen it alone). It is clear, however, that cumulative frequency is not predictive of response patterns, even though there are both lemma frequency effects and a subtle density effect. Characters of low frequency have slower response times than characters of high frequency. Low frequency, high density characters have generally slow response times, though they are faster than low frequency, low density characters.

This effect was not expected to be as large for higher frequency characters, however, as such characters are expected to benefit less relative to their own (already high) frequency from having more homophones. Word frequency, it has been noted (e.g. Gahl, 2008; Hay, 2002), has a roughly logarithmic effect on responses. For example, a character of 3,000,000 occurrences in the corpus is likely of little effective difference to a character of 3,001,000 occurrences. For a character of only 1000 occurrences however, a character that has 1,000 more occurrences in the corpus is twice as frequent, and we might expect this to make a bigger difference in terms of response times. An extremely high frequency character will typically not receive much of a relative 'frequency boost' from homophones which will tend to be of lower frequency. These homophones should not add much, in a sense, to the strength of the lexical representation (assuming of course that the words share the same representation - it is also possible that characters with lower frequencies share a representation separate from higher frequency characters). For very low frequency characters however, the addition of only one homophone can cause a massive increase in phonetic/phonological frequency. We might therefore expect a stronger relative effect of density on response times in the lower ranges of frequency, though no evidence to that effect is indicated by the results, and density nevertheless confers some advantage on high frequency characters in lexical processing. If

this density effect is indeed relatively insensitive to simple cumulative frequency, and is due rather to some combination of the number of related types (and their respective frequencies), it suggests that homophones are processed/stored as separate entities at the phonological level, or that the critical variables for word recognition are in the first instance lemma-associated (i.e. lemma frequency) and, secondarily, associated with the number of other lemmas that share the same, or quasi-identical, form.

The benefit conferred from having more homophones is not due to simply sharing the same phonological representation, with a higher cumulative frequency and higher levels of mutual excitation. This would predict a beneficial effect on lower frequency words only. At the same time, the greater potential competitive field posed by a large cohort does not appear either to be inhibiting the selection of words as it does in the case of high frequency words in large phonological neighbourhoods (Luce and Pisoni, 1998). Thus, a different mechanism is at work from that guiding lexical retrieval from a pool of phonological neighbours. Retrieval, in the case of both low or high frequency characters, is dominantly guided by lemma frequency, but also appears to be influenced by strength of lemma association, indicating an explicit association, presumably via the phonological level, between these lemmas. These are not frequency inheritance effects, but a kind of identity effect nevertheless.

5 Results: Mismatch Condition

In total, there were 90 'mismatched' stimuli, of three types - 28 complete mismatches (both segmental and tonal composition of target characters mismatched the stimulus *pinyin*); 31 segmental mismatches (segmental composition of target character and stimulus *pinyin* mismatched, while tone matched), and 31 tonal mismatches (segmental composition of target character and stimulus *pinyin* matched, while tone mismatched). Recall that it was predicted, as per Experiment 1 as well as previous research indicating that tonal information is not as critical as segmental information in evaluating lexical identity (c.f. Ye and Connine, 1999;

Lee, 2007; Taft and Chen, 1992), that Complete Mismatches would be identified and rejected quickly and accurately, while 'Tonal Mismatches', owing to the presumed relatively weak role of tone in lexical identity, would be rejected as matches more slowly and with more errors (as they are deemed closer to homophones). Segmental Mismatches (i.e. only tone matched between stimulus *pinyin* and characters) were predicted to fall somewhere in-between, and a weak 'tonal homophone cohort' was predicted. A boxplot of mismatch type by response time is given in Figure 6.5.

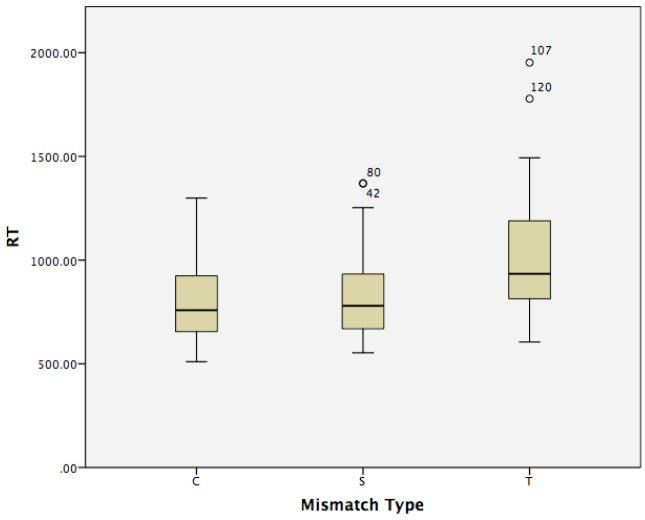


Figure 6.5. Boxplot of mismatch types by RT. 1 (Complete Mismatch); 2 (Segmental Mismatch); 3 (Tonal Mismatch).

When the segments of the target failed to match those of the stimulus, regardless of tone, participants were both faster and more accurate in rejecting the pair as a match. Characters completely mismatching their stimuli (in both tone and segments) were rejected as matches with a mean response time of 810ms, with 0.9% errors. Characters which matched the stimulus *pinyin* in tone, but not in segments (segmental mismatches) had a mean response time of 833ms, with 0.57% errors. When only the tone failed to match between the stimulus

pinyin and the target character, responses were both slower (mean RT: 1017ms) and more error prone (5.98%). The apparent lack of facilitation caused by the target and stimulus sharing a tone is perhaps not surprising. There are only four lexical tones in Mandarin, meaning any 'tonal neighbourhood' would be immense - constituting roughly one quarter of the lexicon. That segmental mismatches are roughly as easy to reject as matches as complete mismatches, and tonal mismatches more difficult, resonates with previous research, which has found that tone is *not* as critical in lexical identity (or is at least processed later) as segmental composition (Ye and Connine, 1999; Lee, 2007; Taft and Chen, 1992).

In order to ascertain whether there is an interaction between mismatch type and lexical frequency, a graphical analysis was first conducted. This is displayed in Figure 6.6. From the figure, it appears that characters in the Complete Mismatch condition were responded to the fastest in general (mean RT: 810ms), and showed a possible effect of frequency - low frequency Complete Mismatches (i.e. 'bin' 4) have a mean RT of 850ms, while high frequency Complete Mismatches ('bin' 1) have a mean RT of 770ms. This is represented in the figure as the blue line.

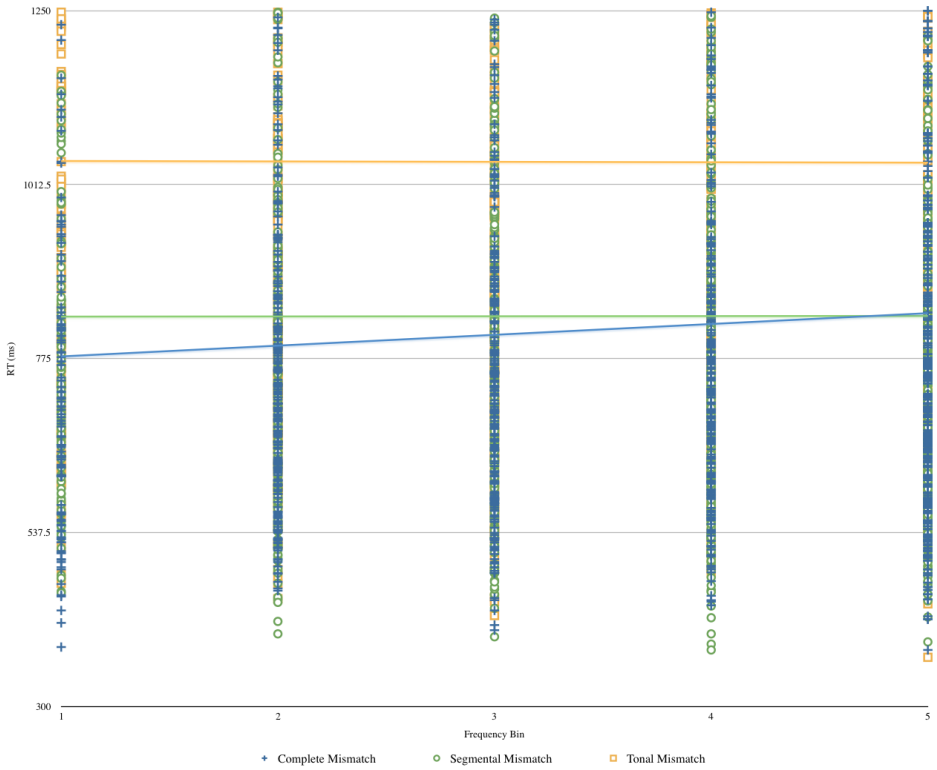


Figure 6.6. Response times (y-axis) by frequency bin (x-axis) to the three different mismatch types (coloured lines): complete mismatch - blue; segmental mismatch - green; tonal mismatch - yellow.

A factorial repeated-measures ANOVA was conducted on the mismatch data. Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of Mismatch Type, $\chi^2(2) = 38.92$, $p<.001$, and the interaction of Mismatch Type and Frequency Category, $\chi^2(2) = 70.47$, $p<.001$. Therefore degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity ($\varepsilon = .609$ for the main effect of Mismatch Type and .673 for the interaction).

All effects are reported as significant at $p < .05$. There was a significant main effect of Mismatch Type on RT, $F(1.22, 47.53) = 50.36$. There was also a significant main effect of Frequency on RT, $F(4, 156) = 2.51$. However, there was no significant interaction effect

between Mismatch Type and Frequency, $F(5.38, 209.87) = 1.953$, $p = .052$. Contrasts revealed that RTs for Complete and Segmental Mismatches were not significantly different, $F(1, 39) = .897$, $p = .35$. However, RTs for Tonal Mismatches were significantly slower than for Segmental Mismatches, $F(1, 39) = 51.55$.

6 Discussion

As expected, for characters in the matching condition, character frequency had a speeding effect on response times in evaluating homophony, in that higher frequency characters were responded to faster than lower frequency characters. When the results were viewed graphically by 'bin', a further facilitatory effect of full homophone density on response times was observed. When frequency was held relatively stable, it was found that increasing homophone density decreased response times. Contrary to expectations, this effect held across the frequency spectrum. It was expected that any density effect would mirror that established for neighbourhood density (Luce and Pisoni, 1998), and thus benefit lower frequency words, and be nonexistent in higher frequency words. This was not found to be the case for homophone density, suggesting a different set of relations between homophones than between near neighbours, and a different set of dynamics involved in any co-activation effects. It is also possible that the frequency 'bins' or categories utilised in the analysis were of inappropriate number or frequency range, and so gave an unrepresentative picture of effects across the density range; however, in trying to model these results, a number of different possible 'bin' sets were tested, all with similar results, so we do not believe the observed density effects to be an artefact of the method used.

For the mismatch condition, Tonal Mismatches (i.e. only tone fails to match) were found to be the slowest and most error prone, and the only category of mismatches that appeared to cause hesitation or indecision. It is important to note that this result does not justify a conclusion that tonal distinctions are unimportant in Mandarin, or that tonal

distinctions are categorically different from segmental distinctions. There are a number of reasons for this, the most critical being that the Tonal Mismatches differed only in one tone, while the Segmental Mismatches differed by *all segments*.

In order to explore this issue in more depth, an experiment (or series of experiments) would need to be conducted in which segmental mismatches of varying types (e.g mismatched by a single segment or feature) are compared to tonal mismatches of varying types (e.g. tone 1 vs. tone 2, tone 1 vs. tone 3, etc.). The perceptual salience of the contrasts would also need to be taken into consideration (as well as position within the syllable, e.g. initial vs. final), with the expectation that less salient contrasts would result in slower/less accurate responses. The results of such an investigation could provide insight into the gradient vs. categorical nature of such contrasts. I would expect to find that match/mismatch judgments would reflect a spectrum of contrast, with tones (and perhaps feature contrasts of low perceptual salience) at one end, and highly salient feature contrasts at the other.

In judging homophony for complete mismatches, a *pinyin* syllable is first read, e.g. *rén*. This presumably activates the cloud of exemplars corresponding to *rén*, with higher frequency *rén* characters achieving higher levels of activation than lower frequency characters. Following this, a phonologically-unrelated character is encountered, e.g. 马 *mǎ* 'horse'. The character activates the exemplar cloud for *mǎ*, none of which overlap to any meaningful degree with *rén*, and the pair is deemed a mismatch. This judgement is made relatively quickly, as the exemplar clouds bear little resemblance to each other.

For tonal mismatches, an exemplar cloud is activated upon reading the *pinyin* stimulus (e.g. *má*). Upon recognising the character 马, the exemplar cloud for *mǎ* is activated. It is more difficult (and slower) to determine whether the two clouds are different, as they overlap substantially (i.e. are phonetically quite similar). Our results indicate that character frequency does not affect the speed of this process, and we may only speculate as to why - it may be that process of distinguishing two highly similar exemplar clouds is just slow and difficult, and

masks the smaller frequency advantage gained from high frequency characters having faster activation of their own cloud. If only characters matching in both segments and tone are activated by the stimulus (the *pinyin*), we would expect responses to tonally mismatched characters to be relatively slow in comparison to Complete Mismatches for example. The stimulus pinyin should activate only (or at least primarily, leaving aside more general neighbourhood activation effects) those characters that fully match, for example all *wán* characters. The target character (which is, say, a *wān* character) is then encountered. This character is substantially similar phonetically to the stimulus, and we might expect the realisation of the mismatch to take more time than a character which fails to match the stimulus in *all* respects. We would also, however, expect the time-course of this realisation to be affected by the frequency of the particular character. These mismatched characters receive little or no activation from the stimulus, and therefore the amount of time it takes to access the target character (and determine it is a mismatch) is a function of lexical frequency. High frequency characters should be recognised faster than low frequency characters, and therefore rejected faster.

Instead, for the Tonal Mismatches, no effect of frequency was found on response times. This potentially indicates that the tonally-specified stimulus pinyin is providing activation to all characters matching in segmental composition, regardless of tone. When a target character was encountered matching in segments but not in tone, it had already been receiving activation from the stimulus, regardless of tone. It was still difficult (both in terms of response time and accuracy) to identify as a mismatch, again as expected. But its recognition was no longer a function of lexical frequency. Characters which mismatch a stimulus only in tone are more difficult to reject as homophones, indicating that tone is not a terribly strong cue for homophone identity (though it is still a cue, just a weak one).

Segmental mismatches (i.e. same segments between *pinyin* and character, but different tones), occupy a place somewhere between these extremes. The exemplar clouds are

substantially different, and responses of 'do not match' were made approximately as quickly as for complete mismatches. Unlike complete mismatches, however, no effect of character frequency was found. It may be that the tonal similarities result in very weakly overlapping exemplar clouds, which negates the benefit of high character frequency. Alternately, the apparent frequency effect for complete mismatches may be spurious. It may just be the case that character frequency does not affect the speed of a task such as the one employed in Experiment 2.

7 Summary

Experiment 2 has provided a number of different strands of evidence indicating that lexical retrieval is boosted not only by increased frequency, as found for cohorts of phonological neighbours, but, unlike phonological neighbourhoods, by increased homophone density as well. Lexical frequency effects have been shown to operate on the production of homophones in English (Gahl, 2008). Furthermore, high density in phonological neighbourhoods is known to induce differences in production, e.g. decreased duration and other effects of phonetic reduction (e.g. Kilanski, 2009; Gahl et al, 2012; Wright 1997, 2004; Munson and Solomon, 2004; Munson, 2007; Scarborough, 2009). A further experiment was therefore conducted to investigate the effects on production of lexical frequency and homophone density in Mandarin (Experiment 3), detailed in the following chapter.

Chapter VII. Experiment 3

1 Motivation for Experiment 3

In Experiment 2 it was found that, in the evaluation of words as homophonous or not, higher frequency words generated faster (correct) responses than lower frequency words, consolidating the findings of Experiment 1 which showed a preference for high frequency members of a homophone cohort when presented with an ambiguous stimulus. The faster retrieval of a higher frequency homophone helps clarify the mechanisms behind the preferential choice effect observed in Experiment 1. In a neutral context, a higher frequency member of a homophone cohort will also be the most probable interpretation, although this in and of itself does not guarantee the speediest response, since in theory all possible outcomes could be activated and then weighed for probability. The faster reaction times observed in Experiment 2 strongly indicate that the more commonly encountered word is more strongly/ quickly activated. Experiment 2 also provided a clearer picture of the subtle effects of homophone density found in Experiment 1. In the first experiment, higher density was seen as providing a boost to the lexical retrieval of low frequency homophones, with effects on high frequency words obscured by the experimental material. Experiment 2 confirmed this subtle effect for low frequency words, and also confirmed its application for all frequency levels (the five 'frequency categories' defined in Experiment 2).

This latter result was surprising in two respects. First, it shows a divergence from observed density effects for phonological neighbours (Luce & Pisoni, 1998; McClelland & Elman, 1986; Vitevitch & Luce, 1998), where increased competition causes delay in lexical access, especially for high frequency words. Second, while density boosts to low frequency words can, in theory, be ascribed to cumulative (or inheritance) frequency effects, this is more problematic to explain for high frequency words, since the potential added benefits of cumulative frequency for a very high frequency word will be minimal.

These results were interpreted as indicating that Mandarin homophones do not share a single phonological representation which is the locus of frequency effects, as response times to characters were largely determined by individual character frequency (i.e. no frequency inheritance as conventionally described, with all homophones sharing a single representation and therefore a homophone-cumulative frequency), but that lexical access is still affected by the presence, and number of, homophones - perhaps through strength of association with multiple lemmas.

The results of Experiment 2 are indicative of the effect of frequency and density on visual word recognition, but further questions remain as to whether and how such lexical variables affect, and may be utilised in, spoken language. If homophones do not share a single phonological representation, but do still affect each other, what is the nature of their relationship? Does each homophone have its own representation, or do only those homophones of sufficient frequency warrant an independent representation? How will the status of a character's representation affect fine phonetic detail in production? It is known from previous studies that phonological neighbourhood density affects production and recognition processes differently (see Gahl et al, 2012, for an overview). While higher neighbourhood density, as we have discussed, is reported to inhibit lexical retrieval, especially for higher frequency words (Luce and Pisoni, 1998; McClelland and Elman, 1986; Vitevitch and Luce, 1998), it has been shown to have a facilitative[8] effect in production (Dell and Gordon, 2003), resulting in a range of phonetic effects in words with higher density phonological neighbourhoods, including degree of vowel dispersion, degree of coarticulatory effects, and duration (Kilanski, 2009; Gahl et al, 2012 ; Wright, 1997, 2004; Munson and Solomon, 2004; Munson, 2007 ; Scarborough, 2009).

From the perspective of a usage-based framework, we should expect higher frequency words to have shorter relative durations than lower frequency words. Bybee (2001) proposes

[8] Or at least *diversifying*, since there is not complete consensus in the literature as to whether high density generates greater phonetic reduction or indeed the reverse, i.e. productions that increase intelligibility.

that this is due to 'articulatory routinisation', where with repeated use the articulation process is streamlined (see also Pierrehumbert, 2001). This has important implications for the extensive homophony in Mandarin – are homophone cohorts phonetically distinguishable by their duration?

If the locus of frequency is the phonological form, and homophones share a single phonological form, we might expect homophones to be just that - fully homophonous - with low and high frequency words having the same durations, perhaps related to their cumulative frequencies (i.e. frequency inheritance). If, however, each character has a unique form, consisting of encountered tokens or exemplars, we would expect to find systematic differences in duration based on character/lemma frequency, even for characters that are canonically considered homophonous. A phoneme-based model would potentially have difficulty accounting for such durational differences.

Lemma frequency is known to affect production in English, with higher frequency words having shorter durations than lower frequency homophones (Gahl, 2008). Gahl (2008) compared the durations of pairs of homophonous English words, and found that higher frequency homophones had shorter durations. This indicates that for a language with relatively shallow homophony (i.e. a word has at most a few homophones), the homophones may have independent representations. Higher frequency causes one homophone to shorten, leading to differences in production. This appears to contradict phoneme-based models, though Gahl (2008) notes despite her findings, it may still be the case that the lower frequency words in the study were benefiting from having a higher frequency homophonous mate, indicating some weak form of frequency inheritance. Given Gahl's work on English, we expect that individual character frequency will also affect production in Mandarin, with high frequency characters having shorter durations.

Unlike English, however, where homophones are relatively scarce, Mandarin has widely variable levels of homophony. Given the results of Experiment 2, which found a

reduction effect on response times for increasing full homophone density, and the studies showing density effects on the production of phonological neighbours, we also predicted density to affect the fine aspects of production in homophones, such as relative duration.

To examine this possibility, a final experiment was conducted (Experiment 3) to examine the possible effect of lexical frequency and homophone density on the detailed phonetic realisation of what are canonically considered homophones in Mandarin, with the general expectation that higher frequency and higher density will generate greater 'reduction', i.e. shorter word duration. The exact realisation of this reduction is potentially quite complex, both in the number of phonetic indicators (e.g. vowel reduction and centralisation, tonal reduction, etc.), and in the consistent measurement of those indicators. Syllable duration has been chosen for examination as having the greatest likelihood of exhibiting reliably measurable and interpretable indices of phonetic reduction.

2 Experiment 3 Method

2.1 Stimuli

Twenty characters (10 full homophone pairs) were selected for use in Experiment 3. In each pair one character is of relatively high frequency and one of relatively low frequency, with varying relative and absolute frequency relationships between the characters. Full homophone density for the pairs also varies from high to low (see Table 7.1 for a complete listing of stimuli).

The part of speech relationship within the pairs varies, as follows: one noun/noun pair, three noun/adjective pairs, two verb/adjective pairs, one verb/noun, one particle/noun, and one preposition/noun pair. The syllables were chosen not only for their distributional properties, but also for their segmental composition. There are two syllables with no consonants, 'ai' [aɪ] and 'wu' [u]; six syllables with consonants in the onset, 'dui' [tueɪ],

‘li’ [li], ‘shi’ [ʂɨ], ‘chu’ [ʈʂʰu], ‘zhi’ [ʈʂɨ], and ‘ji’ [tɕi]; and two syllables with consonants in the onset and the coda, ‘xin’ [ɕin] and ‘ren’ [ʐən].

Pair No.	Syllable	Character	Frequency	Log10 Freq	Full Homophone Density	Segmental Homophone Density	Character Gloss	Part of Speech
1	shì	是	13583.82	-1.87	25	45	is	copula
		侍	43.21	-4.36	25	45	servant	noun
2	rén	人	9701.65	-2.01	2	8	person	noun
		仁	89.66	-4.05	2	8	benevolent	adjective
3	lǐ	里	2793.09	-2.55	5	30	inside	preposition
		鲤	7.68	-5.11	5	30	carp	noun
4	ài	爱	561.78	-3.25	4	12	love	verb
		隘	8.15	-5.09	4	12	narrow	adjective
5	chū	出	3922.50	-2.41	2	14	go out	verb
		初	304.28	-3.52	2	14	junior	noun
6	duì	对	3654.39	-2.44	3	4	about	adverb
		兑	21.30	-4.67	3	4	cash out	verb
7	jī	机	1764.91	-2.75	20	54	machine	noun
		圾	16.11	-4.79	20	54	rubbish	noun
8	zhī	之	3162.92	-2.50	10	39	(literary possessive)	particle
		汁	18.51	-4.73	10	39	juice	noun
9	xīn	心	2037.08	-2.69	7	9	heart	noun
		馨	17.04	-4.77	7	9	fragrant	adjective
10	wú	无	1865.88	-2.73	5	27	no	adverb
		梧	6.33	-5.20	5	27	tree	noun

Table 7.1. Stimuli list for Experiment 3.

The target syllables were embedded in sentences of two types - a neutral carrier designed to provide only running speech context, but no semantic or syntactic context, and a biasing sentence providing both semantic and syntactic context. The target syllables were embedded in sentences in order to avoid possible effects of single-word elicitation, which is arguably ill-suited to the gathering of realistic tokens of the targets, as these often occur in stretches of speech and not in isolation.

It may be the case that in Mandarin the relationship between distributional statistics such as lexical frequency, and the phonetic realisation of a particular ‘word’, is different in citation form than in running speech. By using both full sentences and neutral carriers, it is also possible to examine whether and how semantic and syntactic context affect phonetic realisation. Guion (1995) (cited in Gahl, 2008), for example, found a duration difference

correlated to frequency in English, but only when the words were embedded in full sentences, and not neutral carriers.

In order to examine these relationships, a carrier sentence was devised, with the intention of providing running speech context but no bias toward a particular meaning of a syllable. The carrier used in the experiment was: 我在纸上写下 **'X'** 这个字 (I wrote the character 'X' on the paper). Twenty unique 'biasing' sentences were also devised, one for each character. An attempt was made to keep each pair of sentences fairly similar, with the sentences of similar length, and the target syllable falling in a similar position (see *Appendix 3* for a full listing).

2.2 Experimental Procedure

The target syllables were each elicited a total of six times from each subject. First, subjects read through a randomised list of the 'biasing' sentences (though all participants read the sentences in the same order). Once they read through the full list of twenty sentences, they started again at the beginning. All participants read the biasing sentences through three times.

The second task was to read through a list of the carrier phrases three times. These were presented in a different random order from List 1. This gave a total of six repetitions of each target character for each speaker. The full sentences were read first, to avoid the potential problem of participants knowing which characters were the target of the experiment (though this became obvious in List 2). Each list was read through three times, to check for a shortening effect of repetition on duration (Wiener et al, 2012).

2.3 Participants

The same 40 participants from Experiment 2 participated in Experiment 3. In Experiment 2, the importance of testing a group of participants with a homogeneous speech variety was discussed. This was even more crucial for Experiment 3, which was interested in examining

the fine phonetic detail of certain syllables in production. Researchers must be cautious when investigating production in Mandarin, as speakers of 'Standard Mandarin' often vary considerably by regional origin. Phonological differences in syllables, such as the lack of a distinction between alveolar and velar nasal codas in some southern varieties of Mandarin (e.g. Hangzhou-accented Mandarin), can result in drastically different homophone densities (c.f. Standard Mandarin *lin* and *ling* are both produced as *ling* by Hangzhou speakers).

3 Results

3.1 Neutral Carrier Condition

The results for the 'carrier condition', where the target word occurred in a neutral carrier phrase (我在纸上写下 'X' 这个字; 'I wrote the character 'X' on the paper'), are summarised in Table 7.2.

Pair No.	Syllable	Character	Log10 Freq	Full Homophone Density	Rep. 1 Duration	Rep. 2 Duration	Rep. 3 Duration	Mean Duration	Rep. Diff. (3-1)	Mean Duration Difference (H-L)
1	shì	是	-1.87	25	313.84	310.31	314.12	312.75	0.28	-10.59
		侍	-4.36	25	335.36	321.65	313.01	323.34	-22.35	
2	rén	人	-2.01	2	302.23	296.53	292.97	297.24	-9.26	-1.79
		仁	-4.05	2	298.94	296.54	301.60	299.03	2.66	
3	lǐ	里	-2.55	5	264.69	236.64	258.80	253.38	-5.89	-0.01
		鲤	-5.11	5	270.94	254.09	235.13	253.39	-35.81	
4	ài	爱	-3.25	4	229.39	225.27	220.80	225.15	-8.58	-5.14
		隘	-5.09	4	241.54	228.71	220.64	230.30	-20.90	
5	chū	出	-2.41	2	291.62	278.64	280.08	283.45	-11.54	-8.68
		初	-3.52	2	294.38	294.60	287.42	292.13	-6.96	
6	duì	对	-2.44	3	202.55	199.29	197.65	199.83	-4.90	-6.89
		兑	-4.67	3	209.99	205.40	204.76	206.72	-5.22	
7	jī	机	-2.75	20	256.78	248.74	253.63	253.05	-3.15	4.43
		圾	-4.79	20	260.22	248.12	237.53	248.62	-22.69	
8	zhī	之	-2.50	10	242.83	231.15	241.72	238.57	-1.10	-10.00
		汁	-4.73	10	251.76	245.99	247.97	248.57	-3.79	
9	xīn	心	-2.69	7	334.42	318.96	342.93	332.10	8.51	-14.59
		馨	-4.77	7	358.29	343.81	337.97	346.69	-20.32	
10	wú	无	-2.73	5	253.97	253.81	251.30	253.02	-2.67	1.36
		梧	-5.20	5	256.31	256.54	242.16	251.67	-14.15	

Table 7.2. Mean durations (ms) for characters in the carrier condition. 'Rep. Diff. (3-1)' column gives a subtraction of the third repetition from the first - negative numbers indicate the third repetition is shorter. 'Mean Duration Difference (H-L)' column gives difference between high and low frequency characters in a pair - a negative number indicates that the high frequency character is shorter in duration than the low frequency character.

Given previous findings of frequency-based durational differences between homophones in English (Gahl, 2008), and the effect of frequency on response times in Experiment 2, it was expected that a reduction in duration would also be found for the higher frequency character of Mandarin homophone pairs. In 8 of the 10 pairs, the higher frequency character is shorter in duration than the lower frequency character. For 6 of these pairs, the difference is 5ms or greater; for 3 of the pairs, 10ms or greater. The overall mean difference is 5.19ms. If there is indeed an effect of frequency on duration for these pairs, it is a relatively small one. A paired *t*-test (see Table 7.3) showed the difference in duration between high and low frequency homophones was significant ($t(379) = -3.428$, $p < 0.001$).

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	HDuration_ms - LDuration_ms	-6.07709	34.55521	1.77265	-9.56254	-2.59164	-3.428	379	.001

Table 7.3. Paired *t*-test results for Experiment 3, 'carrier condition'.

That the difference in duration by frequency is small is not entirely unexpected. In Gahl's (2008) study on English homophone durations, high frequency words were found to be on average 28ms shorter than their low frequency homophones. Gahl's (2008) study analysed a corpus of spontaneous speech, recorded from telephone conversations. It has been noted that words shorten in duration when they are predictable from context (Jurafsky et al, 2001a, b), and this may account to some extent for the even smaller effect found in our 'carrier condition', which provided running speech context, but no semantic or syntactic context.

Repetition

A visual summary of the results for duration by repetition for the carrier condition is given in Figure 7.1.

Figure 7.1. Mean responses to pairs with a homophone density of greater than or equal to 5 (top) and fewer than 5 (bottom). The numbered columns represent the mean duration for each word. Homophone pairs share a number - 'a' indicates the higher frequency character of the pair, and 'b' the lower frequency character. The blue bar in each set of 3 columns represents the mean duration for the first repetition of the character, the green bar repetition 2, and the yellow bar repetition 3. The pairs are arranged in descending full homophone density (left to right).

For the first repetition of the characters in a given pair, the higher frequency character is generally of shorter duration than the lower frequency character. This can be seen in Figure 7.1, where the blue column for the 'a' set of a given number (representing the first occurrence of the target character) is shorter than the blue column of the corresponding 'b' set (e.g. the blue column in 1(a), representing the mean duration of the first occurrence of the higher frequency homophone of *shì* 是, is shorter than the blue column in 1(b), which is the first

occurrence of the lower frequency character *shì*, 侍). This indicates that, for the first production of the characters at least, the higher frequency homophones are of shorter duration than their lower frequency mates, as expected. It was also expected that the duration of characters would shorten with repetition, given previous findings of this nature (Bard et al, 2000). This would present in Figure 7.1 as the green bar in a set being shorter than the blue bar, and the yellow being shorter than both (this can be seen in 4(b), for example).

For characters with fewer than 5 full homophones, there is a consistent trend of duration decreasing with repetition, even if in most cases these differences are small and likely to be below the perceptual threshold. For characters with 5 or more full homophones, only the lower frequency homophone of each pair exhibits this pattern. The high density, higher frequency characters generally show a slight reduction in duration for the second occurrence, and a *longer* duration for the third occurrence (compare 9(a) from Figure 7.1, which shows this characteristic 'U' shape, to its lower frequency counterpart 9(b), which shows a regular reduction in duration with repetition). Low frequency, high density characters generally have the longest durations, and shorten the most with repetition.

Homophone density appears to affect phonetic realisation in the following way: high frequency, high homophone density characters have short durations, and do not consistently shorten with repetition. These characters appear resistant to the effects of repetition, presumably because they are already fairly routinised productions by the first repetition. As high frequency characters they are expected to have independently strong representations, and having many full homophones is expected to accentuate this strength (in terms of speed of access, as found in Experiment 2, as well as in production, resulting in greater reduction). High frequency, high density characters may not exhibit an effect of repetition on duration because they are already highly streamlined in terms of both lexical access and production.

High frequency, low density characters exhibit some repetition-induced shortening (see 2(a), for example). Such characters are of high frequency, and therefore have strong independent

representations. However, they do not benefit from a dense 'full homophone neighbourhood'. Repetition is still capable of affecting duration in these cases, as production has not been maximally streamlined by the characters' distributional statistics (e.g. character frequency and full homophone density). High frequency, high density characters appear to be 'strong' for two reasons - primarily, they are strong because they are high frequency, but they also receive some benefit from being in high density full homophone cohorts. Low frequency, high density characters have the longest durations and shorten the most with repetition, implying that they have the 'weakest' representations.

3.2 Sentence Condition

The results for the 'Sentence Condition' are summarised in Table 7.4.

Pair No.	Syllable	Character	Log10 Freq	Full Homophone Density	Rep. 1 Duration	Rep. 2 Duration	Rep. 3 Duration	Mean Duration	Rep. Diff. (3-1)	Mean Duration Difference (H-L)
1	shì	是	-1.87	25	147.00	135.69	134.47	139.05	-12.53	-85.70
		侍	-4.36	25	243.77	218.75	211.72	224.75	-32.04	
2	rén	人	-2.01	2	249.59	231.07	234.96	238.54	-14.63	27.92
		仁	-4.05	2	217.28	209.83	204.75	210.62	-12.53	
3	lǐ	里	-2.55	5	221.57	228.47	213.97	221.34	-7.60	16.59
		鲤	-5.11	5	210.97	201.02	202.24	204.75	-8.73	
4	ài	爱	-3.25	4	137.49	136.68	129.42	134.53	-8.07	-3.49
		隘	-5.09	4	145.73	133.62	134.69	138.02	-11.04	
5	chū	出	-2.41	2	209.42	192.74	179.09	193.75	-30.33	11.09
		初	-3.52	2	190.86	180.96	176.14	182.66	-14.71	
6	duì	对	-2.44	3	135.70	122.03	121.68	126.47	-14.02	-2.18
		兑	-4.67	3	130.48	126.57	128.89	128.65	-1.58	
7	jī	机	-2.75	20	118.49	110.04	108.99	112.51	-9.50	-8.40
		圾	-4.79	20	120.77	118.35	123.60	120.91	2.83	
8	zhī	之	-2.50	10	127.35	121.05	119.16	122.52	-8.19	-30.49
		汁	-4.73	10	160.04	151.47	147.53	153.01	-12.51	
9	xīn	心	-2.69	7	290.74	244.90	240.50	258.71	-50.24	14.63
		馨	-4.77	7	259.31	236.48	236.46	244.08	-22.85	
10	wú	无	-2.73	5	150.96	130.88	130.32	137.39	-20.63	5.36
		梧	-5.20	5	134.88	129.80	131.39	132.03	-3.49	

Table 7.4. Mean durations (ms) for target characters in the sentence condition of Experiment 3. The 'Rep. Diff. (3-1)' column gives a simple subtraction of the third occurrence from the first - a negative number indicates that the third occurrence is shorter than the first. The 'Mean Duration Difference (H-L)' column provides the difference between the high frequency and low frequency characters in a pair - a negative number indicates that the high frequency character is shorter in duration than the low frequency character.

Previous research has indicated that word frequency does not reliably affect duration in isolation or carrier phrases (Gahl et al, 2012; Bard et al, 2000). Given that we found a small but systematic effect for words in neutral carrier sentences, at least for the first repetition, and also in the presence vs absence of repetition effects, we should therefore expect to find a greater effect of frequency on duration in the sentence condition. However, in the sentence condition only 5 of the 10 pairs of characters exhibit the expected shortening of duration for higher frequency characters (negative duration in the 'Mean Duration Difference' column, Table 7.5). The range of variation, when compared to the 'carrier condition' is, however, much greater. Recall that the largest mean negative difference between a high frequency character's duration and its low frequency homophone counterpart (meaning the high frequency character was shorter than the low frequency character) was -14.59ms, while the largest mean positive difference was 4.43ms. In the 'sentence condition', the largest mean negative difference between a high and low frequency character pair is -85.7ms, while the largest positive difference is 27.92ms. Mean duration differences between high and low frequency homophones are just generally more pronounced in the sentence condition, both negatively and positively. The mean duration difference within a homophone pair is -5.47ms however, quite similar to the carrier condition's -5.19ms. A paired *t*-test (see Table 7.5) showed the difference in duration between high and low frequency homophones was significant ($t(378) = -2.49$, $p < 0.05$).

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1 HDuration - LDuration	-5.4590448	42.6815627	2.19240461	-9.7698815	-1.1482081	-2.490	378	.013

Table 7.5. Paired *t*-test results for Experiment 3, 'sentence condition'.

Repetition

A visual summary of the results for duration by repetition for the sentence condition is given in Figure 7.2.

Figure 7.2. Mean responses to pairs with a homophone density of greater than or equal to 5 (top) and fewer than 5 (bottom). The numbered columns represent the mean duration for each word. Homophone pairs share a number - 'a' indicates the higher frequency character of the pair, and 'b' the lower frequency character. The blue bar in each set of 3 columns represents the mean duration for the first repetition of the character, the green bar repetition 2, and the yellow bar repetition 3. The pairs are arranged in terms of descending full homophone density (left to right).

When compared to the same set of results from the carrier condition, the pattern for the sentence condition is more variable, and less clear. The characters generally exhibit shortening due to repetition, unlike in the carrier condition where high frequency, high density characters were somewhat immune to this effect. It is also apparent from the figure

that for a number of the homophone pairs, the higher frequency character is not shorter in duration than its low frequency counterpart. This is a somewhat surprising result. In the carrier condition, the relationship between high frequency and reductions in duration was more consistent. Given previous research (Bard et al, 2000; Gahl et al, 2012) we expected the opposite, with a greater effect of frequency on duration for characters in the sentence condition.

4 Discussion

Based on previous studies showing a relationship between lexical frequency and word duration for pairs of homophones (Gahl, 2008), and effects of phonological neighbourhood density on word duration more generally (as summarised in Gahl et al, 2012) in production in English, and related findings of the effects of frequency and homophone density on lexical retrieval in Experiment 2, it was expected that increasing lexical frequency, and to some extent also increasing homophone density, would have shortening effects on duration in Mandarin, particularly in the full sentence condition. These expectations were borne out, but only systematically so for the neutral carrier condition. For homophone pairs, characters of higher frequency were consistently, significantly shorter in duration than their lower frequency counterparts, especially on their first repetition, and were resistant to repetition effects. The picture for the sentence condition is less clear. Where a frequency-related difference was found, it was typically large when compared to the carrier condition. For example, the high frequency *shì* character 是 was 10.59ms shorter in duration than its low frequency counterpart 侍 in the carrier condition (mean duration 313ms and 323ms, respectively). In the sentence condition, not only was the mean duration for all *shì* characters shorter than in the carrier condition (139ms and 225ms, respectively), the high frequency *shì* character was 85.7ms shorter than its low frequency counterpart.

It is possible that the confused effect observed in the sentence condition is due to confounding factors in the stimuli. In creating the sentences there are a considerable number of factors to balance. For example, it was considered important to keep the target characters for each pair within the same immediate phonetic context, to avoid confounds caused by coarticulation both in terms of segments and in tone. This can actually be fairly difficult, for a number of reasons. There is substantial compounding in Mandarin, and different homophones will form compounds with different syllables than their mates. An example from Experiment 3 is *xīn* (high frequency character 心; low frequency character 馨). In this case, the immediate phonetic context is the same in terms of segments and tones (preceded by yòng and followed by xiāng). The preceding character is even the same in both (用). However, 心 combines with the the preceding syllable to form the compound 用心 'carefully', while 馨 forms a compound with the following character 馨香 'fragrant'. This is potentially problematic in general, as stress can interact with position in a compound, and affect duration (see Moore, 1993). It is also problematic in that there is a part of speech difference (another potential confound).

Even if a suitable bimorphemic-compound homophone pair is located with the same part of speech, it can be difficult to insert both into a natural-sounding sentence with the same phonetic context - particularly with nouns, which may require different measure words (an example from **Chapter II** is 一只水獭 *yī zuò shuǐ tǎ* 'one MW (buildings) water tower'; vs. 一座水塔 *yī zhī shuǐ tǎ* 'one MW (animals) water otter').

When trying to keep all of these factors controlled, consistent phrasal position can become difficult to maintain. An example from Experiment 3 is the *lǐ* pair. The low frequency character 鲤 occurs in a compound, and as the first syllable in the sentence, while the high frequency 里 character is not in a compound, and occurs at the end of a phrase. Phrasal position has been found to affect duration in Mandarin as well (Shih, 1988).

In addition, the experiment tried to utilise character pairs of different frequencies, and with a wide variety of full and segmental homophone densities. While it may be easier to balance most of the potential confounding factors for a syllable with many homophones to choose from such as *shì*, which has 25 full homophones, we are still constrained by the frequencies of those characters. If the goal is to compare characters within certain frequency ranges, there may be quite limited options, even for a syllable such as *shì*. It can be very difficult indeed for a limited syllable like *duì*, which has a full homophone density of only 3.

Finally, there is the issue of contextual probability. It may be technically possible to form a pair of sentences and account for all of the preceding confounding factors, but the results may be somewhat unusual. For example, the gloss for one sentence used in the experiment is, 'This has changed the way people handle **rubbish**.' Compare this to its more sedate counterpart, 'I think the list price of this **tractor** is too high.' Although it is potentially amusing to imagine the breakthrough in garbage handling that occasioned the first sentence, it is difficult to claim that it is a very likely utterance.

The endeavour of creating balanced sentences for homophone pairs seems fairly complex when compared to a similar task for English. It is possible, however, that this relative difficulty in creating similar syntactic or semantic sentences for homophones is in some way related to the relatively high morphemic ambiguity in Mandarin.

It is difficult to explain, in an abstract model, why two characters conventionally considered to be homophones would systematically differ in terms of duration (and possibly other phonetic indicators such as degree of vowel dispersion, though these were not examined in Experiment 3). If all homophonic lemmas link to a single phonological representation, which is then used as the basis for phonetic production, it is unclear why production would differ by lemma and not, say, only by the phonetic or prosodic context in which the character occurs (where we would naturally expect differing productions as a result of coarticulatory effects or different degrees of prosodic prominence, for example). In Experiment 3, the

immediate phonetic context for both high and low frequency homophones was identical, yet differences in production remain. This suggests that high and low frequency homophones in Mandarin cannot be said to have precisely the same phonological representation.

In a usage-based framework, such frequency-based differences in production are more easily incorporated into our model by assuming that each lemma is associated with its own exemplar cloud. We expect that the exemplars for two lemmas conventionally considered homophones will be substantially similar, and therefore the two clouds will overlap to a large extent. For two homophonic lemmas of low frequency, the exemplars may be effectively identical. A high frequency lemma will have many more exemplars, and be subject to greater routinisation, resulting in shorter productions. Experienced tokens for a high frequency lemma will also be shorter in duration, skewing the distribution of the exemplar cloud, and allowing for systematic differences in production.

The predictions made by abstract and usage-based models for homophony, and how we might best incorporate the findings of Experiments 1-3 in a particular account of homophony, are discussed more fully in the next chapter.

Chapter VIII. Discussion

In this thesis, we examined the structural and distributional properties of homophony in Mandarin, and how these properties affect lexical processing and production. The results of our investigations help to address some of the questions outlined in the first chapters of this work: 1) how is homophony distributed in Mandarin, and does this distribution reflect how the presence of homophones affects speech, either in the lexical access of characters or in fine phonetic detail in production?; 2) what is the status of tone in Mandarin homophony, and how does it affect lexical access?; 3) how are homophones represented in the Mandarin lexicon, and what might this reveal about lexical representation more generally?

1 Distribution of Mandarin Homophony

Given the extensive nature of homophony in Mandarin, one of the objectives of this thesis was first to investigate its distribution, and potentially gain insight into how the lexicon is structured. It was found that homophony is more densely clustered than it structurally needs to be. There are approximately 400 unique base syllables in Mandarin, when stripped of tonal distinctions. As the number of unique segmental strings is small, it would be reasonable to expect tone to be maximally exploited by having each syllable paradigm fully utilised with respect to syllable and tone combinations - that is, given the 400 unique base syllables and the four lexical tones, we would expect around 1,600 tonally specified unique syllables if the morphological and phonotactic possibilities were exploited to the full. Duanmu (2007) gives the maximum total number of tonally specified syllables actually occurring in Mandarin at 1,255. In analysing a large written corpus (Da, 2007), in which low frequency words were excluded, we found 399 unique base syllables, but only 1089 unique tonally specified syllables. This syllable inventory is quite limited cross-linguistically - as an example, English

has around 9,000 unique syllables (Duanmu, 2007). The constraints of this limited syllable inventory are further compounded by the fact that Mandarin morphemes are almost exclusively monosyllabic, rendering the possible pool of unique morphemes much reduced. In contrast, English not only has many more possible syllable strings, but also freely allows polysyllabic morphemes. Other languages with even smaller syllable inventories than Mandarin also fare better in this respect because of the abundance of polysyllabic morphemes, e.g. Hawaiian, which has only around 200 syllables, but where most morphemes are two syllables (Krupa, 1971) (yielding, potentially, some 40,000 unique disyllabic combinations for distinguishing morphemes). There is, as has been reiterated a number of times throughout this thesis, remarkable pressure in Mandarin for extensive homophony among morphemes, which means, given the lemma-like status of monosyllabic morphemes in Mandarin, extensive homophony among lemmas. In our abridged corpus, there are on average 3.2 homophones per tonally specified unique syllable, and since homophony is far from evenly distributed, homophony density can be much higher than this.

At first glance it might be thought that tone is the critical disambiguating factor, and indeed is often described as critically distinguishing morphemes (see Winskel et al, 2016 for an example). However, with even the most superficial of analyses it is immediately clear that tone in Mandarin fails to perform this role very effectively, and certainly not consistently or conclusively. Most Mandarin morphemes have homophones, even with tone. From the outside, this would seem to present a problem for Mandarin speakers. Given such extensive homophony, for most syllables, how is effective communication possible? If we think of homophony as a source of potential ambiguity, and ambiguity as placing a burden on lexical processing, any model of word recognition valid for Mandarin must be capable of explaining how this degree of homophone density is accommodated and structured.

In order to devise, or evaluate, models of Mandarin homophony, a more profound understanding of the nature of this homophony must first be forged. This was the purpose of

Chapter III. While references provide general information about the nature and distribution of homophony in Mandarin, for example that it is extensive and sometimes uneven (see Duanmu, 2007), to our knowledge no investigation of a more detailed nature had previously been carried out. Our analysis found a number of curious characteristics in homophone distribution and other structural properties.

First, it was found that a roughly comparable number of characters are associated with Tones 1 and 2, while rather fewer are associated with Tone 3, and rather more with Tone 4 (see **Chapter III**, *4.2 Distribution of Tone*). There are only 12 characters in the corpus that are associated with Tone 5, and these are exclusively grammatical particles, interjections, and onomatopoeia. The finding that Tone 4 characters are the most numerous, and Tone 3 characters the least numerous, is directly mirrored in child acquisition of Mandarin tone, where Tone 4 has been found to be acquired earliest in development, and Tone 3 latest (Wong, 2013). It is speculated that this asymmetry may be grounded at least in part in phonetic naturalness. Tone 4 consists of a falling pitch. All other things being equal, and if no effort to the contrary is made, during an utterance pitch is known naturally to decline, with vocal fold vibration decelerating as the sub-glottal/supra-glottal air pressure difference decreases on the out-breath. There are, therefore, grounds for assuming this is an 'easier' tone for infants to acquire (at least in terms of laryngeal coordination), and perhaps more resistant to erosion and change, and thus more common in the Mandarin lexicon. Tone 3, on the other hand, contains both a falling and a rising component, and is therefore more phonetically complex and arguably less 'natural' (and more difficult). The incidence of different tones in the lexicon would appear then to reflect relative difficulties in child acquisition.

The analysis also makes clear that not only is Mandarin homophony unevenly distributed for certain syllables - it is *grossly* and *systematically* uneven. There are a number of unique base syllables (i.e. disregarding tone) that have only one associated character. On the other end of the scale, there are syllables like *ji*, which occurs with all four tones, and has

a total of 54 associated characters. Not only is there great disparity in homophony, but there is a tendency for base syllables with all four tones to have *more* full homophones associated with each tone.

When considering the lexicon as a whole, the full tonal paradigm of four tones is largely under-utilised. On average, fewer than 3 tones are associated with each segmentally unique syllable, but that is not to say that the majority of unique base syllables have around 3 associated tones, indeed the distribution is far from even. Of the 399 unique base syllables in the corpus, a surprising 59 have only one associated tone. For such syllables, tone is essentially superfluous for lexical access - there are no tonal minimal pairs for around 15% of all unique Mandarin base syllables. Nevertheless there is still homophony for many of these syllables. The unique base syllable *mie*, for example, has only one associated tone in the corpus (tone 4), *miè*. There are two *miè* characters associated with this syllable, 灭 and 蔑 (i.e. a homophone density of 2). In a sense, this homophony is, structurally speaking, avoidable - there are three tonally specified syllables (*miē*, *mié*, and *miě*) with no characters (lemmas) at all (not to mention no homophones). In other words, the full spectrum of syllable + tone possibilities is underused, while particular combinations are overused. In addition, some of these mono-tonal syllables have only a single associated character (recall that the average for the corpus is 3.2 characters per tonally specified syllable). For example, only one tone (tone 4) occurs with the syllable *miu*, and only one character occurs with the syllable *miù:* 谬 'absurd'. For four potential syllables (*miū, miú, miǔ*, and *miù*), only one is utilised, and that only has a single character. There are 33 unique base syllables in the corpus with only one associated character. That is, for a language in which extensive homophony is inevitable, some 8.3% of all unique base syllables (33 out of 399) are utilised for only 33 morphemes. In other words, there are 33 characters which not only have no full homophones, but also have no segmental homophones - an astonishing 'waste' of morpheme-level contrast in a language with such extensive homophony. This is in contrast to a syllable like *ji*, which

occurs with all four tones, each of which has a number of characters - *jī* is associated with 20 characters; *jí*, 14; *jǐ*, 3; and *jì*, 17. The difference between these two syllables (*miu* and *ji*), illustrating as it does the great unevenness in the relationship between characters/morphemes, segmental strings, and tones in the lexicon, is perhaps surprising - and results in even denser concentrations of homophony than are structurally necessary.

Some degree of unevenness in homophone distribution is to be expected, and disparities in homophone density between particular syllables have been noted by previous researchers (e.g. Duanmu, 2007). We found however, perhaps surprisingly, that homophone distribution was not just uneven, but systematically so. From a starting point of a restricted pool of phonological distinctiveness (low number of base syllables), morphemes are unnecessarily 'squashed' into an even smaller sub-pool, and yet even then homophony is more dense than it needs to be. The more tones associated with a unique base syllable, the more homophones associated with each related tonally-specified syllable - segmental homophony and full homophony appear to interact (see **Chapter III** for full details). The mean full homophone density in the corpus is 3.2 homophones per tonally specified syllable. As mentioned, 59 of the 399 unique base syllables are associated with only one tone, so if characters were spread evenly we would expect to find a total of around 189 characters associated with these syllables (i.e. based on an average of 3.2 characters per mono-tonal base syllable). In fact, there are only 123 characters associated with these 59 syllables in the corpus. Thus, not only do these syllables under-employ tonal distinctions but they also attract less homophony. Meanwhile, there are 128 syllables with a complete tonal paradigm (i.e. syllables in which all four tones are represented in the corpus). We would therefore expect around 1638 characters to be associated with these syllables (at 3.2 characters per syllable). There are actually 1834 characters associated with these syllables in the corpus. Thus, not only do these syllables fully employ tonal distinctions, they also attract more homophony. In other words, base syllables without a full complement of tones have fewer than the expected

number of characters (homophony), while syllables with the full paradigm have more. This may be surprising. Recall that tonal contrasts in Mandarin have been generally found to be more difficult to process than segmental contrasts for native speakers. Syllables differing only in tone are, at some level, more difficult to judge as mismatching than those differing in segments, and are closer to being 'the same' (i.e. homophones).

Those syllables with a full tonal paradigm are arguably already in a quasi-homophonic relationship with the rest of the paradigm, *and* compound this by having more full homophones too. For a tonally specified syllable with 3 homophones, having a 'tonal neighbour', a syllable with the same segments and a different tone, which also has 3 homophones, should then result in some degree of competition. To successfully identify a particular morpheme, there is competition between three full homophones, as well as three segmental homophones (which are substantially similar, and cause some degree of competition during lexical access). For a character which has a base syllable with all four tones, this difficulty is further compounded, as there are three tonal neighbours, each with 3 full homophones. Characters which are associated with tonally-specified syllables that have more 'tonal neighbours' (i.e. syllables with more tonal minimal pairs) should experience greater inhibition due to increased competition. We might then suspect that, as the number of tones associated with a base syllable increases, the number of tolerable full homophones for the average tonally specified syllable would go down. Taking the mean full homophone density for Mandarin of 3.2, we might then predict that base syllables with all four tones have 10 associated characters (e.g. only 2.5 full homophones per syllable, less than the mean), while base syllables with only one associated tone have 4 associated characters (4 full homophones per syllable, more than the mean). In fact we find the opposite - there are 2.08 characters associated with mono-tonal syllables in the corpus, and 3.58 characters per tonally-specified syllable with unique base syllables having the full complement of tones. This distributional pattern should lead us to conclude that the presence of 'competitors' for a

particular character, both full *and* segmental homophones, is not particularly problematic and, indeed, may even confer benefits in lexical access, production, or both.

If homophony were to pose a significant obstacle to lexical access, it would be difficult to imagine how such a situation of extensive and (apparently) unnecessarily densely clustered homophony could have arisen historically. In a play-off between the pressures leading to the loss of syllable types and the reduction of tones, and the pressure to retain sufficient linguistic contrast, if the loss of overt contrasts had been too catastrophic, one assumes the phonetic erosion would not have been allowed to exert such extreme influence in the first place.

That there is such extensive homophony at all indicates that having homophones is not hugely costly for communication. That there appear to be systematic biases, structurally, toward condensed homophony within certain parts of the lexicon, indicates that there may even be processing advantages or efficiencies that come from having high homophone density in certain parts of the lexicon. Evidence from the three experiments conducted for this thesis would appear to support this conclusion insofar as they show homophone density to exert an enabling influence in certain aspects of lexical access and production, even while character frequency remained by far the single most important actor in these processes.

For example, in Experiment 1 (free response), the most frequent response prompted by a given unique base syllable (e.g. *yi*) was generally the highest frequency character, regardless of full homophone density. However, while character frequency was found to be largely determinative of participant responses, some interesting details emerge. For example, for the syllable *wan*, the two highest frequency characters were both very close in frequency, but of differing full homophone density. The highest frequency *wan* character, 完 *wán* (frequency: 743; full homophone density: 2) was not given as a response by any subjects, while the second highest frequency character, 万 *wàn* (frequency: 697; full homophone density: 4) was given by 64% of subjects. At least in terms of character selection, when two

characters of similar frequency were in competition, having *more* homophones appears to confer some advantage, suggesting the overall frequency of the phonological form, or forms with very close associations, plays a supplementary role.

There is no doubt that character frequency is the single most important factor in disambiguating homophones in a neutral context. In Experiment 2 (testing reaction times), again, a strong effect of lexical frequency was found, with characters of higher frequency responded to faster than those of lower frequency. Density was also again seen to exert a subtle influence. It had been predicted that increasing full homophone density would facilitate responses to primarily lower frequency characters (through inheritance effects), with a smaller, or even non-existent, effect of homophone density on higher frequency characters, as such characters benefit less relative to their own (already high) frequency from having more homophones. Somewhat surprisingly, as full homophone density increased, response times were found to decrease *regardless* of frequency category. Low frequency, high density characters were found to have generally slow response times, though they were faster than characters of equivalently low frequency, but of low homophone density. While characters do appear to benefit from having higher full homophone densities in terms of reduced response times, responses are still 'slower' than for independently high frequency characters. At first pass, we might ascribe these patterns to a form of inheritance effect, with lower frequencies benefitting from the overall higher cumulative frequency of the homophone cohort, as provided by the very high frequency cohort mates. However, such an explanation cannot be applied to density effects on high frequency characters, since the additional effect of adding in the much lower frequencies of their cohort mates is proportionately very small. The influence exerted by a large cohort cannot, therefore, be explained as simple numerical inheritance.

Finally, in Experiment 3 (production study) effects of character frequency and full homophone density were also found. High frequency, high homophone density characters had

the shortest durations overall, and were not systematically shortened by repetition, and thus can be assumed to be readily retrieved and executed without the need for further familiarisation and motor routinisation afforded by subsequent repetitions. If we compare this with high frequency, low homophone density characters, we can identify the same subtle boost afforded by phonetically close (quasi-identical) neighbours as observed in Experiment 2. High frequency, low homophone density characters were somewhat slower overall, and were found to shorten with repetition, showing the absence of such a boost when the cohort is small, while nevertheless retaining the specific 'advantages' afforded by being of high frequency.

This trend, however, was reversed for low frequency characters: low frequency, high density characters were the longest in duration overall, and shortened the most with repetition. Significantly, the combination of low frequency and high density creates the condition of greatest competitive pressure, and we can speculate how it is such competitive pressure that results in an even slower implementation of a low frequency character. Curiously, competition among members of a cohort appears to interfere less or even not at all to high frequency characters, which presents an interesting divergence in the way that frequency and density interact, and furthermore a divergence in effects for lexical access in perception and for production. For high frequency characters, increased density of homophony appears further to aid both recognition/lexical access, and streamline production. However, for low frequency characters, while increased density aids lexical access, it impedes the streamlining of production.

One possible explanation for these findings is that high frequency characters are stored differently but not entirely separately – i.e. they stand out from the homophone crowd in some way – but also receive a boost from the phonological (close-)identity with a larger cohort. We can view them as like salient peaks in an identifiably cohesive mountain range. Low frequency characters make up part of this range, but without a form that is discernibly

distinct from the rest. In an indirect way, this approach captures what could impressionistically be viewed as a hybrid model, with both cohesive phonological identity on the one hand, and usage-related individualism on the other. In a hybrid account such as this, high frequency characters are 'treated' (processed/stored/produced) differently, and speed and routinisation are boosted by having a large backdrop of phonological mates. Low frequency characters form a general pool at the phonological level, and are thus slower to process and less streamlined in production. The implications for this apparently hybrid account, and in particular whether this is best served by abstractionist or phonetically-rich theories of representation are discussed below.

The results of the three experiments generally align with the indications given by the structure of the Mandarin lexicon itself - increasing homophone density confers advantage on characters, across a number of tasks (character selection, match/mismatch, and at least for high frequency characters, also in production). It seems likely, given the processing advantages of increasing homophone density, that the current structural distribution of homophony in Mandarin is not detrimental, and that the lexicon has been shaped in part by such effects. We turn next to *why* unevenness of homophony may confer some processing advantage in Mandarin - what do the distributional properties and experimental results tell us about the structure of the lexicon, and what it means to be a homophone in Mandarin.

2 What is Homophony in Mandarin?

2.1 Full Homophones

The distribution of homophony in the Mandarin lexicon indicated that higher full homophone density confers advantage in lexical recognition and, for high frequency characters at least, also in production. The results of Experiments 1-3 generally confirmed this expectation. It is

not immediately obvious, however, from previous work, why this should be the case. Prevailing models of lexical access and word recognition tend to treat homophony as a problem to be overcome, and not as a particular structural 'landscape' with its own particular characteristics to be exploited or otherwise. To understand why homophony may be beneficial, it is important to first summarise the various ways of thinking about lexical representation and structure.

A strong assumption underlying usage-based models is that the form and structure of the lexicon is shaped by an individual's language experience: frequency of usage is encoded in the representation/lexical organisation. In theories advocating phonetically-rich forms of representation, representation expresses a range of phonetic variation, emerging from a set of near-identical or very similar exemplars. With increasing frequency, the representation of the form grows stronger, and lexical access is easier and faster, and the articulation process is streamlined, resulting in greater phonetic reduction, and propensity to other effects, such as semantic bleaching, over time (Pierrehumbert, 2001; Bybee, 2001).

Within such a model, homophony can be represented as a scalar or gradient phenomenon – with identity expressed by degree. This allows an entire homophone cohort to be more or less loosely associated, but also differentiated along frequency lines. So, within an exemplar cloud of quasi-identical tokens, those associated with the most frequently used characters will also stand out – just as the aforementioned 'peaks' against a cohesive mountain range. They can both benefit from the overall size of the exemplar cloud, and stand apart, i.e. be accessed more quickly, and also produced more quickly and with greater motor dexterity. Lower frequency characters benefit from the overall size of the exemplar cloud, but will take longer to be associated with a particular lemma and less routinised on production.

For models of abstract representation, homophones are represented as sharing a single phonological form. This cannot account for the subtle interaction of frequency and density effects observed in the experiments reported here. It is then a question of how the lemma

units connect to phonological representations. Fine phonetic differences in production (e.g. durational differences) for words considered to be homophonous are relatively easy to incorporate into an exemplar model, by assuming that each lemma has its own exemplar cloud, which overlaps more or less with the clouds of homophones within the larger cloud. In an abstract model, such differences are more difficult to explain. If all 'homophonous' lemmas share a single abstract phonological representation, it is unclear why they would systematically differ in production.

In Experiment 3, we found that high frequency characters were shorter in duration than their low frequency full homophone counterparts, particularly when comparing the first repetition. This is difficult to explain in an abstract, phonetically impoverished model, particularly if assuming that full homophones share a phonological representation. In an exemplar model, the result can be explained as the two characters having largely overlapping yet independent sub-sets of exemplar clouds, composed of similar, but systematically differing (in terms of duration at least) exemplars.

This account may be adequate for a language with low homophone density, such as English, but what is the situation when there are not two, but many homophones, as is frequently the case in Mandarin? Does each lemma have an independent exemplar cloud, or do some (presumably lower frequency) characters share a single cloud? A strong exemplar account would perhaps insist on each character/lemma having a unique representation, as provided by a uniquely discernible exemplar cloud. However, with such large homophone cohorts as we have identified in Mandarin, this is surely implausible, and not supported by the findings reported here. Instead, it is proposed that low frequency members of a homophone cohort form a tightly knit cloud of exemplars which have associations with multiple lemmas (selected on the basis of frequency, or more typically in normal communication, a higher order probabilistic choice informed also by context). These low frequency characters would be treated (processed/produced) in similar ways. This inner cloud of exemplars is in many

respects behaving like a single abstract phonological form: while the basis on which it is built is scalar, or at least gradient, in allowing for the possibility of variation in exemplars, the overlap is so tight that there is de facto phonological identity. The critical difference lies in two important aspects.

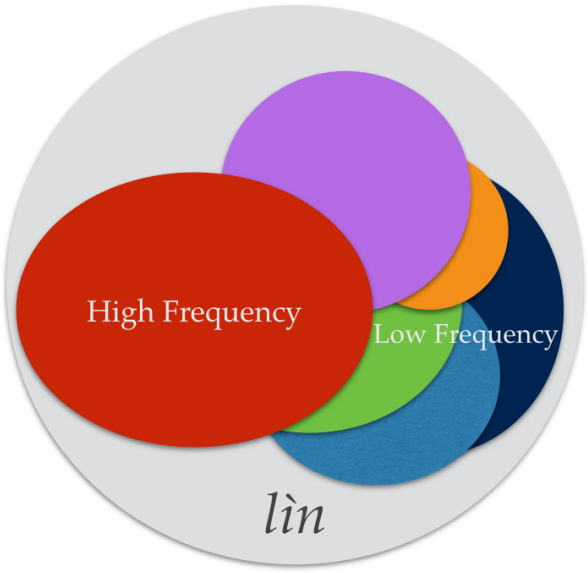


Figure 8.1. Schematic diagram representing *lìn*, with tightly overlapping exemplar sub-clouds for low frequency homophones, and a less overlapping exemplar sub-cloud for a very high frequency homophone in the same cohort (together forming a two-tier exemplar cloud.

First, the underlying scalar nature of this approach allows for an association with the quasi-identical exemplars of a very high frequency homophone: the overlap is not so tight as among lower frequency homophones, there are measurable phonetic distinctions in outputs, but there is nevertheless a high degree of identity (much more than for a mere phonological neighbour). Figure 8.1 provides a simple schematic representation of this concept.

Second, the gradience inherent (but not superficially evident) in this model allows for flexibility in the relationships as usage changes with time. What are tightly overlapping

exemplar sub-clouds could become less overlapping, if the usage of a particular character were to increase. In keeping with exemplar accounts more generally, the approach allows for categorical effects to emerge from what is underlyingly gradient. The novelty of what is being proposed here is that an exemplar cloud can be associated with multiple lemmas, but also differentiated within that cloud.

It is not possible to confidently ascertain, from Experiments 1-3 alone, whether lower frequency characters in a full homophone set share a single representation, or have independent representations, though it would be amenable to further experimental investigation looking at the details of comparable homophone mates across cohorts. Essentially, the question is: do characters pattern in concert with lemma frequency, or cumulative homophone frequency? If two homophonous characters *a* and *b*, with small lemma frequency differences (e.g. 1.2 and 1.3 per million), exhibit frequency-based differences in responses (say in terms of RT), independent representations are indicated. If character *a* patterns similarly to a non-homophonous character *c* from a different homophone cohort (which has the same cumulative frequency of *a*+*b*, 2.5 occurrences per million), then a shared representation for homophones would be indicated.

The fact that low frequency characters with more full homophones were responded to faster in Experiment 2 would seem to indicate shared representations along these lines. However, characters of all frequencies exhibited this effect in Experiment 2, low frequency no more than high frequency characters. Parallel results were found in Experiment 3, where high frequency and high density characters consistently had the shortest durations (and were most resistant to repetition effects), while high frequency low density characters exhibited repetition-induced shortening (indicating, perhaps, that their representations did not provide for maximum streamlining of production, as the high frequency, high density characters).

It is not immediately clear why extremely high frequency characters should benefit from having more full homophones (of lower frequency), for either independent or shared

representation formulations. If the high frequency character shares a representation with its homophones, it should be only marginally affected by the addition of low frequency homophones (e.g. cumulative frequency is not greatly impacted by the addition of proportionally much lower frequency characters). If the high frequency character has an independent representation, why is it affected by the presence of homophones at all? The presence of homophones appears to affect a character, regardless of its frequency, in a way that does not appear to be due simply to sharing a representation (i.e. through cumulative frequency alone). A larger cohort appears to provide a boost which is not to do with actual frequency; the only obvious alternative is that multiple lemma associations boost activation (and processing/production) in the high frequency character. Low frequency characters are also boosted by more lemma associations, but interestingly these multiple associations confer an inhibitive effect in production, presumably because a low frequency character suffers in a competitive pecking order. It is not clear why these competition effects are evident in production but not in lexical processing on the receptive side (though this point is addressed in the final chapter).

2.2 Segmental Homophones

There is evidence that tone is processed and organised differently than segmental material in Mandarin. It has been shown that the perception of tones is slower and more error prone than for vowels, and that tonal information is not fully available as quickly as segmental information, and cannot be used for word disambiguation early in the recognition process (Ye and Connine, 1999; Cutler and Chen, 1997; Taft and Chen, 1992; Lee, 2007). In our Experiment 2, subjects were asked whether a *pinyin* stimulus and a character matched. For mismatches, when the segments of the target failed to match those of the *pinyin* stimulus, regardless of tone, participants were both faster and more accurate in rejecting the pair as a match. Characters which completely mismatched the stimulus, and those which mismatched

in segments but had matching tone, were indistinguishable in response time and accuracy, while responses to tonal mismatches were both slower and more error prone. It was more difficult for Mandarin speakers to classify segmental homophones as mismatches than other types of mismatches. From this accumulated evidence, we may conclude that tone is not weighted equivalently to segmental information in Mandarin lexical processing.

This has implications for the definition of the relationship between homophones and near homophones, the role of tone, full homophone density and neighbourhood density, and how these affect lexical processing in Mandarin. There appears to be an intermediary level between full homophones and phonological neighbours, called *segmental homophones* throughout this thesis, which are segmentally identical, but differ in tone. Is the apparent existence of this level due to the fact that tone is phonologically separate from segments (i.e. analysed on a separate tier in an abstractionist view), or because differences in tone are phonetically 'less different' than differences in segments?

For theories of abstract representation, a gradient judgment of homophony might be explained as activating either both segmental and tonal tiers (full homophony) or just the segmental tier (near homophony). Abstract phonological treatments of tone tend to classify tone distinctions in Mandarin as equivalent to phoneme distinctions in importance. Ye and Connine (1999), for example, posit a separate toneme tier in a TRACE architecture, in addition to the feature, phoneme, and lexical tiers.

If tone is less critical to lexical identity than segments, we expect segmental homophones to be 'closer' in some sense than (segmental) phonological neighbours. How is this reconcilable with usage-based models of lexical representation? If tone is less salient than segments in some sense, this could lead to tokens with the same segmental composition (but differing tones) being 'mapped' closer together in overlapping exemplar clouds, than tokens differing in segments. The phonetic overlap in tokens for the segmental homophones 豌 (wān;

'pea') and 碗 (wǎn; 'bowl') is extensive, but phonetic identity is compromised by different F_0 patterns.

For theories of phonetically-rich representation (e.g. Exemplar Theory) a gradient judgment would have to be able to differentiate between exemplar clouds which acoustically match/over-lap in a holistic way (i.e. do not discriminate between the phonetic exponents of the hypothetical segmental and tonal tiers), and exemplar clouds which acoustically match/over-lap only for the phonetic exponent of the hypothetical segmental tier. This potentially causes a problem for Exemplar Theory, as it is not immediately clear how these different kinds of phonetic information could be discriminated without some form of abstraction being involved. In theory we could posit a differentiation between F_0 and all other acoustic information, though we know that the phonetic expression of tone is not limited to F_0. If we evaluate acoustic stimuli according to phonetic proximity against an acoustic trace in our memory, then the more complex the mismatch is, the more likely and more quickly it will be rejected. Arguably a segmental difference constitutes a bigger difference (more abrupt and different along more than one parameter) than a tonal difference. There may also be more 'noise' in the signal with regard to tones, since they are very prone to coarticulatory variation (c.f. tonal sandhi), and in a way that potentially could interfere more with phonetic identity between exemplars than segmental coarticulation, although this remains to be empirically tested. That means the store of traces for a particular lemma may not have as much phonetic cohesion with respect to tonal information, and thus tone is less reliable.

An exemplar approach is, more generally, quite suited to a gradient evaluation of homophony that is *scalar* (i.e. not a series of different categories, but a continuum of greater or lesser homophony). Thus, the exemplar clouds associated with e.g. 豌 *wān*; 'pea' and 碗 *wǎn*; 'bowl' would overlap to a significant degree, thereby triggering a judgment of likeness; likewise the exemplar clouds associated with syllables which are phonological neighbours. e.g. *wān* and *yān*, would also overlap to a significant degree, thereby also triggering a

judgment of likeness. Any distinction between these would depend on the degree of likeness involved, which may be quantitatively different (i.e. the difference between two tones may be phonetically less marked than the difference between two segments), hence phonological neighbours may appear to be a different category from segmental homophones, through phonetic accident rather than phonological analysis. An example of how this might work is given in Figure 8.2.

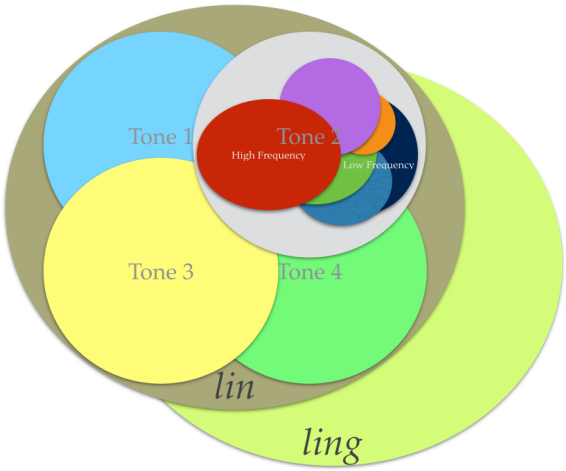


Figure 8.2. Schematic diagram representing overlapping exemplar clouds for segmental homophones, and a less overlapping exemplar cloud for a phonological neighbour.

Figure 8.2 gives an example of how the exemplar clouds for two phonological neighbours (i.e. differing in terms of one segment) overlap. In the diagram, *lin* is superposed on *ling*. We expect that exemplars from both clouds will be judged as being substantially similar, with differences along various phonetic parameters. For example, we might expect a *lìn* (falling tone) exemplar to bear more similarity to *lìng* than to *líng* (rising tone); certainly more than to any *ming* exemplar. On this model, tokens are classified by phonetic parameters (and not lemma identity), and mapped to a 3D space. Each exemplar is connected to its lemma, but is affected by adjacent exemplars in concert with phonetic distance. For example, on encountering a token of *lìn*, all *lìn* exemplars are activated, particularly those bearing most

similarity to the encountered token (which might have a skewed, more distinct distribution as a high frequency character). Concurrently, all other *lin* exemplars activate in concert with how similar they are to the encountered token, followed (in terms of activation level) by all *ling* characters, and so on (see Figure 8.3).

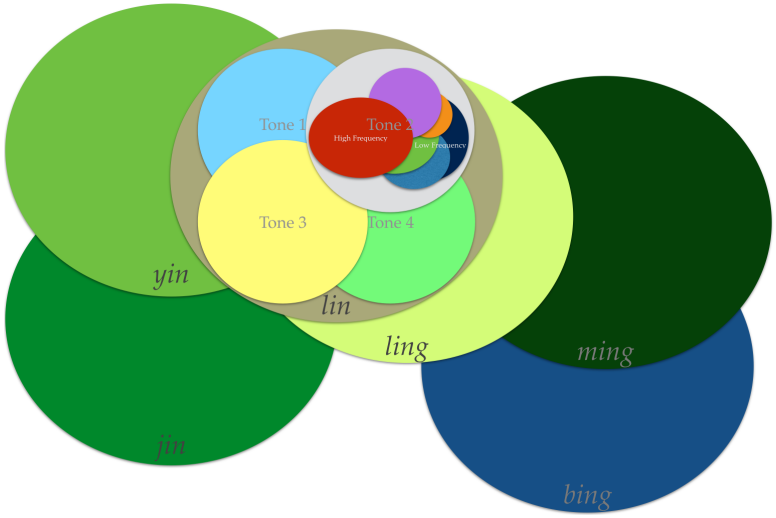


Figure 8.3. Overlapping exemplar clouds for full homophones, segmental homophones, and phonological neighbours.

All of the lexicon, on this model, can be conceived of as an enormous, loosely associated cloud, and to adopt an astronomical metaphor, along the lines of a galaxy. Individual bodies exert influence (gravity) on each other due to proximity, modulated by size (i.e. frequency). The solar system constitutes a cluster (full homophones), with a number of large bodies, and many smaller bodies. This cluster interacts with other nearby clusters (segmental homophones), which can be pictured as nearby solar systems. This supercluster (segmental homophones) interacts with nearby superclusters (phonological neighbours), etc. Gravitational effects extend across vast distances, but the analogy breaks down in that there is not usually actual overlap in the planets of different solar systems (for example). Encountered tokens are given coordinates in the map according to various phonetic parameters. A

difference along certain parameters, such as tone, results in less coordinate distance than a difference along other parameters, e.g. vowel quality.

Similarity in phonological form amongst lemmas is not necessarily inhibitive (though this depends on task). We found in this thesis that, at least in perception, the more lemmas sharing a similar phonetic form to the target, the more facilitation in recognising the target (in terms of response times). This mirrors the distributional structure of Mandarin, in which morphemes cluster both in terms of full homophones and segmental homophones. In our final chapter, we will discuss some of the implications of this.

Chapter IX. Conclusion

Languages with very extensive homophony, such as Mandarin, raise certain challenges for existing models of lexical organisation and access, and indeed also for our understanding of homophony itself. Our investigation into the structure and distribution of homophony in Mandarin suggests that homophony that is deep (i.e. very high density), extensive (i.e. involving much of the lexicon) and unevenly distributed (structural clustering) is not only *not* a cumbersome obstacle to communication, but may even confer some functional advantage to lexical organisation. The analysis of the homophonic lexicon revealed a systematic bias towards syllables with high full and segmental homophone density, with syllables bearing the full range of tones attracting more homophones. In a series of experiments, we found that this structural bias was mirrored in language behaviour, with characters having high homophone density exhibiting speeded response times and (at least for higher frequency characters) shorter productions. The conceptual apparatus provided by classic models tends to frame homophony – whether by accident or design - as a processing burden to overcome, through frequency-ordered lists or competition for activation. We put forward two tentative, usage-based approaches where both inhibition in production (for low frequency characters) and facilitation in lexical recognition, are incorporated - though other explanations are, doubtlessly, possible.

The first approach takes both lexical access and production to be two-stage processes, and in this respect is similar to the two-step interactive model of lexical access as proposed by Dell, 1986, and Dell et al, 1997 (in the context of phonological neighbourhoods), which posits a 'lemma' level and a phonological level, with activation able to flow in both directions. In languages such as English (which have little homophony), high density phonological neighbourhoods have been shown to inhibit access, and this has been explained as arising from increased competition at the lemma level. Why is this not the case for

Mandarin homophones, where activation of a form (or set of quasi-identical forms) should equally trigger a large number of competing lemmas? One potential explanation is that for phonological neighbours (without homophones), competition actually takes place at the phonological level, between slightly different forms (which are then also associated with different lemmas, but this is incidental to the process at this point). Thus, while it has been assumed that the inhibition arises from competition at the lemma level, we could argue that competition is actually occurring at the phonological level ('*what is it that I have heard?* [hat], [kat], [sat], [ðat] etc.?' rather than '*what is the meaning of what I have heard? Hat, cat, sat, that, etc.?')*, which is where the short delay arises, and then once the form has been identified, this directly accesses the correct associated lemma. It is at this second step that relative frequency of the lemma concerned has a further differential effect on processing speed.

For homophones, on the other hand, the process is somewhat different. When the listener/reader hears/reads the stimulus, activating a particular phonological form (or something in the phonetic ballpark, or general exemplar cloud), this 'automatically' places the listener/reader very squarely in the 'right place' on the phonological level, and thus there is no competition at this point. From this point, there are associations with multiple lemmas, hence competition ensues, the outcome of which is determined – in the absence of any context – by lexical frequency. Frequency (and context) directs the listener to the most likely character, as it gives the probabilistically most likely character, so frequency is a critical determiner of speed of lexical access.

So where does density come in? For phonological neighbours, density determines the amount of competitive pressure in getting to the 'right place' on the phonological level, and hence higher densities exert a greater inhibitory effect. With homophones, however, there is no real competition between homophones at the phonological level. However, we not only do not see any inhibitory effect of density for homophones, but a positive boost. In higher

density homophone cohorts, both low and high frequency words are accessed more rapidly. This cannot be simply a question of higher cumulative frequency, activating a larger exemplar cloud, since if that were the case we would not expect higher density to boost activation of high frequency words. The boost must therefore originate from the greater number of *lemma* associations. Homophone density increases the likelihood of the whole phonological form (loosely associated) being an actual word, as a higher density cohort means more lemma associations and therefore a more likely word. Higher homophone density makes a particular form more likely to be a linguistic 'hit', speeding access of said form. This model would explain both frequency and density effects (higher frequency and higher density always speed access).

Given that we have found *greater* concentrations of full homophones in larger segmental homophone cohorts, as well as a lesser role for tone than segmental information in lexical identity, it would be illuminating to investigate whether segmental homophones are similarly beneficial in lexical access - as segmental homophones (syllables only differing in tone) are judged by Mandarin speakers to be quite similar, perhaps higher concentrations of segmental homophones make particular segmental forms (and therefore particular lemmas) more likely than other, (segmentally defined) phonological neighbours.

An alternative explanation for the observed density and frequency effects is to view density and frequency as competing pressures, which combine to affect high and low frequency words in different ways. On this view, encountered tokens are classified in a scalar way, with links to a particular lemma. Very similar tokens form a tightly knit cloud; less similar tokens are more loosely associated. For a syllable which has many homophones, then, we propose a large cloud containing exemplars for a number of lemmas. As the tokens are all very similar phonetically, they cluster together - yet individual tokens are associated only with their respective lemma. When a new token is encountered for a highly dense syllable of this type, two forces interact. First, the new token lights up a large portion of the entire

homophone cloud, as it is quasi-identical to most of the stored exemplars for all the homophones. A high frequency lemma has many more connections to tokens in the cloud which have been activated than a low frequency lemma, and is therefore accessed more quickly (when context does not interfere with interpretation).

The presence of so many activated exemplars enhances the individual activation of every exemplar. The relative frequencies of high frequency characters to lower frequency characters means that low frequency characters experience a greater facilitation due to this mutually-excitatory activation than high frequency characters when interpretation is forced. This aligns with our findings from Experiments 1 and 2. In Experiment 1, when subjects were presented with a syllable and asked to write a character (lemma), the highest frequency character was generally given. This is because, devoid of any additional biases, the highest frequency homophone is most activated by the excitation of the homophone exemplar cloud. In Experiment 2, it was found that responses to low frequency characters were *faster* when homophone density was high. The choice of lemma was forced (in that a specific character was given), but speed of response was facilitated by the presence of many exemplars mutually activating those exemplars linked to the 'correct' lemma. By this account, we should expect responses to high frequency characters in Experiment 2 to experience only minimal facilitation due to occurring in high homophone density neighbourhoods, as the impact of low frequency homophones on the exemplar cloud is relatively lower - yet we found approximately equal speeding of response times for high frequency characters in high density neighbourhoods. One potential explanation for this finding is that, though increasing homophone density is beneficial in one sense, through increasing the size of the homophone exemplar cloud and providing greater activation for exemplars (with low frequency characters benefitting in particular), it is also negative in that it increases competition amongst lemmas. Low frequency lemmas (meaning those associated to fewer exemplars in the homophone cloud), which benefit more from large exemplar clouds, experience greater inhibition from

competition with many other homophonic lemmas than high frequency lemmas. This results in a kind of balance in terms of activation benefits for high and low frequency characters, as homophone density increases. High frequency characters experience marginal gains from existing in a slightly larger exemplar cloud, but as their exemplars dominate that cloud, experience few drawbacks from having many lower frequency homophones. Low frequency characters experience large gains from existing in (relatively) enormous exemplar clouds, but experience more dramatic inhibition from having many similar or higher frequency homophones. The end result is that very large, diverse cohorts are linguistically sustainable. Low frequency characters are not smothered by association with many (higher frequency) characters, but are in fact generally enhanced, and high frequency characters receive a boost as well (perhaps explaining why, historically, they continue to exist in such massive cohorts, and have not 'slipped away' phonetically to form distinct syllables).

The two possible models above attempt to explain density and frequency effects in lexical retrieval of homophones, and in particular why these may differ from those observed for phonological neighbours. But what about production? In production, high frequency was seen to always have an enabling effect (in the sense that articulations are more rapid, and thus probably in some sense more 'routinised'). According to an 'intelligibility based' account (c.f. Clark and Brennan, 1991; Galati and Brennan, 2010), higher frequency words are more predictable (because they are more probable), and can afford to be quicker (provide less phonetic information), while lower frequency words do not have this 'luxury' and may need to be articulated more carefully/slowly. We also know that high frequency words tend to shorten and undergo phonetic erosion more readily (Bell et al, 2009; Bybee, 2001), and the mechanisms underlying this can be described as being 'speaker-internal'.

Hence on the surface, we may home in specifically on the shorter durations of high frequency and high density forms as being due to either speaker-internal processes of articulatory erosion or listener-orientated, or a combination of the two. In other words, the longer

durations of the other words would be seen as the norm, and it is higher predictability that causes a divergence in production. However, in the case of homophones, it is not just frequency but also cohort size (density) that is an issue, with the potential added pressure of having to (or possibly simply choosing to) distinguish between multiple associated target lemmas. We have observed effects not just of frequency (higher frequency = shorter duration) but also of density, with high density inducing even greater shortening (and less shortening from repetition) in high frequency words but conversely even less shortening in low frequency words. Thus a more holistic account, taking into account the whole range of the cohort, is needed.

An alternative interpretation is to consider both extremes, i.e. both the more rapid execution of high frequency words and the more 'laboured' execution of the low frequency, high density words. High frequency vs low frequency regulates probability, but density size, by introducing direct competition effects, also affects probability. It is not so much that high density facilitates production (which it arguably does so only for high frequency words), but that the difference between high and low frequency words in a density cohort facilitates *differentiation*, by generating different variants at the two probability extremes. This would explain why density had a different effect on high and low frequency words in Experiment 3. The underlying motivation (speaker-driven intelligibility for the sake of the listener) is nevertheless a unified one. The higher the density of the cohort, the greater pressure there is to differentiate between forms that will compete in the listener's 'ear'. With duration, there are two options - either a rapid form or a slow form. Faster forms are associated with higher probability resulting from higher frequency, but it is also a way of making a more effective differentiation within a larger cohort. Thus, high frequency words are fast, and when in a high density cohort they are even faster. Lower frequency words then also become distinct by virtue of not being fast, and in a high density cohort this is exaggerated even more, and they become particularly slow. These durational differences may emerge 'naturally' from

frequency effects - more repeated words will be phonetically routinised and eroded while less repeated words will not be - but then these differences are exaggerated when competition pressures increase. We might say that frequency effects are speaker-internal processes, and density effects are motivated by a desire for intelligibility (i.e. for the benefit of the listener).

By working through the possible interpretations for lexical access and production, it becomes clear that we get quite different results for lexical access and production because there are different pressures involved. In lexical access, the shared nature (whether identical or greatly overlapping) of the form is not especially problematic, and things are decided by frequency (probability) of a particular lemma/character. Having more lemma associations (higher homophone density) provides a boost overall from the higher probability of meaningfulness. In production, higher frequency means a more streamlined production, which results from speaker-internal motor short-cuts - this is principally about commonness of a particular lemma being produced. Higher density warns the speaker of greater problems for the listener and s/he seeks to differentiate forms even more by increasing the speed of high frequency words and decreasing the speed of low frequency words. There is an active awareness of the (quasi)-shared form, and production strategy is attuned to the acoustic output. On this account, we have a mix of speaker-internal and listener-orientated strategies.

Experiment 2 only examined one character each from a large number of tonally specified syllables. In the future, it would be useful to examine entire syllables, those with both high and low full homophone density, and across a range of frequencies. This could show whether each character has its own representation, or at least set of motor commands (does the lexical frequency for every constituent character affect responses, or is there a high frequency and a low frequency category?). The results of this experiment would feed into further production experiments - if each character appears to generate responses based on its own independent lexical frequency, how is each character produced? The findings of Experiments 3 indicate that lemmas in Mandarin are affected in terms of fine phonetic detail

in production (e.g. duration) by both frequency and full homophone density. While we have posited that speakers produce variant forms for high density cohorts for listener-orientated purposes, it is unknown whether such phonetic detail is actually meaningfully available in auditory perception. A further study, perhaps splicing low frequency (e.g. longer) productions into context sentences where a high frequency homophone is expected, and looking for small effects on perception, would be interesting. Auditory stimuli might provide more accurate information regarding the scalar nature of phonetic identity. For this end, and also for a more complete picture about differentiation of homophones in production, a more detailed examination of further phonetic parameters, such as degree of vowel dispersion and microvariation in tonal contours, would be illuminating.

This thesis looked at one particular unit, the character. Characters can be broken up into smaller parts, e.g. strokes and radicals; and also expanded into larger parts, like compound words. Compounding is common in modern Mandarin (as opposed to the classical language), and is the most common method for creating new words (Norman, 1988). This study looked at single characters, but there are many questions regarding the status of compounds. Compounds serve to reduce 'word-level' homophony to a large degree in Mandarin, and we might wonder what the status of compound words is in terms of lexical organisation, access, and production.

It is also known that complex characters (characters composed of two or more parts, which may have independent semantic or phonetic content) are very common in modern Mandarin (Perfetti and Tan, 1999; Shu and Anderson, 1999), and that the radical composition of characters induces priming effects, both phonological and semantic (Zhou and Marslen-Wilson, 1999). Priming effects due to character composition, and effects of character complexity, were not examined in this thesis, though it is advisable for future work that such potential effects are controlled for. It could also be argued that investigating responses to single characters in isolation presents an unrealistic view of how natural speech, which is

unlikely to consist of unrelated monosyllables, is processed. Despite this, it is our opinion that an understanding of single character lexical access is foundational to understanding lexical access in connected speech and in more natural conditions.

It was also beyond the practical limits of this thesis to consider the historical processes which led to the current high degree of homophony in Mandarin. How did the current structural distribution of homophony in Mandarin come about? Historically, there were both more tonal contrasts and many more unique base syllables. Is it historical accident that the collapse of some linguistic contrasts in Mandarin led to extremely high homophone density in some syllables, and very low homophone density in others? Or is there a linguistic reason for this distribution, i.e. some benefit to lexical processing or storage which exerts pressure toward increasing homophony for certain (presumably high frequency) syllables? Higher frequency syllables exhibit a greater tendency toward phonetic reduction (c.f. Bybee, 2001), and as Mandarin has relatively few syllables which are therefore used relatively more often, reduction and loss of contrasts become progressively more likely. Arguably the much reduced system is a much more efficient use of phonological/morphological structure, when compared with languages without such extensive homophony. Most words are not commonly used, and it is arguably 'wasteful' for a phonology/morphology to have unique forms for their representation, when context and frequency effects can perfectly well disambiguate shared forms as and when the need arises.

Summary

This thesis has provided detail on the structure and nature of homophony in Mandarin. We found that homophones have a surprising tendency to cluster in Mandarin, leading to large concentrations of homophones in some syllables, and no homophony at all in others. In a series of experiments, the uneven nature of homophony in Mandarin was found to be utilised by speakers, both in character perception and in production - possibly indicating how the

language was 'allowed' to become so deeply homophonous in the first place. Despite our expectations, derived from models of homophone processing designed with homophone-scarce languages such as English in mind, extensive and variegated homophony is *not* a barrier to successful communication in Mandarin, but gives rise to a lexicon which is organised in a fundamentally different way from the lexicons of non-homophonous languages, and exploits statistically guided processing in different ways.

Appendix 1. Experiment 2, Match Stimuli

Frequency	Segmental Density	Full Homophone Density	字	Pinyin	Frequency	Full Homophone Density
HH	H	H	是	shì	13583.8	25
			于	yú	3274.7	17
		M	之	zhī	3162.9	10
			以	yǐ	4729.4	7
			时	shí	4329.0	7
			里	lǐ	2793.1	5
	M	M	到	dào	5012.0	5
			后	hòu	2964.3	4
			家	jiā	2647.6	7
			生	shēng	3542.2	5
		L	用	yòng	2781.1	1
			也	yě	3688.8	3
			下	xià	3226.2	3
			出	chū	3922.5	2
	L	M	作	zuò	2819.0	4
			所	suǒ	2716.4	4
		L	对	duì	3654.4	3
			在	zài	10434.9	2
			上	shàng	5554.9	2
			然	rán	2654.1	2
H	H	H	服	fú	607.4	15
			书	shū	786.1	11
			其	qí	2093.2	13
		M	接	jiē	911.3	6
			具	jù	565.1	10
			无	wú	1865.9	5
			想	xiǎng	1915.5	4
		L	己	jǐ	1303.0	3
	M	H	程	chéng	711.3	11
		M	爱	ài	561.8	4
			况	kuàng	522.9	5
			变	biàn	961.4	6
			常	cháng	1102.6	7
		L	此	cǐ	1654.3	1
			定	dìng	2194.5	2
			多	duō	2501.7	3
			反	fǎn	942.0	2
	L	M	平	píng	993.4	6
			组	zǔ	621.9	4
			决	júe	820.4	7
		L	本	běn	2015.7	1
			白	bái	772.4	1
			德	dé	877.3	2
			存	cún	576.3	1

Frequency	Segmental Density	Full Homophone Density	字	Pinyin	Frequency	Full Homophone Density
M	H	H	精	jīng	504.5	11
			卫	wèi	303.3	12
			育	yù	345.6	17
		M	简	jiǎn	271.5	9
			较	jiào	430.4	6
			举	jǔ	358.0	4
			限	xiàn	342.9	9
		L	显	xiǎn	456.1	2
			局	jú	439.3	2
	M	H	木	mù	282.7	11
		M	帝	dì	343.0	7
			财	cái	291.0	4
			除	chú	463.6	5
			刚	gāng	335.6	8
		L	独	dú	333.4	3
			选	xuǎn	428.5	1
			突	tū	438.7	3
			随	suí	429.0	2
	L	M	宗	zōng	263.8	5
			列	liè	428.0	5
			农	nóng	459.2	4
			责	zé	396.7	5
		L	族	zú	385.2	2
			错	cuò	324.1	3
			跟	gēn	394.5	2
			考	kǎo	429.4	3
L	H	H	币	bì	117.3	18
			智	zhì	197.7	16
			励	lì	75.4	14
		M	笔	bǐ	176.9	4
			抚	fǔ	52.5	9
			径	jìng	95.7	9
			鼠	shǔ	54.4	8
		L	姐	jiě	216.9	2
			详	xiáng	78.4	3
	M	M	幸	xìng	193.2	4
			泰	taì	93.8	4
			班	bān	198.0	6
		L	茶	chá	100.8	3
			抽	chōu	119.1	1
			肥	féi	60.2	1
			鬼	guǐ	150.2	3
	L	M	宾	bīn	68.3	5
			垂	chuí	62.1	4
			测	cè	206.5	5
			括	kuò	210.4	4
		L	壮	zhuàng	78.7	3
			暂	zàn	93.3	2
			跳	tiào	162.7	2
			柔	róu	64.3	2

Frequency	Segmental Density	Full Homophone Density	字	Pinyin	Frequency	Full Homophone Density
LL	H	H	歼	jiān	46.3	12
			毅	yì	39.9	25
		M	拘	jū	32.3	6
			雁	yàn	18.0	8
			襄	xiāng	10.0	8
			巫	wū	29.0	6
		L	赎	shú	11.8	2
	M	H	卉	huì	6.5	11
		M	哀	aī	50.3	5
			黯	àn	11.5	5
			镑	bàng	19.0	4
			饼	bǐng	23.1	7
		L	札	zhá	7.6	3
			秧	yāng	5.2	2
			熏	xūn	9.9	2
			淌	tǎng	9.9	3
	L	M	舱	cāng	44.5	4
			澈	chè	6.2	4
			蛙	wā	11.2	4
			凳	dèng	15.2	4
		L	擦	cā	49.6	1
			凑	còu	28.8	1
			囊	náng	34.8	1
			瞥	piē	14.7	2

Appendix 2. Experiment 2, Mismatch Stimuli

Frequency	Segmental Density	Full Homophone Density	Mismatch Type	字	Pinyin	Presented Pinyin	Frequency	Full Homophone Density
HH	H	M	T	一	yi1	yí	15844.2	5
	M	M	C	道	dao4	rén	2777.0	5
			T	就	jiu4	jīu	4004.8	6
			S	他	ta1	shuō	8287.8	4
		L	T	自	zi4	zí	3176.9	2
			C	你	ni3	gè	3662.6	2
			S	可	ke3	chuān	3755.5	2
	L	L	C	年	nian2	wǒ	3126.0	2
			T	国	guo2	guò	5117.5	1
			S	能	neng2	ér	3455.6	1
H	H	H	C	十	shi4	bú	600.9	25
			T	记	ji4	jí	724.5	17
			S	西	xi1	jiāo	1265.3	20
		M	T	理	li3	lí	2067.5	5
			C	眼	yan3	shī	789.8	4
			S	主	zhu3	qǐ	2075.8	4
	M	M	C	安	an1	xíng	951.5	4
			T	保	bao3	bào	854.0	4
			S	备	bei4	dàn	556.3	8
		L	C	各	ge4	tōng	1004.6	2
			S	管	guan3	tī	886.8	2
			T	兵	bing1	bìng	555.6	2
	L	M	C	做	zuo4	gōng	911.9	4
			S	南	nan2	quán	722.3	4
			T	明	ming2	mìng	1609.4	6
		L	T	告	gao4	gāo	718.2	1
			S	二	er4	biāo	1389.3	1
			C	民	min2	rì	1744.0	1
M	H	H	S	致	zhi4	jì	406.3	16
			C	息	xi1	yán	511.3	20
			T	福	fu2	fū	290.2	15
		M	C	织	zhi1	wéi	360.9	10
			S	武	wu3	yǎn	425.2	7
			T	意	jing4	jīng	275.4	9
	M	M	C	够	gou4	tán	395.7	4
			S	河	he2	liú	361.9	6
			T	假	ban1	bàn	331.0	6
		L	C	苏	su1	zhòng	355.7	3
			S	米	mi3	huǒ	361.8	1
			T	环	huan2	huān	290.7	1
	L	M	C	装	zhuang1	nán	457.8	4
			S	坐	zuo4	duàn	344.3	4
			T	策	ce4	cē	272.1	5
		L	C	冷	leng3	pài	277.7	1
			S	拿	na2	réng	321.0	1
			T	退	tui4	tuī	266.3	1
L	H	H	C	鸡	ji1	zhèng	83.9	20
			T	坚	jian1	jiàn	253.2	12
			S	殊	shu1	xī	116.4	11
		M	C	乌	wu1	qì	105.4	6
			S	席	xi2	zhú	195.2	5
			T	县	xian4	xián	199.5	9
		L	T	鼻	bi2	bī	92.3	1
			S	祥	xiang2	péng	55.7	3
	M	H	S	慧	hui4	jìn	71.7	11
			T	穆	mu4	mū	55.0	11
		M	2	奥	ao4	aò	170.1	4
			C	刘	liu2	gān	252.4	8
			T	貌	mao4	máo	61.9	5
			S	荣	rong2	qín	164.4	10
		L	C	喊	han3	duó	118.3	2
			S	巧	qiao3	rěn	108.9	2
			T	冬	dong1	dōng	85.0	3
	L	M	C	迈	mai4	fá	54.2	4
			S	凝	ning2	lún	59.3	4
			T	胸	xiong1	xióng	90.3	5
		L	T	饶	rao4	rǎo	73.6	1
			S	恰	qia4	pèi	95.7	2
			C	偶	ou3	màn	89.6	2

Frequency	Segmental Density	Full Homophone Density	Mismatch Type	字	Pinyin	Presented Pinyin	Frequency	Full Homophone Density
LL	H	H	C	棋	qi2	hào	32.7	13
			S	莉	li4	dà	43.0	14
			T	寂	ji4	jī	51.3	17
		M	C	恕	shu4	yān	24.8	9
			S	轿	jiao4	wù	20.5	6
			T	捷	jie2	jiè	47.9	10
		L	C	迂	yu1	lī	10.4	2
	M	H	S	晦	hui4	zì	8.9	11
			T	绘	hui4	huí	46.9	11
		M	C	蛾	e2	dùn	5.9	7
			S	帆	fan1	gē	20.1	5
			T	蜡	la4	lā	16.9	4
		L	C	褥	ru4	shá	5.6	2
			T	娶	qu3	qù	19.4	3
			S	棉	mian2	kuí	38.5	3
	L	M	S	沦	lun2	máng	14.2	5
			C	恳	ken3	huái	26.4	4
			T	拳	quan2	quān	48.2	4
		L	C	惹	re3	biāo	26.1	1
			S	涩	se4	qià	12.3	3
			T	憎	zeng1	zèng	16.5	2

Appendix 3. Experiment 3, Sentence Stimuli

shi4	是	我 家 是 从 南 方 搬 来 的。
		wo3 jia1 ***shi4*** cong2 nan2 fang1 ban1 lai2 de.
		my family ***is*** from south move here *part.
		My family moved here from the South.
	侍	他 们 家 侍 从 对 人 很 和 善。
		ta1 men jia1 ***shi4*** cong2 dui4 ren2 hen3 he2 shan4.
		their family ***servant*** to people very kind.
		Their family's **servants** are very kind to people.
ren2	人	据 我 所 知，这 个 人 正 直 又 诚 实。
		ju4 wo3 suo3 zhi1 zhe4 ge4 ***ren2*** zheng4 zhi2 you4 cheng2 shi2.
		according to I *part. know, this *MW ***person*** having-integrity and honest.
		According to what I know, this **person** is honourable and honest.
	仁	虽 然 这 是 个 仁 政 国 家，那 国 王 太 软 弱 了。
		sui1 ran2 zhe4 shi4 ge4 ***ren2*** zheng4 guo2 jia1 na4 guo2 wang2 tai4 ruan3 ruo4 le.
		although this is *MW ***benevolent*** policy country that king too weak *part.
		Although this is a country with **benevolent** policies, the king is too weak.
li3	里	我 老 家 的 那 片 湖 里 鱼 有 很 多。
		wo3 lao3 jia1 de na4 pian4 hu2 ***li3*** yu2 you3 hen3 duo1.
		my home town *part. that *MW lake ***inside*** fish has very many.
		The lake in my hometown has a lot of fish **in** it.
	鲤	鲤 鱼 是 一 种 吉 祥 的 象 征。
		li3 yu2 shi4 yi1 zhong3 ji2 xiang2 de xiang4 zheng1.
		carp fish is one type good luck *part. symbol.
		The **carp** is a symbol of good luck.
ai4	爱	蝙 蝠 侠 爱 的 女 人 并 不 知 道 他 的 身 份。
		bian1 fu2 xia2 ***ai4*** de nv3 ren2 bing4 bu4 zhi1 dao4 ta1 de shen1 fen4.
		batman ***love*** *part. woman also not know his identity.
		The woman batman **loves** does not know his identity.
	隘	我 希 望 自 己 不 再 有 那 么 狭 隘 的 心 胸。
		wo3 xi1 wang4 zi4 ji3 bu4 zai4 you3 na4 me xia2 ***ai4*** de xin1 xiong1.
		I hope self not anymore have such ***narrow*** *part. heart.
		I hope I will no longer be so **small**-minded.
chu1	出	作 为 一 名 资 深 记 者 她 时 常 进 出 中 东。
		zuo4 wei2 yi1 ming2 zi1 shen1 ji4 zhe3 ta1 shi2 chang2 jin4 ***chu1*** zhong1 dong1.
		as one *MW senior reporter she often enter ***exit*** middle east.
		As a senior reporter, she is often in and **out** of the Middle East.
	初	今 天 是 他 进 初 中 以 来 最 高 兴 的 一 天。
		jin1 tian1 shi4 ta1 jin4 ***chu1*** zhong1 yi3 lai2 zui4 gao1 xing4 de yi1 tian1.
		today is he enter ***junior*** middle since most happy *part. one day.
		Today is his happiest day since he entered **junior** high school.

dui4	对	就 连 她 这 样 乐 观 的 人 也 对 现 状 失 去 了 信 心。
		jiu4 lian2 ta1 zhe4 yang4 le4 guan1 de ren2 ye3 ***dui4*** xian4 zhuang4 shi1 qu4 le xin4 xin1.
		even she such optimistic *part. person also ***regarding*** current situation lose *part. confidence.
		Even an optimistic person like her has lost confidence **regarding** the current situation.
	兑	那 些 承 诺 都 是 空 谈，永 远 也 兑 现 不 了。
		na4 xie1 cheng2 nuo4 dou1 shi4 kong1 tan2 yong3 yuan3 ye3 ***dui4*** xian4 bu4 liao3.
		those promises all are empty talk forever also ***trade***-cash (fulfil) not *part.
		Those promises are all empty talk and will never be **fulfilled**.
ji1	机	我 觉 得 这 台 拖 拉 机 的 定 价 太 高 了。
		wo3 jue2 de2 zhe4 tai2 tuo1 la1 ***ji1*** de ding4 jia4 tai4 gao1 le.
		I think this *MW drag ***machine*** *part. list price too high *part.
		I think the list price of this **tractor** is too high.
	圾	这 已 改 变 人 们 对 垃 圾 的 处 理 方 式。
		zhe4 yi3 gai3 bian4 ren2 men2 dui4 la1 ***ji1*** de chu3 li3 fang1 shi4.
		this already change people toward ***rubbish*** *part. handle method.
		This has changed the way people handle **rubbish**.
zhi1	之	这 是 泰 坦 尼 克 号 沉 没 之 前 拍 摄 的 最 后 一 张 照 片。
		zhe4 shi4 tai4 tan3 ni2 ke4 hao4 chen2 mo4 ***zhi1*** qian2 pai1 she4 de zui4 hou4 yi1 zhang1 zhao4 pian4.
		this is Titanic sink ****lit. poss.*** before photograph *part. last one *MW photograph.
		This is the last photograph taken **before** the sinking of the Titanic.
	汁	在 蘸 墨 汁 前 他 先 把 毛 笔 用 清 水 洗 了 一下。
		zai4 zhan4 mo4 ***zhi1*** qian2 ta1 xian1 ba3 mao2 bi3 yong4 qing1 shui3 xi3 le yi1 xia4.
		when dip ink-***liquid*** before he first *part. hair pen use clear water wash *part. a little.
		Before he dipped the brush in **ink**, he first rinsed it with clear water.
xin1	心	设 计 师 在 裙 子 上 用 心 镶 嵌 了 一 块 宝 石。
		she4 ji4 shi1 zai4 qun2 zi3 shang4 yong4 ***xin1*** xiang1 qian4 le yi1 kuai4 bao3 shi2.
		designer *part. dress on use ***heart*** inlay *part. one piece precious stone.
		The designer very **carefully** inlaid a gem in the dress.
	馨	他 用 馨 香 的 桂 花 酿 出 一 壶 好 酒。
		ta1 yong4 ***xin1*** xiang1 de gui4 hua1 niang4 chu1 yi1 hu2 hao3 jiu3.
		he use ***fragrant*** *part. osmanthus flower brew out one jar good alcohol.
		He used **fragrant** osmanthus to make a good jar of wine.
wu2	无	分 手 后 他 们 再 无 同 台 表 演 的 机 会 了。
		fen1 shou3 hou4 ta1 men2 zai4 ***wu2*** tong2 tai2 biao3 yan3 de ji1 hui4 le.
		breakup after they again ***no*** same stage perform *part. opportunity *part.
		After breaking up, they do **not have** the opportunity to perform on the same stage anymore.
	梧	我 很 喜 欢 坐 在 梧 桐 树 下 发 呆。
		wo3 hen3 xi3 huan1 zuo4 zai4 ***wu2*** tong2 shu4 xia4 fa1 dai1.
		I very like sit at ***plane*** tree under day dream.
		I like sitting under the **plane** tree and daydreaming.

References

Bard, E.G., Anderson, A.H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., and A. Newlands (2000). 'Controlling the intelligibility of referring expressions in dialogue.' *Journal of Memory and Language* 42:1-22.

Bell, A., Brenier, J., Gregory, M., Girand, C., and D. Jurafsky (2009). 'Predictability effects on durations of content and function words in conversational English.' *Journal of Memory and Language* 60:92-111.

Bybee, J. (2001). *Phonology and language use*. Cambridge University Press.

Caramazza, A., Miozzo, M., Costa, A., and Y. Bi (2001). 'The whole-word frequency effect: Implications for the representations of homophones.' *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27:1430-1450.

Chao, Y. R. (1930). 'A system of tone letters'. *Le maître phonétique* 45:24-27.

Chao, Y.R. (1943). 'Languages and Dialects in China.' *The Geographical Journal* 102 (2):63-66.

Chao, Y.R. (1961). 'What is Correct Chinese?' *Journal of the American Oriental Society* 81(3):171-177.

Chao, Y.R. (1968a). *A Grammar of Spoken Chinese*. University of California Press.

Chao, Y.R. (1968b). *Language and Symbolic Systems*. Cambridge University Press.

Chen, R. (1995). 'Communicative Dynamism and Word Order in Mandarin Chinese.' *Language Sciences* 17(2):201-222.

Chen, T.H. and D. Massaro (2008). 'Seeing Pitch: Visual information for lexical tones of Mandarin Chinese.' *Journal of the Acoustic Society of America*, 123(4):2356-2366.

Chen, M. and J. Yuen (1991). 'Effects of Pinyin and Script Type on Verbal Processing: Comparisons of China, Taiwan, and Hong Kong Experience.' *International Journal of Behavioral Development* 14(4):429-448.

Clark, H.H., and S.E. Brennan (1991). 'Grounding in communication.' In *Perspectives on socially shared Cognition* (L.B. Resnick, B.M. Levine, M. John, and S.D. Teasley, *editors*), pp.127-149. Washington, DC, US: American Psychological Association.

Coblin, W.S. (2000). 'A Brief History of Mandarin.' *Journal of the American Oriental Society* 120(4):287-288.

Crystal, D. (2008). *A dictionary of linguistics and phonetics* (6th edition). Wiley-Blackwell.

Cutler, A., and H.C. Chen (1997). 'Lexical tone in Cantonese spoken word processing.' Attention, Perception and Psychophysics 59:165-179.

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

Dell, G.S. (1986). 'A spreading-activation theory of retrieval in sentence production.' *Psychological Review* 93(3):283-321.

Dell, G.S., and J.K. Gordon (2003). 'Neighbors in the lexicon: Friends or foes?' In *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (N.O. Schiller and A.S. Meyer, editors), pp. 9-37. Mouton de Gruyter.

Dell, G.S., Schwartz, M.F., Martin, N., Saffran, E.M., and D.A. Gagnon (1997). 'Lexical access in aphasic and nonaphasic speakers.' *Psycho*logical Review 104:801-838.

Duanmu, S. (1998). 'Wordhood in Chinese.' In *New Approaches to Chinese Word Formation: Morphology, Phonology, and the Lexicon in Modern and Ancient Chinese* (Packard, J, editor), pp.135-196. Mouton de Gruyter, Berlin.

Duanmu, S. (2005). 'Chinese (Mandarin), Phonology of.' Encyclopedia of Language and Linguistics (Second Edition). Elsevier Publishing House.

Duanmu, S. (2007). *The Phonology of Standard Chinese* (Second Edition). Oxford University Press.

Feldman, L. and W.W. Siok (1999). 'Semantic Radicals in Phonetic Compounds: Implications for Visual Character Recognition in Chinese.' In *Reading Chinese Script* (Wang, J., Inhoff, A., and H.C. Chen, *editors*), pp. 19-35. Lawrence Erlbaum Associates, London.

Foss, D.J. (1970). 'Some Effects of Ambiguity upon Sentence Comprehension.' *Journal of Verbal Learning and Verbal Behavior* 9:699-706.

Gahl, S. (2008). '"Time" and "thyme" are not homophones: Word durations in spontaneous speech.' *Language* 84:474-496.

Gahl, S., Yao, Y., and K. Johnson (2012). 'Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech.' *Journal of Memory and Language* 66:789-806.

Galati, A. and S.E. Brennan (2010). 'Attenuating information in spoken communication: For the speaker, or for the addressee?' *Journal of Memory and Language* 62:35-51.

Goldsmith, J. (1990). *Autosegmental and metrical phonology*. Oxford: Basil Blackwell.

Gordon, B. (1983). 'Lexical Access and Lexical Decision: Mechanisms of Frequency Sensitivity.' *Journal of Verbal Learning and Verbal Behavior* 22:24-44.

Guion, S. (1995). 'Word Frequency effects among homonyms.' *Texas Linguistic Forum* 35:103-116.

Hamed, M. and F. Wang (2006). 'Stuck in the forest: Trees, networks and Chinese dialects.' *Diachronica* 23(1): 29:29-60.

Hay, J. (2002). 'From Speech Perception to Morphology: Affix Ordering Revisited.' *Language* 78(3):527-555.

Hogaboam, T.W. and C.A. Perfetti (1975). 'Lexical Ambiguity and Sentence Comprehension.' *Journal of Verbal Learning and Verbal Behavior* 14:265-274.

Hong, E.L. and G. Yelland (1996). 'The Generality of Lexical Neighbourhood Effects.' In *Cognitive Processing of Chinese and Related Asian Languages* (H.C. Chen *editor*), pp. 187-203. Hong Kong: Chinese University Press.

Hoosain, R. (1991). *Psycholinguistic Implications for Linguistic Relativity: A Case Study of Chinese*. Lawrence Erlbaum Associates, Inc.

Hoosain, R. (1992). 'Psychological Reality of the Word in Chinese.' In *Language Processing in Chinese* (H.C. Chen and O.J.L. Tzeng, *editors*), pp. 111-130. Elsevier.

Hu, C. (2007). 'Hu Shi and the First Advocate of the New Punctuation, Wang Yuanfang.' *Chinese Studies in History* 40(4):86-92.

Institute of Language Teaching and Research (1986). *Modern Chinese Frequency Dictionary*. Beijing Language Institute Press.

Jescheniak, J., and W. Levelt (1994). 'Word frequency effects in speech production: Retrieval of syntactic information and of phonological form.' *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20:824-843.

Jescheniak, J., Meyer, A.S., and W. Levelt (2003). 'Specific-word frequency is not all that counts in speech production: Comments on Caramazza, Costa, et al. (2001) and new experimental data.' *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29:432-438.

Jurafsky, D., Bell, A., Gregory, M.L., and W. Raymond (2001a). 'The effect of language model probability on pronunciation reduction.' Paper presented at the *International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2001*, Salt Lake City, UT.

Jurafsky, D., Bell, A., Gregory, M.L., and W. Raymond (2001b). 'Probabilistic relations between words: Evidence from reduction in lexical production.' In *Frequency and the emergence of linguistic structure*, (J. Bybee and P. Hopper, *editors*), pp. 229-254. Amsterdam: John Benjamins.

Kilanski, K. (2009). *The effects of token frequency and phonological neighborhood density on native and non-native speech production*. Seattle: University of Washington.

Krupa, V. (1971). 'The Phonotactic Structure of the Morph in Polynesian Languages.' *Language 47(3):668-684*.

Kuo, Y.C., Rosen, S., and A. Faulkner. (2008). 'Acoustic cues to tonal contrasts in Mandarin: Implications for cochlear implants.' *Journal of the Acoustic Society of America*, 123(5): 2815-2824.

Lai, Y. and Zhang, J. (2008). 'Mandarin Lexical Tone Recognition: The Gating Paradigm.' *Kansas Working Papers in Linguistics* 30:183-194.

LaPolla, R. (2010). 'Language Contact and Language Change in the History of the Sinitic Languages.' *Procedia - Social and Behavioral Sciences* 2(5):6858-6868.

Lee, C. (2007). 'Does Horse Activate Mother? Processing Lexical Tone in Form Priming.' *Language and Speech* 50(1):101-123.

Levelt, W., Roelofs, A., and A. Meyer (1999). 'A theory of lexical access in speech production.' *Behavioral and Brain Sciences* 22(1):1-38.

Li, C.W.C. (2004). 'Conflicting notions of language purity: the interplay of archaising, ethnographic, reformist, elitist, and xenophobic purism in the perception of Standard Chinese.' *Language and Communication* 24:97-133.

Li, X (2006). '李小凡"汉语方言的轻声变调"中国方言学报.' 'Neutral tone sandhi in Chinese dialects.' *Journal of Chinese Dialects* 1:41-52.

Li, Y.-H. A. (1999). 'Plurality in a Classifier Language.' *Journal of East Asian Linguistics* 8(1):75-99.

Li, L. and C.H. Tat (2014). 'Acquisition of Chinese quardra-syllabic idiomatic expressions: Effects of semantic opacity and structural symmetry.' *First Language* 34(4):336-353.

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,* pp. 228-232.

Li, P. and M. Yip (1998). 'Context effects and the processing of spoken homophones.' *Reading and Writing: An Interdisciplinary Journal* 10:223–243.

Liu, L., Yang, W., Wang, H., and Y. Chang (1989). 'Tone recognition of polysyllabic words in Mandarin speech.' *Computer Speech and Language* 3:253-264.

Liu, S., and A.G. Samuel (2004). 'The role of Mandarin lexical tones in lexical access under different contextual conditions.' *Language and Cognitive Processes* 22(4):566-594.

Luce, P. and C. McLennan (2005). 'Spoken Word Recognition: The Challenge of Variation.' In *The Handbook of Speech Perception* (D. Pisoni and R. Remez, *editors*), pp. 591-609. Blackwell.

Luce, P. A., and D.B. Pisoni (1987). 'Speech perception: Recent trends in research, theory, and applications.' In *Human communication and its disorders* (H. Winitz, *editor*), pp.1-87. Norwood, NJ: Ablex.

Luce, P., and D. Pisoni (1998). 'Recognizing Spoken Words: The Neighborhood Activation Model.' *Ear and Hearing* 19:1-36.

Lukatela, G., Eaton, T., Lee, C., Carello, C., and M.T. Turvey (2002). 'Equal Homophonic Priming with Words and Pseudohomophones.' *Journal of Experimental Psychology* 28:3-21.

Marchman, V., and E. Bates (1994). 'Continuity in lexical and morphological development: a test of the critical mass hypothesis.' *Journal of Child Language* 21(2):339-366.

Marian, V., Bartolotti, J., Chabal S., and A. Shook (2012). 'CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities.' *PLOS ONE* 7(8): e43230.

Marslen-Wilson, W. D. (1987). 'Parallel processing in spoken word recognition.' *Cognition,* 25:71-102.

McClelland, J. L., and J.L. Elman (1986). 'The TRACE model of speech perception.' *Cognitive Psychology* 18:1-86.

Moore, C. B. (1993). 'Phonetic Observations on Stress and Tones in Mandarin Chinese.' *Working Papers of the Cornell Phonetics Laboratory* 8:87-117.

Morton, J. (1969). 'Interaction of information in word recognition.' *Psychological Review* 76(2):165-178.

Munson, B. (2007). 'Lexical access, lexical representation, and vowel production.' In *Laboratory phonology 9: Phonology and Phonetics* (J. Cole and J.I. Hualde, *editors*), pp. 201-228. Berlin: Mouton.

Munson, B., and N.P. Solomon (2004). 'The effect of phonological neighborhood density on vowel articulation.' *Speech, Language, and Hearing Research* 47:1048-1058.

Norman, J. (1988). *Chinese*. Cambridge University Press.

Norris, D. (1994). 'Shortlist: A connectionist model of continuous speech recognition.' *Cognition* 52:189-234.

Peng, D., Liu, Y., and C. Wang (1999). 'How is Access Representation Organized? The Relation of Polymorphemic Words and their Morphemes in Chinese.' In *Reading Chinese Script: A cognitive analysis* (A. Inhoff, J. Wang, and I. Chen, *editors*), pp. 65-89. Lawrence Erlbaum Associates, London.

Perfetti, C. A., & Tan, L. H. (1999). 'The constituency model of Chinese character identification.' In *Reading Chinese Script: A cognitive analysis* (A. Inhoff, J. Wang, and I. Chen, *editors*), pp. 115-134. Lawrence Erlbaum Associates, London.

Pierrehumbert, J. (2001). 'Exemplar dynamics: Word frequency, lenition and contrast.' In *Frequency and the emergence of linguistic structure* (J. Bybee and P. Hopper, *editors*), pp. 137-157. Amsterdam: John Benjamins.

Pisoni, D. B., Nusbaum, H. C., Luce, P. A., and L.M. Slowiaczek (1985). 'Speech perception, word recognition and the structure of the lexicon.' *Speech Communication* 4:75-95.

Psychology Software Tools, Inc. [E-Prime 2.0] (2012). Retrieved from www.pstnet.com.

Rumelhart, D., and P. Siple (1974). 'Process of Recognizing Tachistoscopically Presented Words.' Psychological Review 81:99-118.

Scarborough, R.A. (2009). 'Lexical similarity and speech production: Neighborhoods for nonwords.' In paper presented at the NELS workshop on phonological similarity.

Sereno, J.A. and A. Jongman (1997). 'Processing of English inflectional morphology.' *Memory & Cognition* 25:425-437.

Sherr-Ziarko, E (2015). 'Word Frequency Effects on Homophonous Words in Mandarin Chinese.' In *The Scottish Consortium for ICPhS 2015 (Ed.), Proceedings of the 18th International Congress of Phonetic Sciences*. The University of Glasgow.

Shih, C (1988). 'Tone and Intonation in Mandarin.' In *Working Papers of the Cornell Phonetics Laboratory* 3:83-109.

Shu, H. and R. C. Anderson (1999). 'Learning to Read Chinese: The Development of Metalinguistic Awareness.' In *Reading Chinese Script: A cognitive analysis* (A. Inhoff, J. Wang, and I. Chen, *editors*), pp. 1-18. Lawrence Erlbaum Associates, London.

Simpson, G. (1981). 'Meaning Dominance and Semantic Context in the Processing of Lexical Ambiguity.' *Journal of Verbal Learning and Verbal Behavior* 20:120-136.

Sproat, R., Shih, C., Gale, W., and N. Chang (1996). 'A Stochastic finite-state word-segmentation algorithm for Chinese.' *Computational Linguistics* 22(3):377-404.

Swinney, D.A. (1979). 'Lexical Access during Sentence Comprehension: (Re)Consideration of Context Effects.' *Journal of Verbal Learning and Verbal Behavior* 18:645-659.

Taft, M., and H.C. Chen (1992). 'Judging Homophony in Chinese: The Influence of Tones.' In *Language Processing in Chinese* (H.C. Chen and O.J.L. Tzeng, *editors*), pp. 151-172. Elsevier.

Taft, M., Liu, Y., and X. Zhu (1999). 'Morphemic Processing in Reading Chinese.' In *Reading Chinese Script: A cognitive analysis* (A. Inhoff, J. Wang, and I. Chen, *editors*), pp. 91-113. Lawrence Erlbaum Associates, London.

Tan, L.H., Xu, M., Chang, C.Q., and W.T. Siok (2013). 'China's language input system in the digital age affects children's reading development.' *Proceedings of the National Academy of Sciences of the United States of America* 110(3):1119-1123.

Tang, C. and V. van Heuven (2009). 'Mutual intelligibility of Chinese dialects experimentally tested.' *Lingua* 119:709-732.

Teng, E., Lee, P.H., Yang, K.S., and P. Chang (1979). 'Lateral Preferences for Hand, Foot and Eye, and Their Lack of Association with Scholastic Achievement, in 4143 Chinese.' *Neuropsychologia* 17:41-48.

Tyler, L.K. and U.H. Frauenfelder. (1987). 'The process of spoken word recognition: An introduction.' In *Spoken Word Recognition* (Frauenfelder, U.H. and L.K. Tyler, *editors*). MIT Press.

Vitevitch, M.S., and P.A. Luce (1998). 'When words compete: Levels of processing in perception of spoken words.' *Psychological Science* 9:325-329.

Vu, H., Kellas, G., and S. Paul (1998). 'Sources of sentence constraint on lexical ambiguity resolution.' *Memory and Cognition* 26(5):979-1001.

Wan, I.-P., and J. Jaeger (1998). 'Speech errors and the representation of tone in Mandarin Chinese.' *Phonology* 15:417-461.

Wan, I.-P., and J. Jaeger (2003). 'The Phonological Representation of Taiwan Mandarin Vowels: A Psycholinguistic Study.' *Journal of East Asian Linguistics* 12:205-257.

Whalen, D. H., and Y. Xu (1992). 'Information for Mandarin tones in the amplitude contour and in brief segments.' *Phonetica* 49:25-47.

Whaley, C.P. (1978). 'Word-Nonword Classification Time.' *Journal of Verbal Learning and Verbal Behavior* 17:143-154.

Wong, P. (2013). 'Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan.' *Journal of the Acoustical Society of America* 133(1):434-443.

Wong, W. and S.P. Law (2008). 'The Relationship Between Semantic Short-Term Memory and Immediate Serial Recall of Known and Unknown Words and Nonwords: Data From Two Chinese Individuals With Aphasia.' *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34(4):900-917.

Wiener, S., Speer, S., and C. Shank (2012). 'Effects of frequency, repetition and prosodic location on ambiguous Mandarin word production.' In *Proceedings of the 6th International Conference on Speech Prosody*, pp. 528-531. Shanghai, China.

Winskel, H., Ratitamkul, T., and A. Charoensit (2016). 'The role of tone and segmental information in visual word recognition in Thai.' *The Quarterly Journal of Experimental Psychology* 70(7):1282-1291.

Wright, R. (1997). 'Lexical competition and reduction in speech: A preliminary report.' *Indiana University research on spoken language processing progress report* 21:471-485.

Wright, R. (2004). 'Factors of lexical competition in vowel articulation.' In *Papers in laboratory phonology VI* (J. Local, R. Ogden, and R. Temple, *editors*), pp. 75-87. Cambridge University Press.

Ye, Y. and C. Connine (1999). 'Processing Spoken Chinese: The Role of Tone Information.' *Language and Cognitive Processes* 14(5/6):609-630.

Zhang, H. (2007). 'Numeral Classifiers in Mandarin Chinese.' *Journal of East Asian Linguistics* 16:43-59.

Zhang, B. and D. Peng (1992). 'Decomposed Storage in the Chinese Lexicon.' In *Language Processing in Chinese* (H.C. Chen and O.J.L. Tzeng, *editors*), pp. 131-149. Elsevier.

Zhou, X., and W. Marslen-Wilson(1995). 'Morphological structure in the Chinese mental lexicon.' *Language and Cognitive Processes* 10:545-600.