# The Effects of Sound Information Provided by Annotations on Mandarin Chinese Vocabulary Acquisition Through Reading

JING CHU

Thesis submitted for the degree of PhD

2017

Department of Linguistics

SOAS, University of London

# Declaration for SOAS PhD thesis

I have read and understood regulation 17.9 of the Regulations for students of the SOAS, University of London concerning plagiarism. I undertake that all the material presented for examination is my own work and has not been written for me, in whole or in part, by any other person. I also undertake that any quotation or paraphrase from the published or unpublished work of another person has been duly acknowledged in the work which I present for examination.

Signed: _____ Date: _____

# Abstract

This study is an investigation of the effects of the rarely researched sound-related (Pinyin/audio) annotations on incidental vocabulary learning under the following important, but less studied, conditions: 1) L2 Mandarin Chinese reading; 2) testing productive word knowledge along with receptive knowledge; and 3) when criteria that are sensitive to partial knowledge are applied to measure word knowledge gain. The research consists of two related experiments with samples of 25 and 41 beginner-level students respectively. The first experiment has a between-subject design and compares the effects of the text-only and text + Pinyin annotation on incidental vocabulary learning in a pen-and-paper environment, while the second experiment compares the effects of text + Pinyin, text +audio, text + Pinyin + audio annotation with a self-designed online reading program in a within-subject design.

The only statistically significant difference was found between the text-only and text + Pinyin annotations in the meaning-based Pinyin production test in the first experiment ($U = 28.00$, $z = 2.897$, $p < 0.02$, $r = - 0.58$). However, there is a tendency for the text +Pinyin annotation to be more helpful in terms of assisting the acquisition of not only sound-related knowledge of words, but also knowledge relating to character form. Similarly, positive effects of text + Pinyin + audio annotation were also observed in the second experiment.

Both experiments demonstrated various types of sound-related word knowledge gain and the amount ranged from 10.67% to 33.8%. More importantly, the results suggested that criteria that are sensitive to partial knowledge were crucial to incidental vocabulary acquisition research because the differences between the scores marked under such criteria and the scores marked without such criteria were significant ($F (6.12, 58.5) = 6.12$, $p= 0.03$). In addition, the number of total strokes in a word might affect the results of incidental vocabulary acquisition.

# Contents

# List of tables and figures

# Acknowledgements

Conducting this PhD research and overcoming the related difficulties and tough times have been very rewarding and special for me, both academically and personally – and I would like to express my sincere gratitude to everyone who has generously supported me in one way or another.

My thanks go first to my main supervisor, Dr Noriko Iwasaki, for her continuous support of my PhD study and her patience, motivation, and understanding throughout this process. I would also like to thank the other members of my supervisory committee, Dr Barbara Pizziconi and Dr Lianyi Song. I am especially grateful to Dr Song for his help with finding participants for my study's first experiment.

I would like to thank my examination committee (namely Dr Clare Wright, Dr Qian Kan, and Prof Jae Hoon Yeon) for not only their insightful comments and encouragement, but also the hard questions they asked to motivate me to widen my research to consider additional perspectives.

My sincere thanks also go to Dr Hong Lu, who provided me with the opportunity to join the teaching team at the *Confucius Institute for Business London* at the *London School of Economics and Political Science* – where I gained much experience of Mandarin teaching in the context of the UK higher education system. I thank my fellow teachers for the stimulating discussions, the sleepless nights we spent working together, and all of the fun we had there. Without such teaching experience and the financial support that accompanied it, I would not have been able to conduct this research.

Last but not least, I would like to thank my friend Theresa Booth for her kind help with polishing the language in this thesis, as well as my parents for supporting me spiritually throughout the thesis writing process and my life in general.

# Chapter 1: Introduction

When second language (L2) learners are asked what the essential aspects of learning another language are, the most common answers are generally pronunciation, grammar, vocabulary, and perhaps the written character form found in such languages as Mandarin and Japanese. No matter which language is being learned, vocabulary plays a crucial role in mastering it. However, this importance was not reflected in either research on second language acquisition (SLA) or in language teaching practices; moreover, research on L2 vocabulary acquisition was neglected before the 1980s (Meara, 1980). Learners have mainly acquired words in a new language through their own efforts. Traditionally, researchers seemed to think that teaching vocabulary was not necessary, in the belief that "the new language would somehow magically fall into place" (Folse, 2004, PV) as long as learners were provided with comprehensible input. Fortunately, this situation is gradually changing, and studies are now being conducted against the backdrop of vocabulary acquisition moving to the centre of SLA research – especially in relation to incidental learning of vocabulary, which has become an object of considerable interest in the field of SLA since the 1990s. This research, which focuses on learning L2 Mandarin Chinese vocabulary incidentally through reading, was catalysed by these developments. This chapter provides an overview of the study and orients the reader to the organisation of the thesis.

This research is primarily concerned with whether sound or sound-related information concerning unknown words help L2 Mandarin learners with incidental vocabulary acquisition through reading, using criteria that are sensitive to partial knowledge. Moreover, it also addresses whether annotations that provide different types of sound-related information, such as Pinyin form and spoken form of word (provided by audio annotation), affect results of incidental vocabulary acquisition to a different extent in an online reading environment. In this study, sound-related information is provided by two types of annotation, namely Pinyin and audio. Both

annotation types are included in the study's second experiment (which is conducted in an online environment), while only the Pinyin annotation is employed in the first experiment (which is undertaken in a pen-and-paper context). In addition, learners' attitudes to the different types of annotation are also investigated to decide whether learners have different preferences for them.

This study was initially motivated by the growing attention being paid to L2 vocabulary research, especially incidental vocabulary acquisition through reading. The attention on incidental vocabulary acquisition originated from a few studies conducted by Nagy and his associates (1985, 1987) in the field of first language learning, which then attracted much research interest in relation to SLA. Nagy, Herman and Anderson (1985) investigated the amount of L1 (English) incidental word knowledge gained through natural reading; multiple-choice tests across different levels of difficulty revealed that the probability of learning a word is 15% to 22%, whereas interviews indicated a range from 11% to 19%. Based on this study, they proposed the incidental vocabulary learning hypothesis. This hypothesis suggests that incidental vocabulary learning happens while reading, although only small increments of a word can be acquired in this procedure. Moreover, such partial increases in knowledge of a word can be expected with even a single encounter with that word. However, it should be borne in mind that these researchers' interest in word knowledge was restricted exclusively to the meaning of words.

Following Nagy et al. (1985), research on incidental vocabulary learning (which is also referred to as incidental vocabulary acquisition) was extended and attracted attention from many SLA researchers (e.g. Al-Seghayer, 2001; Chun & Plass, 1996; Duan & Yan, 2004; Ellis & He, 1999; Gao & Liu, 2009; Folse, 2004; Paribakht & Wesche, 1999; Huckin & Coady, 1999; Hulstijn, Hollander, & Greidanus, 1996; James, 2009; Laufer & Hill, 2000; Wu & Xu, 2009; Yeh & Wang, 2003;). Nonetheless, no definition of incidental vocabulary learning/acquisition has been universally accepted. The challenge of providing such a definition stems from a set of ambiguous notions borrowed from

psychology, inter alia attention, consciousness, and implicit and explicit learning (Schmitt, 2000; Haynes, 1998). Researchers have tried to define it from a mental processing perspective (Nation, 2001), through its by-product nature (Paribakht & Wesche, 1999; Nation, 2001), and using its methodological meaning (Laufer & Hulstijn, 2001); further discussion can be found in Section 2.2.1.2. The current study follows Laufer and Hulstijn's dichotomy of separating the confusing concepts of “methodology meaning” and “educational meaning” and in operational terms regards incidental vocabulary learning as learning in which pre-learning instructions do not forewarn learners that a subsequent vocabulary test exists.

Against the backdrop of increasing interests in research on incidental vocabulary learning, this research focuses on L2 Mandarin Chinese, a language that has rarely been investigated in this context. Other reasons for choosing this language are as follows: firstly, the L2 Mandarin Chinese context fits well with my past teaching experience; secondly, I am personally interested in conducting research on topics related to improving L2 Mandarin Chinese teaching and learning.

Although some excellent reviews on incidental vocabulary learning exist within SLA (see, for example, Huckin & Coady, 1999; Khatib & Zourzadeh, 2012; Waring & Nation, 2004), each to some extent reflects the authors' personal research interests and expertise – which do not include Mandarin Chinese. This situation confirms Folse’s (2004) statements about European languages (e.g., German, English, and Spanish) dominating research in this field. The reasons for not including studies on Mandarin Chinese in these reviews could be either that such studies do not exist or that it was difficult to assess studies published in languages such as Mandarin Chinese due to language barriers. To bridge this gap in the literature and build a solid foundation for this research, studies relating to incidental vocabulary acquisition in the Mandarin Chinese literature were therefore reviewed; details are presented in Section 2.2.2. The results of this review suggest that L2 Mandarin Chinese is clearly under-researched as a target language, as only seven of all the 68 studies in incidental vocabulary acquisition

in Chinese literature I considered were found to target it. Such a finding clearly demonstrates the need to conduct research on incidental vocabulary learning in the context of L2 Mandarin Chinese.

In addition to the target language considered, this research also differs from previous studies by focusing on sound-related word knowledge (and not word meaning) when investigating the incidental acquisition of words in the context of reading. The reasons for this choice stem mainly from the following three factors. Firstly, while reading builds receptive knowledge of a word's written form (i.e. helps learners to recognise a written word) – which is the type of knowledge usually tested in this field – researchers have proven that reading also enables learners to gain other types of word knowledge incidentally; this includes both association (Horst, Cobb, & Meara, 1998) and collocation (Webb, Newton, & Chang, 2013). Moreover, several studies have inadvertently demonstrated the possibility for learners to acquire productive knowledge of a word's written form (Bowles, 2004; Yanguas, 2009) and receptive knowledge of Kana in Japanese (James, 2009) when certain types of vocabulary tests are used (see Section 3.2 for details of these studies). As such, it is reasonable to believe that other types of word knowledge can be acquired incidentally through reading. Considering the deep orthography of the character form of Mandarin Chinese, which requires learners to put extra effort into mastering a word's character form, it would be interesting to know if learners can acquire word knowledge related to sound of word knowledge through reading as well. If yes, reading could be regarded as a more efficient resource than normally expected for L2 Mandarin Chinese learners.

Secondly, to improve incidental vocabulary learning in a reading context, previous researchers most commonly added glosses or multiple types of annotations to a reading text, as these methods have been proven to be effective (Abraham, 2008; Hulstijn et al., 1996; Watanabe, 1997). This initially meant simply adding word meaning as a marginal gloss. However, researchers' interests then shifted to comparing the effect of different types of glosses on incidental vocabulary learning, for example,

comparing the effect of word meaning provided in L1 with that in L2, and comparing the effect of text annotation and that of pictorial annotation (Hulstijn et al., 1996; Yoshii, 2006). With the development of modern technologies, research on incidental vocabulary acquisition also started to be conducted in a computer-assisted language learning (CALL) environment. Moreover, audio and video annotations were included in several studies, such as Chun and Plass (1996) and Yeh and Wang (2003). Regardless of which type of annotation is the best in terms of assisting learners to acquire word knowledge incidentally through reading, however, to the best of my knowledge the effect of audio annotation has frequently been ignored in previous studies.

Furthermore, studies investigating the relationship between knowing a word's pronunciation or Pinyin and understanding its meaning have suggested that these aspects are strongly correlated. They have also revealed that L2 Mandarin Chinese learners – particularly those who lack learning experience with non-alphabetic languages – rely on sound-related information in learning Chinese (Everson, 1998; Qian, 2003). As such, sound-related information provided in annotations may have a fundamental impact on a learner's ability to acquire word knowledge in Mandarin Chinese, in relation to not only knowledge relating to sound of a word but also to its written form (i.e. character form). Further details of these issues are largely clustered in Chapter 3 (which reviews word knowledge and assessments) and Chapter 2 (which relates to the development of research on incidental vocabulary). Considering the three points discussed above, this study hence sets sound-related annotation as its research target and attempts to determine the effect of such annotation on incidental vocabulary acquisition.

Apart from that, the incidental vocabulary learning of beginner learners of Mandarin Chinese in an online environment was investigated in this research. Learners at beginner's level were chosen mainly because they tend to need more help to overcome the challenges caused by the complexity of Mandarin Chinese's spoken and written systems. Mandarin Chinese is difficult for learners who have learnt only

Western languages, such as English, German, Spanish, etc. To learn a word in Mandarin Chinese, in addition to learning how to pronounce it and how to read and write its unique written form, that is, Chinese characters, beginner learners would have to learn a Pinyin system, in which the Latin alphabet and tone marks are used to record the pronunciation of Chinese characters (see Section 3.3 for more details of the unique nature of L2 Mandarin Chinese learning). Thus, the workload might be very heavy for these learners. Moreover, without much learning experience, they should face a great challenge at this stage. Therefore, beginner learners were selected for this research with the objective of finding methods that would help them to reduce the difficulties caused by the complicated spoken and written systems.

The online program was developed by the researcher to accommodate the audio annotation of words, which is difficult to present in a paper-and-pen environment. In addition, with the ease of access and affordances of modern technology, online programs are being increasingly used by all education providers and learners. It is therefore time for Mandarin Chinese language teachers and researchers to explore Mandarin Chinese online learning and it is hoped that online learning programs could consequently be designed based on features of both CALL and Mandarin Chinese learning (see Section 2.2.3.3 for more details on CALL).

In relation to participants' attitudes towards various types of annotations, Chun and Plass (1996) have reported that learners' attitudes might reflect the effects that different annotation types have on incidental vocabulary learning. Their study reveals that picture + text annotations are better than video + text annotations and text-only annotations and that more participants report pictures as retrieval cues for remembering words. These findings inspired my interests in participants' attitudes to different types of annotations in my study. It is interesting to know if learners regard Pinyin or audio annotation as important in reading and whether their attitude alters the relevant annotation's effect.

Now that background information concerning research into incidental vocabulary

acquisition has been provided and existing research gaps in the field have been identified, another important issue needs to be addressed to ensure the quality of this research on incidental vocabulary acquisition: how to evaluate the results of related studies. To answer this question, it is critical to understand some issues related to word knowledge, such as the types of word knowledge that can be acquired incidentally through reading, which of these types are examined in this research, and how this study measures knowledge gain and subsequently uses it to explain the results of incidental vocabulary acquisition or draw comparisons with other studies. To the best of my knowledge, these issues have not been emphasised sufficiently in previous studies.

Researchers who explore vocabulary learning have made many efforts to identify what L2 learners need to know about a word from an SLA perspective, such as form (e.g. written, spoken, word parts), meaning (e.g. form and meaning, concept and referents, associations), and usage (e.g. grammar functions, collocations, constraints on use) (Richards, 1976; Laufer, 2012; Nation, 2001). However, previous research on incidental vocabulary learning has been restricted to word meaning and form – or more precisely to retrieving word meaning from a word's written form. In contrast, producing a word's written form according to its meaning has been mentioned in the literature far less, and knowledge related to a word's sound has been barely addressed at all. A possible reason for this situation could be that it is very easy to assume that word meaning is what matters most for L2 learners in the context of paper-based reading. Important studies conducted in this field, such as Nagy, Anderson and Herman (1987) and Chun and Plass (1996), with testing receptive knowledge of words only in their studies may have also influenced this situation.

Apart from the limited types of word knowledge being investigated in the relevant literature, partial word knowledge (which refers to incomplete knowledge of certain types of word knowledge) is another issue that requires consideration in the current study. For example, L2 learners can know exactly what a word means or not know its

precise meaning but have a general idea; the latter situation could be considered as having partial knowledge of a word's meaning. It is well known that learners need to gradually acquire a word through multiple encounters with it, and it is obvious that participants may only have a partial knowledge of words after reading the articles in this study.

In order to evaluate the effects of incidental vocabulary acquisition, no matter which types of annotations were investigated, up to now vocabulary tests have been the only method for measuring word knowledge gain and thus comparing results across studies. Different vocabulary tests and marking criteria could lead to varying results and have consequently influenced previous researchers' views on the effects of incidental vocabulary acquisition, particularly from the perspective of the types of word knowledge being assessed and the amount of word knowledge gained in the process of incidental learning. However, these issues have not been emphasised enough in the field of incidental vocabulary acquisition.

The measurement of incremental word knowledge hence becomes an important issue. Word knowledge, tests related to different types of word knowledge, and criteria for partial word knowledge are presented and analysed in Chapter 3. Features of words in Mandarin Chinese, including structures of Pinyin and Chinese characters, are also included in Chapter 3 to demonstrate the partial knowledge-sensitive criteria for Mandarin Chinese words, as related guidance cannot be easily found in the existing literature on incidental vocabulary acquisition.

Before turning to practical issues related to the design of this research, it is important to mention rationales for studying incidental vocabulary acquisition and factors that affect the results of related research. Previous studies have employed a few theories and hypotheses to explain the mechanism of incidental vocabulary acquisition, such as the input hypothesis (Krashen, 1985), the noticing hypothesis (Schmidt, 1990), the theory of depth of processing (Craik & Lockhart, 1972), the involvement load hypothesis (Laufer & Hulstijn, 2001), and the cognitive theory of

multimedia learning (Mayer, 2009). However, they are not closely related to each other, and not much substantial progress has actually been made to answer questions related to the mechanisms of incidental vocabulary acquisition.

As to the influential factor issue, a number of factors that may affect incidental vocabulary learning have been mentioned in previous studies. However, the factors that have attracted the most attention from researchers are the frequency of encountering a word and effects of various reading comprehension tasks, although neither has been fully understood. Both factors are covered in detail in Section 2.2.3.5.

Now that important issues relating to conducting research in incidental vocabulary learning have been clarified, this section briefly introduces the design of the current study. In order to answer the questions addressed above, two experiments that use an incidental research design are conducted: the first aims to determine effect of the Pinyin annotation to find out whether L2 learners can acquire sound-related information of Chinese words incidentally (with reading as the main focus), whereas the second mainly compares the effects of different types of sound information on incidental vocabulary learning. The first experiment involves a control group that has no access to sound-related information pertaining to target words while reading, and the second one adopts a modified “crossover research design” when comparing the effect of text + Pinyin, text + audio, and text + Pinyin + audio annotations to counterbalance the order of articles, combinations of target words and annotations, and learner differences. The amount of word gain is calculated according to the percentages of correct answers to the vocabulary posttest under marking criteria that are sensitive to partial knowledge. Both experiments involve questionnaires that are mainly aimed at collecting information about participants’ language backgrounds and attitudes towards different types of word knowledge. Data collected from both experiments are analysed quantitatively using the SPSS program, with statistical tests being carefully selected according to the data’s distribution and homogeneity features. The results of the experiments are also later discussed and analysed with a view to

identifying factors that may affect incidental vocabulary acquisition through reading from the perspective of Mandarin Chinese word features.

The results of this study generally provide evidence that incidental learning of sound-related knowledge of words happens in reading and demonstrates the amount of knowledge gain. The results of the first experiment suggest that Pinyin annotation may assist word acquisition in not only sound-related knowledge but also other types of word knowledge relating to the character form of words. Moreover, the results of the second experiment suggest that text + Pinyin + audio annotations are the most helpful of all annotations, although the differences among the annotation types are not statistically significant. In addition, the results of both experiments suggest a word effect indicating some words are easier than others to acquire. A further attempt is thus conducted to identify factors that may account for such a word effect.

This study will contribute to a deeper understanding of many issues relating to incidental vocabulary learning in L2. Firstly, although Mandarin Chinese and sound information were the primary focuses of this research, new topics were added to this field. Secondly, assisted by the self-developed online reading programme, the sound-related information of different modalities (Pinyin and audio) was compared. Although no significant difference was reported in this study, research into the effect of sound-related information could be conducted based on suggestions made in this research related to the improvement of research design.

In addition, in this research the understanding of the incremental nature of incidental vocabulary acquisition was highlighted. This had been pointed out by Nagy et al. (1987) long ago but restricted to acquiring word meanings. With the intention of evaluating the learning results more accurately, an attempt was made to creatively connect to this issue various vocabulary post-tests and criteria that are sensitive to partial word knowledge. This study thus serves as a starting point to elaborate further on the measurement of incidental vocabulary acquisition in the context of L2 Mandarin Chinese learning. It is also expected that researchers interested in other L2 languages

will be attracted to these issues.

Another point that needs to be addressed is that research into incidental vocabulary learning through reading in L2 Mandarin Chinese is still in its infancy; thus many questions remain unanswered in this field. Apart from the issues related to measuring word knowledge mentioned above, it is also not clear which factors will affect the results of research in this field. This research thus provides some preliminary investigation in Chapter 8, starting with factors relating to Chinese character recognition. Based on the results, further empirical studies are suggested to address several factors when choosing target words, including the number of strokes in a whole word, the first character and whether a word contains repeated components. It is hoped that by starting with those factors, the rules of selecting target words will be gradually built up.

The last part of this chapter deals with the detailed structure of this thesis. This thesis consists of nine chapters, which present an introduction, literature review, the methodology used, results, a discussion, a summary, and implications of the research. Following the background information on research into incidental vocabulary learning and reasons for conducting this research provided in this chapter, Chapters 2 and 3 present the literature review. After briefly introducing the development of research on incidental vocabulary acquisition, Chapter 2 mainly reviews studies closely related to this topic, with a special focus on studies involving various types of annotations (including multimedia annotations). It also discusses factors that may affect the learning of Mandarin Chinese vocabulary and characters, in order to identify some character factors that may affect the results of incidental learning. Chapter 3 then focuses on evaluating the results of incidental vocabulary learning, starting with a word knowledge framework that provides a basis for understanding the incremental nature of knowledge gain in incidental vocabulary learning. Details of studies that compare the effects of different types of glosses (especially those that involve audio annotation) are provided thereafter. These details include a further explanation of previous studies'

vocabulary assessments, such as tests and marking criteria.

Chapter 4 presents and explains the study's research questions and major methodological concerns. As this study involves two experiments, one in a pen-and-paper environment and the other in an online environment, detailed methodologies for each experiment are provided separately in Chapter 5 and 6 with the results for each experiment.

Thereafter, a general discussion is presented in Chapter 7 basing on the findings of both experiments. Chapter 8 then presents a further discussion that aims to identify character factors that may influence incidental vocabulary learning. Finally, Chapter 9 summarises the study's significant findings and limitations, describes implications for research in incidental vocabulary learning, and provides a call for future research.

# Chapter 2: The development of research on incidental vocabulary acquisition

With the development of research into L2 vocabulary acquisition, incidental vocabulary acquisition (and particularly the effects of varying types of annotations added to reading materials) has attracted much interest from researchers. This chapter introduces the development of research into this topic and presents a review of related studies. Section 2.1 provides facts about and reasons for the rise of research into L2 vocabulary acquisition to understand its development and the lack of research on L2 incidental vocabulary acquisition. Section 2.2 then presents detailed information on related research, starting from the incidental vocabulary learning hypothesis (which were proposed in arguably the first study conducted in this field). Considering the lack of reviews of relevant studies on L2 Mandarin Chinese, which is the target language in this study, a brief review of Chinese literature is conducted. Finally, empirical studies conducted in both pen-and-paper and CALL environments are discussed, with a special focus on the effects of different types of annotations.

## 2.1 The rise of research into L2 vocabulary acquisition

After a period in which little research was carried out on vocabulary in SLA, the topic has been receiving increasing attention since the mid-1980s. Lu (1984) suggested that the Interlanguage Symposium held in Edinburgh and publication of its proceedings in 1984 can be seen as a sign that L2 vocabulary acquisition was beginning to receive wide attention and returning to the centre stage of language studies.

The rise of research on L2 vocabulary learning can be seen in the increasing number of studies being conducted and the expansion of the range of research topics from both theoretical and empirical perspectives (Sun, 2007; G. Song, 2002; H. Zhou, 2008). Studies have been undertaken to understand vocabulary knowledge (Richards, 1976; Nation, 1990), mental lexicon (J. Nan, 2000; Aitchison, 2012), vocabulary

knowledge measurement (Lexical Frequency Profile, Laufer & Nation, 1995; Vocabulary Size Test, Nation, 2001, Vocabulary Knowledge Scale, Paribakht & Wesche, 1997), learning strategies (e.g. Fraser, 1999; Gu & Johnson, 1996; Lawson & Hogben, 1996; Moir & Nation, 2002), and the proposed lexical approach, which views vocabulary and lexical units as central to language teaching and learning (Lewis, 1993). Several psychological models of vocabulary learning have also been established, such as Kroll and Stewart's revised hierarchical model (1994), the model of lexical representation and development in a second language (N. Jiang, 2000), the lexical processing model (Paribakht & Wesche, 1997), and the three dimensions of vocabulary development (Henricksen, 1999).

Additional evidence of the increasing prominence of L2 vocabulary studies can be found in the journal *Studies in Second Language Acquisition*, which is published by the Cambridge University Press. Two special issues, published in 1987 and 1989, were focused on vocabulary. In the 1987 issue (volume 9, issue 02, June), researchers pointed out that in addition to grammar, vocabulary is also very important in SLA and merits more attention. In the 1999 issue (volume 21, issue 02, June), researchers put a particular focus on incidental vocabulary learning.

The boom in research on vocabulary learning within SLA can also be seen from an in-depth study of sub-questions in the area of vocabulary learning. At present, instead of asking questions as simple as “Is vocabulary in a foreign language important to a learner?” and “Does vocabulary need to be taught?”, Folse (2004) asserts that “L2 vocabulary research has entered a new phase in which we are no longer looking at whether vocabulary should be emphasized but rather what aspect of vocabulary teaching/learning we should focus on” (p. 26).

The amount of research on L2 vocabulary acquisition has been increasing for many reasons. From the perspective of language learning, it has been accepted that without a sufficient vocabulary, learners cannot speak or write a complete sentence, let alone communicate with others. Llach (2011) reviewed studies on language error

gravity and interestingly suggested lexical error to be the most serious error from the perspectives of both language learners and native speakers, as it is more disruptive to communication and comprehension than others.

Apart from the seriousness of vocabulary error, it is also reportedly the most commonly made error in L2 learning. Ye (2002) analysed errors made by 49 L2 English learners in a writing task and reported a much higher proportion of lexical errors (78%) than of grammatical errors (17%), which confirms Meara's (1984) finding of lexical errors being three to four times more common than grammatical errors. Li (2003) and Wen (2006) supported Ye's finding, noting that over 70% of language errors are lexical. In studies conducted with a larger corpus, namely the Chinese learner English corpus (CLEC), lexical errors accounted for 59.33% of language errors in He (2009) and 50.89% in Yang, Gui, and Yang (2005). As awareness of the importance of vocabulary in language learning has grown, researchers have intensified their efforts in this area, so that they can make vocabulary learning easier and more efficient for L2 learners.

From the perspective of theoretical linguistics, the change of Chomsky's theory, for example, increased the complexity and importance of vocabulary learning as a side effect. In Chomsky's *Minimalist Program*, syntax is simplified with an increase in thesaurus information and syntactic issues such as the parameters have become part of the thesaurus (Cook & Newson, 1996; Gass & Selinker, 2001), which suggests a more important role for vocabulary in language research. In addition, theories from other disciplines (e.g. psychology, speech and language pathology, neurology, education, sociology, and cognitive science) have also been applied to L2 vocabulary acquisition. New disciplines, such as psycholinguistics and cognitive linguistics, have emerged against this backdrop. In these fields, vocabulary is sometimes viewed from different perspectives and given a more important place than in traditional linguistics. For example, cognitive linguists reject the separation of the lexical and syntactic and prioritise syntax. They believe that vocabulary, morphology, and syntax should be treated equally and therefore that more attention should be placed on conducting

research related to vocabulary.

Generally speaking, while research on L2 vocabulary learning is growing, it remains far behind research in other areas in SLA (such as grammar, which long dominated the field of language teaching and acquisition). This area therefore requires further research. However, it is against this backdrop that research into incidental vocabulary learning has emerged and begun to attract researchers' attention. Details of related research are presented in the following sections.

## 2.2 Research into incidental vocabulary learning

This section provides a detailed introduction to research into incidental vocabulary learning/acquisition. Initial research into vocabulary learning in a reading context was driven by the realisation of how few of a language's words that students need to know are learned or taught by direct instruction. As clearly explained by Nagy, et al. (1985), a large amount of vocabulary is not learned from teachers or direct instructions, due to the limited time that students attend language classes when learning a language. That is to say, learners need to use many different ways to expand their lexicon, which is where incidental learning comes into play.

Although the expression *incidental word learning* was initially proposed in research dealing with first language learning (Nagy et al., 1987), it currently appears to be of great interest to SLA researchers. Much research has been classified as incidental vocabulary acquisition; however, different views exist as to what can be called incidental vocabulary learning. Definitions of incidental vocabulary learning and details of early research are presented and discussed below, followed by a brief review of studies into incidental vocabulary learning (where researchers' main topics of interest include the effects of different types of glosses and comprehension tasks attached to reading materials).

## 2.2.1 What is incidental vocabulary acquisition?

### 2.2.1.1 The origins of incidental vocabulary learning

The term incidental learning, and its counterpart, intentional learning, have long been researched in the field of psychology. In the experimental literature, incidental learning designs are often seen in memory studies in the tradition of Craik and Lockhart (1972) and Craik and Tulving (1975). In this experimental design, learners are usually involved in a task without any awareness of upcoming tests on some information within the task. Nagy et al. (1985, 1987) adopted this design to investigate vocabulary learning through reading and arguably initiated the use of the concept "incidental word learning from context" in the literature.

In their watershed study, Nagy et al. (1985) investigated the amount of L1 (English) incidental word learning gained through natural reading. Fifty-seventy school children read two articles (one exposition, one narrative) before receiving an unexpected vocabulary test. Difficult words selected from the articles were assessed in an interview test and then through a multiple-choice test. Both tests were designed to measure knowledge gain at different levels (details of the tests are provided in Section 2.1.3). The results of this study suggest substantial and reliable word knowledge gain in both the interview and multiple-choice tests. The researchers reported the probability of learning a word incidentally through reading as 15% to 22% for the multiple-choice test (across different levels of test difficulty) and 11% to 19% for the interview tests. Slightly more learning was reported from tests at a lower level of difficulty; however, the differences were not statistically significant. Based on this experiment, Nagy et al. (1985) proposed the hypothesis: first, that "incidental learning from context proceeds in terms of small increments, so that any one encounter with a word in text will be likely to produce only a partial increase in knowledge of that word" (p. 236); and second, that "learning from context is more effective than many have assumed. Although a single encounter with a word would seldom lead to a full knowledge of its

meaning, we believe that substantial, if incomplete, knowledge about a word can be gained on the basis of even a single encounter" (p. 237).

The incidental vocabulary learning hypothesis clearly demonstrated these researchers' interest in testing word knowledge acquisition exclusively in terms of word meaning. However, instead of providing a clear definition of incidental word learning, Nagy et al. (1985) proposed only the *incidental vocabulary learning hypothesis*, which indicates that a small amount of word knowledge can be learned incidentally through reading and leaves researchers to dispute definitions.

In accordance with Nagy et al. (1985), research in this area commonly referred to either "incidental vocabulary learning" (e.g. Folse, 2004; Hulstijn, Hollander, & Greidanus, 1996; Laufer & Hill, 2000; Wu & Xu, 2009) or "incidental vocabulary acquisition" (e.g. Al-Seghayer, 2001; Chun & Plass, 1996; Duan & Yan, 2004; Ellis & He, 1999; Gao & Liu, 2009; Huckin & Coady, 1999; James, 2009; Paribakht & Wesche, 1999; Yeh & Wang, 2003). In addition, "learning vocabulary from context" was also occasionally employed (e.g. Jenkins, Stein, & Wysocki, 1984; Nagy et al., 1985; Nassaji, 2003).

Inspired by studies discussed above, the incidental learning design was also adopted in this research to investigate L2 Mandarin Chinese word learning through reading. In addition, following most of the previous studies in this field, the notion "incidental learning" is also used in this study. Notably, because it is beyond the scope of this research to distinguish between learning and acquisition, the terms are used interchangeably in this study.

#### 2.2.1.2 Definition of incidental vocabulary learning/acquisition

The lack of an appropriate definition for incidental vocabulary learning has long been recognised by researchers (Singleton, 1997). Attempts to provide one seem to have been hindered by a set of vague notions borrowed from psychology, including, attention, consciousness, and implicit and explicit learning, and it was beyond the

scope of this research to differentiate these complex concepts.

In actuality, researchers always follow a strict operational method when investigating incidental vocabulary learning empirically. In the tradition of Nagy et al. (1985), most researchers of incidental vocabulary learning have adopted the "incidental experimental design", in which learners are required to complete some task (e.g. reading an article) without being told it will be followed by a vocabulary test. This method of not forewarning subjects of a vocabulary test has occasionally been regarded as the key feature that distinguishes incidental from intentional learning (Eysenck, 1982). Laufer and Hulstijn (2001) later labelled it as the methodological meaning of incidental vocabulary learning, in parallel to the general educational definition that was well known in SLA pedagogy.

Therefore, the current researcher accepted Laufer and Hulstijn's (2001) dichotomy of separating the confusing concepts of "methodology meaning" and "educational meaning", and for operational purposes adopted the "methodology meaning" as the definition of incidental vocabulary learning, in which it is regarded as learning in which the learners are not forewarned in pre-learning instructions that a vocabulary test will follow.

However, it should be noted that the acceptance of the methodological definition of incidental vocabulary acquisition in this study was mainly to avoid factors that cannot be easily measured, such as level of attention and quality of mental processing. It does not mean that attempts to define incidental vocabulary learning using the terms mentioned above are meaningless. In the context of natural reading or reading with simple glosses, it is unlikely that readers pay considerable attention to unknown words, which would disturb the reading process. The situation might change, for example, when an individual is reading annotated materials (especially those with multimedia annotations, e.g. pictures, videos and audio in the context of CALL). In this case, the extra information could attract more attention and thus result in a learning process that might not be the same as in the context of natural reading – which means

that keeping the vocabulary test unknown from the learners is not enough to identify the concept of incidental vocabulary acquisition. More research is therefore required to define this term.

Therefore, in the next part of this section definitions informed by both perspectives will be briefly discussed in order to provide background information about research into incidental word learning. It is hoped that a comprehensive understanding of the underlying process of incidental learning will lead to the provision of a more accurate definition in the future.

Schmitt (2000) used explicit learning as the counterpart to incidental learning by stating that there are “two main processes of vocabulary acquisition: explicit learning through the focused study of words and incidental learning through exposure when one's attention is focused on the use of language, rather than the learning itself” (p. 116). This definition is not entirely adequate, as it mixes the concepts of incidental learning and intentional learning with the concepts of implicit learning and explicit learning, which has been identified as a common mistake in defining incidental vocabulary learning as pointed out by Laufer and Hulstijn (2001).

Haynes (1998) shook off the fetters of the influence of implicit/explicit learning and saw attention as the key element for clarifying incidental learning. She regarded incidental learning as a type of automated learning that involves a learner’s peripheral attention while his or her focal attention is on something else (i.e. occupied by attended learning). However, irrespective of whether it should be viewed as the fundamental difference between incidental and intentional vocabulary learning, attention – especially focal attention – seems not easy to measure. As a matter of fact, operational suggestions concerning how to distinguish this focal attention can hardly be found in related literature, which weakens related definitions that employ attention as an indicator of incidental learning as mentioned above.

Similar to Haynes (1998), Nation (2001) suggested that it might be more accurate to differentiate incidental and intentional vocabulary learning from the quality of

mental processing. This perspective is also problematic, as the mental processes involved are not fully understood and standards for measuring these processes' quality in incidental learning are lacking.

The "by-product" construct was introduced as an alternative to attempting to define incidental vocabulary learning from the perspective of a word being attended or not or the extent to which a word is processed. In Paribakht and Wesche's (1999) definition, incidental learning refers to "the process in which learners focus on comprehending meaning of reading and listening contexts rather on the explicit goal of learning new words and acquire vocabulary only as a by-product; while intentional vocabulary learning means the focal attention of vocabulary learning" (p. 176). Nation (2001) put this idea forward by providing some examples of incidental learning. He suggested that "learning from extensive reading, learning from taking part in conversations, and learning from listening to stories, films, television or the radio" can be regarded as incidental vocabulary learning, whereas incidental vocabulary learning does "not include deliberately learning words and their definitions or translations, even if these words are presented in isolated sentence context" (p. 232).

Laufer and Hill (2000) claim that it is a mistake to assume that incidental learning is unattended learning. Attention is the essential condition for learning to happen and thus required in the process of incidental learning as well as in completing original tasks. However, this kind of split attention does not affect the process of completing original tasks. These researchers also adopted the commonly used notion of "by-product" rather than the vague concept of "attention" in their definition, stating that "incidental vocabulary is learned as a by-product of another activity, such as reading or communication, without the learner's conscious decision, or intention, to learn the words" (p. 3).

In this type of by-product definition, learning vocabulary is not the original intention of language-learning activities. In other words, learners are not intended or planned to learn new words from them. This view of vocabulary learning as a

by-product of other activities is frequently found in papers within the special collection of *Incidental L2 Vocabulary Acquisition* (edited by Wesche and Paribakht, 1999). However, the definitions provided in these papers still involve such vague notions as "focal attention" and "conscious decision". In an alternative version of the definition, the ambiguous term "conscious decision" was taken out. Incidental vocabulary was defined as the "learning of vocabulary as the by-product of any activity not explicitly geared to vocabulary learning", while the contrasting term of intentional vocabulary learning was defined as "any activity geared at committing lexical information to memory" (Hulstijn, 2001, p. 271; see also Hulstijn & Laufer, 2001).

However, it should be noted that it is also difficult to decide whether learners are geared to learn some words in the process of reading in the context of L2 learning, especially when the learners are adults. In the context of L1 reading, it is essential for learners to obtain new knowledge from reading materials; however, in the context of L2 reading, the priority of reading might have already shifted to learning some new words, commonly used expressions, or grammar. In this case, even when learners are not told about a subsequent vocabulary test, they expect or are ready to commit several words to memory. It is therefore difficult to determine the extent to which words learned from reading remain by-products.

In another extreme example, Bruton, López, and Mesa (2011) even asserted that the term "incidental vocabulary" should be abandoned. They claimed that the term is impracticable and not very useful for empirical research into vocabulary development in L2 pedagogy due to the intentional–incidental contrast. They propose instead using the term "induced vocabulary salience", which reflects an easily recognisable external intervention.

### 2.2.2 Review of incidental vocabulary learning research in Mandarin Chinese literature

Some excellent reviews on incidental vocabulary learning are available within the SLA discipline (see, for example, Huckin & Coady, 1999; Khatib & Zourzadeh, 2012; Waring & Nation, 2004). It can be clearly seen from these reviews that research on European languages such as German, English, and Spanish dominate the studies in this field, as pointed out by Folse (2004). These reviews' lack of studies that take L2 Mandarin Chinese as the target language could demonstrate a limited interest in this language. Alternatively, it could be due to the language barrier: related research may have been published in other languages (e.g. Mandarin Chinese) but less often discussed given that such research is not easy to access for western SLA researchers. If this is true, Mandarin Chinese literature could thus be an important source for studies on incidental vocabulary learning that take L2 Mandarin Chinese as the target language and needs to be taken into consideration in research. This section therefore briefly reviews the Mandarin Chinese literature to bridge this gap before describing properties of incidental vocabulary learning and recent developments in this field.

The database used to search relevant Chinese literature is the China National Knowledge Infrastructure/Internet (CNKI) database.[1] Importantly, to build a solid base for this research only articles published in core Chinese journals and the Chinese Social Science Citation Index (CSSCI) have been reviewed.

The brief review revealed that the first research relevant to incidental vocabulary acquisition written in Chinese was published in 1994, by Chen and Peng. Although the

[1] Based on the concept of "National Knowledge Infrastructure" proposed by the World Bank in 1998, Tsinghua University in China and the Tsinghua Tongfang company jointly established this database in June 1999. The CNKI is now one of the most important – and perhaps the largest – e-libraries for Chinese resources, including journals, newspapers, dissertations, books, and patents.

target language in this research was Mandarin Chinese, the participants are not second language learners but Chinese children. The earliest empirical studies on learning vocabulary incidentally through reading articles written in Chinese were conducted by Qian (2003), who used Mandrin Chinese as the target language, and Gai (2003), who focused on English.

This review found 68 studies in Chinese literature (including 44 empirical) relevant to incidental vocabulary learning, all of which were published before 14 January 2014. Table 2.1 shows the total number of studies published since 1995 in five-year increments, with the numbers of empirical and non-empirical studies indicated in parentheses.

Table 2.1 Studies related to incidental vocabulary learning published in Chinese

| Year of publication | Number of studies (empirical/non-empirical) |
|---|---|
| 1995–2000 | 2 (1/1) |
| 2001–2005 | 10 (7/3) |
| 2006–2010 | 31 (19/12) |
| 2011–2014 | 25 (17/8) |
| Total | 68 (44/24) |

The table clearly demonstrates that the number of articles increases drastically after 2006, with mostly empirical research; however, only seven of the total 68 studies used Mandarin Chinese as the target language – the rest chose English. It was also found that among the studies that adopted Mandarin Chinese as the target language, participants were often Asian students. For example, both Qian (2003) and Zhu and Cui (2005) discovered that Mandarin vocabulary learning through comprehensible input was very efficient. Participants involved in Qian's study were all native Japanese

speakers, while Zhu and Cui's study included six Japanese students, eight Korean students, three Thai students, and three U.S. students. These results clearly indicate two things. Firstly, L2 Mandarin Chinese as the target language has been under-researched in the field of incidental vocabulary acquisition through reading. Secondly, not enough attention has been given to participants with a non-alphabetic language as their native language, since learning Chinese could mean an entirely different process for learners with different orthographic backgrounds (due to the transference of language capacity and knowledge from L1 to L2).

### 2.2.3 Empirical studies on incidental vocabulary learning

#### 2.2.3.1 Development of incidental vocabulary learning research

Now that it has been acknowledged that research on incidental vocabulary learning in L2 Mandarin Chinese published in Chinese literature is limited, this section presents an overall description of research conducted in the field of incidental vocabulary acquisition as found in both Mandarin Chinese and English literature. It was traditionally assumed that learning from reading is an important source – if not the major source – for vocabulary growth, as it is the only plausible reason for the large but unexplained volume of L1 vocabulary learning that occurs during a child's school years (Nagy et al., 1985). However, empirical studies failed to support this assumption until the late 1980s. Since that time, research has demonstrated that knowledge of previously unknown words, especially the meaning of those words, can be gained incidentally through reading. This includes studies from both the field of first language learning (Jenkins et al., 1984; Nagy et al., 1985, 1987; Shu, Anderson, & Zhang, 1995) and second language learning (Day, Omura, & Hiramatsu, 1992; Duan & Yan, 2004; Dupuy & Krashen, 1993; Elley, 1991; Hulstijn, 1992; Lei, 2011; Pitts, White, & Krashen, 1989; Qian, 2003; Waring & Takaki, 2003; Wu & Xu, 2009; Zhu & Cui, 2005).

Apart from paper-based reading materials, investigation into incidental vocabulary

learning has also been expanded to multimodality materials since 1990s (namely often-adopted listening material, conversations, and videos). Many researchers have focused on the effect of listening materials (Chang & Li, 2009; Sheng, 2011; van Zeeland & Schmitt, 2013; Wang, Yao, & Xu, 2012; Webb et al., 2013; Yang & Hou, 2012) and video materials (Chen & Peng, 1994; Lin, 2010; Wang, 2005; Wang, 2007; Wang, 2012). Although the effects of different materials have not been fully researched, it is generally believed that incidental vocabulary learning occurs when different multimodality materials are used by learners. In studies that compare learning results between listening and reading input, it has been observed that listening input has a better effect than reading input (Vidal, 2011; Wang, 2007; Wang, Yao, & Xu, 2012). This finding challenges the traditional assumption that reading material is a more stable and reliable source than the other sources mentioned above, which was made based on the fact that unknown words do not disappear (as they do in other sources).

The major problem of incidental vocabulary learning through reading mentioned in previous studies is that the incidental learning results are nothing like the results derived from direct and deliberate vocabulary instruction. Irrespective of whether vocabulary learning was investigated under the circumstances of reading short texts (1000–7000 words) or long text (20,000–21,000 words), readers retained no more than seven previously unknown words from their studies (Cho & Krashen, 1994; Day et al., 1992; Horst et al., 1998; Hulstijn, 1992; Knight, 1994; Paribakht & Wesche, 1997; Pitts et al., 1989). In studies that reported the percentage of word retention in the context of incidental learning from reading, the retention of words learned incidentally from reading remained at a very low rate, ranging from 5% to 20% (Schmitt, 2000; Qian, 2003; Hulstijn, 1992; Coady, 1993).

To some extent it should be admitted that incidental vocabulary learning may be not as effective as expected, especially in natural reading or some extreme research conditions in which learners encounter new words a limited number of times. However, it is possible that word knowledge gain has been underestimated in the

above-mentioned studies due to interrelated reasons related to measuring both different types of word knowledge and partial knowledge of words. Firstly, previous studies have focused on measuring word knowledge gain in using receptive knowledge of written word forms (i.e. recognising the written form of words) and ignored other types of word knowledge. Only a few incidental vocabulary learning studies have investigated other types of word knowledge, such as word meaning association (Horst et al., 1998) and collocation (Webb et al., 2013). Several studies have advertently demonstrated the possibility of acquiring productive knowledge of written word forms (Bowles, 2004; Yanguas, 2009) by adopting certain types of vocabulary tests (as detailed in Chapter 3). Secondly, not all knowledge gain can be captured with the vocabulary tests currently being used in this field or marking criteria that are not sensitive to small increments of word knowledge.

The two issues mentioned above demonstrate that researchers' understandings of incidental vocabulary learning are far from comprehensive; as such, results in incremental vocabulary knowledge gain from incidental learning remaining remain under-researched. As these two issues are crucial for understanding the results of incidental vocabulary acquisition, they are discussed in further detail in Chapter 3.

Now that an overview of the development of research relating to incidental vocabulary learning has been presented, the following sections present further information on the topics that have attracted most interest in this field, namely the effects of different types of annotations.

#### 2.2.3.2 The effects of various types of glosses on incidental vocabulary learning

A subject that has been widely discussed within incidental vocabulary learning research is text modification (i.e. modifying reading material to reflect an interventional text format, for example by adding marginal or in-text glosses). Amongst other issues, research has explored the sequence of presenting different information (Chung, 2002). Text modification is also known as "vocabulary assistance" (Alessi &

Dwyer, 2008), “modified input method” (Y. Zhu, 2004), and “词汇强化条件（cíhuì qiánghuà tiáojiàn [vocabulary enhancement conditions])” (L. Lei, 2011) in Mandarin Chinese.

The glossing of word meaning is the technique traditionally used to prepare foreign language reading materials (Bowles, 2004). However, its value for vocabulary learning was rarely researched before the 1990s, when researchers’ interests shifted to comparing the effects of different types of glosses on incidental vocabulary learning (Chun & Plass, 1996; Fraser, 1999; Hulstijn, 1992; Hulstijn et al., 1996; Jacobs, Dufon, & Hong, 1994; Watanabe, 1997). The emerging research revealed that glosses help learners gain greater word knowledge in comparison to an unannotated text (Hulstijn et al., 1996; Watanabe, 1997). Abraham (2008) stated that computer-mediated glosses have an overall medium effect on L2 reading comprehension and a significant effect on incidental vocabulary learning; however, the findings of his meta-analysis of 11 studies are inconsistent with the findings of Chun (2006).

In a paper-based reading environment, definitions or synonyms of unknown words are often provided as glosses in either L1 or L2. In research that involved 195 English L2 learners from a Japanese university, Yoshii (2006) compared the effect of four types of annotation on incidental vocabulary learning, namely L1 text only, L2 text only, L1 text + picture, and L2 text + picture. The results revealed that both L1 and L2 glosses are effective for incidental vocabulary learning, but no significant difference was found between L1 and L2 glosses in definition-only and recognition tests. The participants’ L2 levels were not clear, but it is generally believed that L2 annotation is more suitable for either words with a concrete meaning or high proficiency learners (Zhu, 2004). Researchers have distinguished between low and high levels of learner proficiency by measuring the number of words mastered by learners. Whereas L1 gloss appears to be more effective for learners who know fewer than 2000 words, L2 gloss seems more suitable for advanced learners (Nation, 2009, p. 104).

Beyond the widely compared L1 and L2 glosses, a few researchers have also tested

the effects of the multiple-choice marginal gloss. This very special type of gloss entails providing word meaning options and having readers make a choice according to the context provided in reading. Some studies indicate that multiple-choice marginal glosses are significantly more effective than either single-synonym glosses or no glosses (experiment 3 in Hulstijn, 1992; Duan & Yan, 2004); however, in Watanabe's (1997) research, the single-synonym gloss group appeared to obtain higher mean scores for vocabulary tests than the multiple-choice group. A possible explanation for this result is that without immediate feedback on their choices, learners may be led to memorise their incorrect choice. In other words, if readers choose a wrong answer, it is likely that they may connect the word with this meaning and remember this incorrect information. Nagata (2000) attempted to solve this problem by using a computer software application to provide on-going, immediate feedback regarding readers' selections. The study that used this design showed that the multiple-choice format is significantly more effective than the single-gloss format for helping learners to recall target words. This might be because answering a multiple-choice question requires deeper semantic processing of word meaning, which increases the effectiveness of the multiple-choice gloss.

Miyasako (2002) combined L1 and L2 conditions with the multiple-choice gloss in a study that involved 187 Japanese high school students learning English as an L2. Six types of annotation were used: (1) L2 multiple-choice annotation, (2) L1 multiple-choice annotation, (3) L2 single annotation, (4) L1 single annotation, (5) no annotation, and (6) control (no reading). The tests featured multiple-choice questions. Subjects were required to see each target word in context in the questions. The results showed that the L2 annotation groups (multiple-choice or single) performed significantly better than the L1 annotation groups (multiple-choice or single) in the immediate posttest.

Moreover, Kost, Foss, and Lenzini (1999) compared the three annotation types (i.e. L1 text-only annotation, picture-only annotation, and L1 text + picture annotation) with

56 L2 German learners from a U.S. university. The results of vocabulary tests (which consisted of one multiple-choice question on word definitions and one question that asked the learner to supply definitions of the target words) indicated that the text + picture annotation was more effective than both the picture-only and text-only annotations. The current study thus needs to address the adoption of pictorial annotation, which provides richer information on target words. This type of annotation is one of the often-discussed multimedia annotations and other types include video and audio annotation. These multimedia annotations, which computers can easily present, are discussed in the following section.

#### 2.2.3.3 Research on incidental vocabulary learning in the context of CALL

In the last two decades, research interests in incidental vocabulary acquisition relating to annotation have shifted from the simple text annotation commonly used in the context of paper-based reading to annotation that provides multimodal information in a CALL environment. Before moving to a discussion of studies on multimedia annotations, which directly inspired the interest of this study in investigating the effects of multimodal sound information, the issue of the lack of empirical studies on the real effect of CALL programs, particularly on multimedia annotation L2 Mandarin Chinese, is addressed to provide background information on the reasons for focusing on sound-related information in this study.

As proposed by Beatty (2003, p. 7), a broad definition of CALL is “any process in which a learner uses a computer and, as a result, improves his or her language,” in which the term “computer” was further defined by Levy and Hubbard (2005) as laptops and desktops, as well as “the networks connecting them, peripheral devices associated with them and a number of other technological innovations such as PDAs (personal digital assistants), mp3 players, mobile phones, electronic whiteboards and even DVD players, which have a computer of sorts embedded in them.” In accordance with this definition, related studies conducted in the context of mobile-assisted language

learning (MALL) are thus also included in this section. As the second experiment of this research is conducted in the context of CALL, the findings of empirical studies in related topics should be considered and used as references. Disappointingly, not many studies of the sort were found, as seen from the review in this section.

The widespread use of technology has led to the emergence of a large number of language learning programs and applications, including software, websites, and courseware. Hubbard (2009) has collected and edited CALL-related studies published before 2009, and the four-volume work serves as a good resource for understanding CALL. Within L2 Mandarin Chinese CALL research, some excellent reviews have also been well documented elsewhere in the literature, for example, Yan and Wang (2013), Liang (2015), Liu and Zhao (2012) and Xu (2015). The first application of CALL to L2 Mandarin Chinese teaching involved delivering Chinese character instruction through the Program Logic for Automated Teaching Operation (PLATO) program in the late 1950s.[2] Thereafter, initial interests in computer-assisted Mandarin Chinese teaching and learning focused on characters in the 1980s. The last 20 years have seen an increase in the number of applications for and research topics on CALL in relation to L2 Mandarin Chinese learning. However, unlike research occasionally conducted on reading (Xu, 2006) and vocabulary (McGraw, Yoshimoto, & Seneff, 2009), most studies have not made contributions to research on sound-related annotations in this field. This point is supported by an empirical study, conducted by Liu and Zhao (2012), examining CALL-related research topics in L2 Mandarin Chinese based on a keyword frequency analysis of articles published in related journals and conference proceedings. According to their list, the recent "keywords" were (in order of frequency): the name of the discipline; the internet; multimedia; virtual world; design; courseware; corpus;

[2] PLATO (1958-1993), which was arguably the earliest online language teaching program, was developed and used by the University of Illinois (US) to teach English, French, Mandarin Chinese, Russian, Greek, Latin, Spanish, and the world language Esperanto.

distance teaching and learning; computer-assisted teaching and learning; and model of instruction. Clearly, the effects of multimedia annotations, or even related topics, on reading and vocabulary learning are absent from this list.

In contrast, turning to practical use of the CALL program in L2 Mandarin Chinese learning, such topics are often taken into consideration. For instance, Bourgerie (2003) has listed 13 online Mandarin Chinese reading feeds and Xu (2002) has analysed 30 teaching software programs, including those using multimedia annotations. It is easy to understand that multimedia annotation – for example text, picture, audio, video – can be readily applied by an online language learning program; however, selecting appropriate resources based on one's teaching experience is not reliable and the ease of adopting multimedia resources should not account for adding them to online language learning programs. It was disappointing to discover that most research focuses on designing and developing CALL products for L2 Mandarin Chinese, but that the real effects of most products have not received enough attention or been investigated by empirical studies (Zheng, 2014; Liang, 2015), while the quality of some of the programs reviewed is problematic (Yao, 2003).

The effects of various types of multimedia annotations have not been sufficiently investigated on vocabulary learning, let alone incidental vocabulary learning through reading. More solid evidence is needed from empirical studies, an issue that was also pointed out by other researchers. Zheng (2014) has asserted that it is important to consider the different features of multimedia resources (audio, picture, video, and cartoons) and how they can be applied to achieve certain teaching objectives. Folse (2004) has also openly called for research into the effects of multimedia annotation (although not specifically for L2 Mandarin Chinese). Along with seven other myths in L2 teaching identified in his book *Vocabulary Myth*, published in 2004, Folse points out that research into incidental vocabulary acquisition should focus on discovering "what effects do certain types of marginal annotations and Internet annotations have on incidental vocabulary learning" (p. 28).

Given the aforementioned situation, researchers had made some efforts to the addressed issues above, starting from comparing the effects of different types of multimedia annotations. However, the focus has been on other L2 languages rather than Mandarin Chinese. Previous research on multimedia annotations has mainly focused on text, still picture/image, video, and mixed annotations (Al-Seghayer, 2001; Akbulut, 2007; Chun & Plass, 1996; James, 2009; Plass, Chun, Mayer, & Leutner, 1998; Shahrokni, 2009; Yanguas, 2009; Yeh & Wang, 2003; Yoshii, 2006; Yoshii & Flaitz, 2002). For example, Yoshii and Flaitz (2002) conducted a study similar to the one conducted by Kost et al. (1999) (see details for Kost et al.'s study in Section 2.2.3.2), but in a computer-assisted reading environment instead of in paper-based reading environment. A total of 151 L2 English learners read a text with three types of annotations, namely L2 text only, picture only, and L2 text + picture. Thereafter they were asked to complete both immediate and delayed vocabulary posttests unexpectedly. The results confirmed that text + picture annotation is better than text- and picture-only annotations.

Related results of the studies mentioned above have divergent conclusions concerning the effects of various types of multimedia annotations on incidental vocabulary learning. In studies that included text-only and text + picture annotations, many researchers seem to believe that text + picture annotation is more effective than text-only annotation. However, results become unclear when video and occasionally audio annotations are added into reading materials. In addition, the different research designs used in studies should also lead us to interpreting results of the effects of various types of multimedia annotation on incidental vocabulary acquisition with caution.

The literature review also reveals that the effect of audio annotation has rarely been investigated, even though this type of annotation has been embedded in many online programs. Table 2.2 below provides an overview of studies that incorporate sound information via annotation in reading texts, including the types of annotation

involved and details of the sound-related information presented in annotation. It should be noted that the table includes only studies that aimed to determine the effect of different types of annotation on incidental vocabulary acquisition; those that focus on other issues, such as identifying the effect of different types of annotations on reading comprehension (e.g. Lomicka, 1998) are excluded.

Table 2.2 Studies involving sound-related information in annotations

| Study | Types of annotation reported in the study | How sound information of target word was provided in annotations |
|---|---|---|
| Chun and Plass (1996) | Text (this type of annotation provides the meaning of target words)<br>Text + picture<br>Text + video | All target words read by a native speaker. |
| Laufer and Hill (2000) | Explanation in L2<br>Translation in L1<br>Sound<br>Root<br>Extra | Pronunciation of target words provided by sound annotation. |
| Al-Seghayer (2001) | Text<br>Text + picture<br>Text + video | All target words read by a native speaker. |
| Yeh and Wang (2003) | Text only<br>Text + picture<br>Text + picture + audio | The audio annotation consists of the word's pronunciation and spelling, as well as a sentence with the word embedded. |
| James (2009) | Text<br>Text + Picture<br>Text + video (silent) | Hiragana of Kanji words and pronunciation of target words in the form of a digitised voice recording. |
| Zhu (2004) | Text + Pinyin | Pinyin. |

It is noticeable that researchers sometimes reported types of annotation that were not exactly the same as the annotations they utilised in their studies; in particular, they sometimes failed to report or discuss audio annotations used in their studies

when examining the effects of different types of annotations (e.g. Chun & Plass, 1996; Laufer & Hill, 2000; Al-Seghayer, 2001; Laufer & Hill, 2000). The following points also need to be kept in mind when assessing these studies. Firstly, the sound information provided in annotations varies. Secondly, the studies employed different types of vocabulary tests and marking criteria (Section 3.2 outlines different types of vocabulary tests involved in these previous studies).

Chun and Plass (1996) conducted a set of three often-cited studies that investigated how vocabulary was learned incidentally through reading and comparing the effects of various types of multimedia annotation with a computer application called CyberBuch (which captures each click on a target word made by participants). The second study of the set was used to draw an explicit conclusion on the effects of types of annotation in their study as it involved a large sample size. According to Chun and Plass (1996), the three types of annotation being compared were text annotation, text + picture annotation, and text + video annotation. The research was conducted with 103 second-year university students studying German as an L2. The results showed 24.1% as the mean percentage of correct answers in the L2-based translation test. The researchers also reported 77% of correct answers in relation to the multiple-choice question in study 3. As to the effects of different types of annotation, the text + picture annotation was better than the text- or picture-only annotation. It is noticeable that while audio annotations (of a native speaker pronouncing the target words) were attached to each type of annotation, these annotations were completely ignored in later discussions.

Another point in the study of Chun and Plass (1996) mentioned above that needs to be addressed is the use of the log file. One of their research questions related to determining the relationship between students' look-up behaviour and performance on vocabulary tests. In the program used for this study, students could choose the type of annotation they liked, and the application recorded their clicks to form the log file. The data from the log file were then examined to confirm students' reports about

which annotation they selected and used as the retrieval cues to remember the words. It is interesting to note that the log file assisted the study by providing data on students' clicks. As such, log files may help with determining how many times participants look at an annotation, which is related to the issue on the frequency of encountering target words discussed in Section 2.2.3.5). Due to this consideration, a log file was also employed in the current study. Details concerning how to apply this tool are covered in Section 4.2.7.

The audio annotation was also ignored in a study conducted by Al-Seghayer (2001). In this study, the researcher reported that 30 ESL students read an English article with three types of gloss (namely printed text definition alone, printed text definition coupled with still pictures, and printed text definition coupled with video clips); however, he failed to mention the sound information attached to each type of annotation or include the audio annotation in his later analysis. The overall results of both the recognition and production tests suggested that the text + video annotation (with 87% correct answers for words) had a significantly better effect on incidental vocabulary learning than both the text-only annotation (53%) and the text + picture annotation (67%).

In both studies conducted by Chun and Plass (1996) and Al-Seghayer (2001), It might be reasonable to assume that the effect of audio annotation was offset (as it actually appeared in each type of annotation) and therefore did not need to be discussed when comparing different types of annotation. However, it would be clearer to indicate an audio annotation's existence by reporting that the types of annotation employed in the two studies, namely text + picture + audio annotation, text + video + audio annotation, and text + audio annotation. In this case, the annotations used in the two studies (e.g. text + picture annotation) can be clearly distinguished from the annotations used in other studies, such as the text + picture annotation in Yanguas (2009), in which no sound information was provided.

Irrespective of the conflict between the effects of text + picture and text + video

annotation suggested by Chun and Plass (1996) and Al-Seghayer (2001), which could possibly be caused by the different visual aids and vocabulary tests used in the studies, these researchers support the view that text + picture annotation works better than text-only annotation, and the latter point has been observed by many other researchers (see Section 1.2.3.2). The better effect of combining multimodality annotations is considered to be connected with Paivio's (1986) dual-coding theory. The basic assumption of this theory is that information being processed within verbal and nonverbal channels simultaneously enhances the incidental learning of vocabulary. As such, this theory explains the improved results of using text + picture annotation (in some cases actual text + picture and audio annotation) as opposed to text (plus audio) annotation. Moreover, annotations that contain a short video could assist vocabulary learning as they provide richer information – although they may conversely not be helpful or even impede the learning process by providing information overload.

With respect to studies that focus explicitly on comparing the effect of audio annotation with other types of annotation, the only study found was conducted by Yeh and Wang (2003) using L2 English learners in Taiwan. The aims of this study were to investigate the effect of three types of multimedia annotations and to determine whether learners with certain perceptual learning styles (i.e. auditory, visual-nonverbal/picture, and visual-verbal/text) benefited more from a particular type of vocabulary annotation. A total of 82 first-year college students participated in the research. They were randomly assigned to read an article about Thanksgiving that was annotated using one of the three types of multimedia annotation: (1) text-only annotation with a Mandarin Chinese translation and English explanation, (2) text annotation with a still image associated with the target vocabulary; and (3) text annotation, image, and audio annotation (with a native speaker first reading the word, then spelling it, and finally reading the sentence in which the target word was embedded). The results of this study indicate that the participants who read the article with text + picture annotation outperformed those read it with text-only or text +

picture + audio annotation, which suggests an inhibitory effect of audio annotation. The results also suggest that perceptual learning styles do not affect the effectiveness of vocabulary annotations.

Audio annotation is not beneficial to incidental vocabulary learning in Yeh and Wang's (2003) study for several reasons. Firstly, the transference of language learning skills from L1 Mandarin Chinese to the L2 English learning process may come into play in this case. It is reported that the study's participants tended to obtain information visually and preferred pictorial stimuli, which may be at odds with the participants' language backgrounds. The Chinese speakers may have developed visual skills in the process of acquiring the Mandarin Chinese language; Huang and Hanley (1995) pointed out that learning to read in Mandarin Chinese depends highly on visual skills. The Taiwanese students in this study may therefore have transferred the visual skills that they developed in learning L1 Chinese to L2 English learning. This dependence on visual information explains why the audio annotation was not very useful in Yeh and Wang's study.

Other reasons for the inhibitory effect of audio annotation in the study of Yeh and Wang (2003) could be the quality of the sound information included. The researchers reported that the audio information provided in their program was too fast for participants. The complexity of the audio annotation could also be problematic. In the program, the audio information was provided in a complicated format: a native speaker first read the word, then spelled it, and finally read the whole sentence in which the target word was embedded. With limited time and focus on the reading comprehension tasks, participants may have ignored the information in the audio annotation.

Apart from the weakening effect of audio annotation in Yeh and Wang's (2003) study, the vocabulary tests employed by these researchers may be not in line with the types of annotation used in their study. As mentioned before, the vocabulary tests adopted by researchers in the field of incidental vocabulary learning vary from study to

study; however, they often focus on testing recognition of written word forms (see Section 3.2). In contrast, the three types of vocabulary tests found in Yeh and Wang's study (i.e. multiple-choice questions on word meanings, word association questions, and a cloze test) do not just measure word form and meaning; they also measure word usage according to Nation's framework of word knowledge, which describes knowledge need to know to mastering a word including, for example, receptive and productive knowledge of word written form, word association, and collocation (details on word knowledge frameworks are provided in Section 3.1). It is good to attempt to test various types of word knowledge; however, it would be better to report the scores for each test separately, as these scores could reveal knowledge gain relating to different types of word knowledge (see more details in Chapter 3). More importantly, a test that targets knowledge related to phonetic knowledge of words is perhaps more in line with the audio annotation employed in their research. However, it should be noted that the study conducted by Yeh and Wang (2003) was the only one found that investigated the effect of audio annotations by comparing them with other types of annotations. As such, it is clear that the effect of audio annotation on incidental vocabulary acquisition through reading is under-researched.

It can be seen from the above discussion that many studies seem to add audio annotation just as an accessory to other types of annotation, which means it is often neglected. This was somehow not very surprising to discover, because although reading has been regarded as a valuable resource for incidental vocabulary learning, it has often been connected with acquiring receptive knowledge of written word forms due to its meaning-understanding nature. Although it is possible to acquire phonetic knowledge of a word incidentally through reading, previous studies have failed to demonstrate the effect of sound annotation on incidental vocabulary learning.

In contrast to the above, by employing the Hiragana recognition test, James (2009) demonstrated that sound-related information in annotation for words might help learners to acquire related knowledge of Japanese words. This research was designed

to investigate the amount of incidental vocabulary learning assisted by multimedia annotation, the effects of different types of multimedia annotation, and the interaction between Japanese orthography and annotation type. The 35 third- or-fourth-year L2 Japanese students involved in the study were required to read a short story glossed by various types of multimedia annotations (namely text, text + picture, and text + video). Every Kanji word in the glossary also featured a Hiragana and an audio recording of a native speaker saying the word as a pronunciation aid. The correct answers for both the Kanji and Hiragana recognition tests revealed a very high overall acquisition rate of 87.7%, with 88.6% for Hiragana recognition and 86.9% for Kanji recognition in the multiple-choice questions.

Notably, James' study also indicates that orthography and annotation type interact significantly, although the interaction is different for Kanji and Hiragana acquisition. Text + video annotation was associated with the highest scores for both Kanji and Hiragana tests, whereas text + picture annotation worked better than text-only annotation with Hiragana recognition but not with Kanji recognition. Hiragana acquisition rates, which are consistent with findings in previous studies, suggested that text + picture annotated words were recalled more readily than words with text-only annotation. On the other hand, for Kanji acquisition, text + video annotation is the best, followed by text-only and text + picture annotation. The fact that Kanji acquisition rates for text + picture annotated words were lower than those for words annotated with text only seems to imply that the added picture content may have been not only unhelpful but even detrimental to Kanji acquisition. Unfortunately, the underlying mechanism causing the difference is not clear. According to James, a speculative explanation could relate to the degrees of phonemic transparency of different orthographies. The unique phonemic opacity of Kanji made the acquisition pattern different from Hiragana and other Roman alphabetic languages in previous studies.

Various types of tests may be employed to measure different types of word knowledge. Using multiple-choice questions that required study participants to choose

the correct English equivalent for prompts in either Kanji or Hiragana led James to report 88.6% and 86.9% for Hiragana and Kanji acquisition, respectively; these results suggest a phonemic mastery of Kanji that is beyond the scope of the semantic acquisition of words often attained in the field of incidental vocabulary learning. It was the only study found that reported incidental vocabulary learning relevant to learning word knowledge relating to the sound of words (in this case, Hiragana or Japanese Kanji words in the context of reading).

The above finding sheds new light on measuring word knowledge of nonalphabetic languages, especially Mandarin Chinese, due to the shared features between the Mandarin Chinese and Japanese written systems. As each language contains non-Roman alphabetic characters and written forms of the language's phonetic code, the sound of a written character can be delivered in a written form (i.e. through Pinyin in Chinese and Kana in Japanese) as well as in spoken form (i.e. through someone reading the word). From this point of view, the relationship between Japanese Kanji and Kana is analogous to Chinese Character and Pinyin. As such it is reasonable to hypothesise that the sound of a Chinese word could also be acquired incidentally through reading by providing corresponding information in marginal annotations. This observation directly inspired my research on investigating the effect of sound-related information, namely Pinyin, in the current study; detailed information on this Pinyin form in Mandarin Chinese is revisited in Section 3.3.

In the context of exploring research on incidental vocabulary within CALL, it is also important to examine studies related to mobile learning. The wide ownership of mobile devices has made the potential of using these types of devices to support language learning very evident. As Kukulska-Hulme and Shield (2008) pointed out in their review of developments in "MALL", the range of approaches and learning activities that use MALL is developing very quickly; in the space of two or three years, it has expanded from a purely teacher-learner, text-based model to one that is beginning to support collaborative multimedia listening and speaking activities and allowing

learners to co-construct knowledge to solve problems and fill information gaps (p. 283). In considering vocabulary learning in the context of MALL, Fisher et al. (2012) listed a number of related studies that cover a wide number of topics, including pushing study vocabulary to learners through Email or SMS, taking advantage of memory cycles for vocabulary learning, providing context-relevant vocabulary for learners, and inducing collaboration in learning. However, they also pointed out that majority of these studies focused on intentional rather than incidental vocabulary learning. In other words, studies on incidental vocabulary learning are less common in the context of MALL. Furthermore, if reading is also considered as an important condition for searching literature, studies can rarely be found in the literature. One relevant study was conducted by Todd and Tepsuriwong (2008), who reported vocabulary gains for Thai readers who used an English reading maze on mobile phones. However, the “game” designed by the researchers only presented one or two sentences per screen instead of using typical e-book or article content and format.

Another study, has a research design closer to that of the previous studies. This study, which was conducted by Fisher et al. (2012), compared reading novels that were presented in three mobile modes, namely a paper book, a mobile phone e-book that incorporated an online dictionary, and a mobile phone e-book with enhanced software that provided adaptive vocabulary learning support (i.e. the English Language Mobile system). The results of this study indicate that neither the effect of the three modes of incidental vocabulary learning nor the participants' preferences for the three modes are statistically significant. The result of gaining fewer than two out of 30 target words in each novel confirms the results of previous research that found low acquisition rates of words through reading (as mentioned in Section 2.2.3.1). However, it should also be noted that employing the “crossover research design” (i.e. a longitudinal study in which subjects engage in a sequence of the three learning modes such that each one participates in all conditions) may affect vocabulary learning in research being incidental. In the research of Fisher et al., participants needed to read three books and

take a post-vocabulary test after finishing each one. In other words, the participants may have been aware of the existence of the post-reading vocabulary test when reading the second and third novels, even without the researchers providing any guidance as to test content. Whether the last two conditions can be qualified as incidental vocabulary acquisition is thus questionable.

In summary, the development of modern technology hassled to research on incidental vocabulary acquisition being expanded to the CALL and MALL contexts, and multimedia annotations have been widely adopted in such environments. However, the effects of different types of multimedia annotations, especially audio annotations, remain unclear. Empirical studies have failed to answer related questions, such as whether audio annotation helps with acquiring certain types of word knowledge or what constitutes a good audio annotation. Further studies are therefore still required. In addition, using mobile technologies to support incidental vocabulary learning through reading has not gained much attention in MALL, although this does not mean that potential does not exist. With the increasing popularity of mobile devices, further investigation into mobile-assisted incidental vocabulary acquisition through reading is merited. However, as this is beyond the scope of this research it is not discussed in further detail.

#### 2.2.3.4 Learners' attitudes towards various types of annotations

Studies that compare effects of different types of annotation sometimes mention L2 learners' annotation preferences. Related studies have been conducted in both pen-and-paper and CALL environments to compare learners' attitudes to various types of annotations, including glosses that appear in different places within the reading material (e.g., marginal glosses and glosses at the bottom of each page or at the end of the text), text annotation that is written in L1 or L2, pictorial annotation, and clickable annotation. Regardless of the real effect of annotation, some students prefer to have annotations in L2 rather in L1 (so long as they can understand them), as they have

more chances to practice the target language (Hulstijn et al., 1996). In other studies, researchers have reported learners' annotation preferences for marginal glosses (Zhu, 2004; Kang, 2006), pictorial stimuli (Yeh & Wang, 2003), highlighted multimedia links (De Ridden, 2002), and multimedia dictionaries (Laufer & Hill, 2000).

In most cases, although participants' preferences are not connected to incidental vocabulary learning results, it is reasonable to believe that understanding learners' preferences can help material developers to improve the reading materials they produce. In one recent study, the researcher attempted to connect learners' annotation preferences with the results of vocabulary acquisition through reading. AbuSeileek (2008) conducted a study that compared participants' attitudes concerning the location of annotations (namely at the end of the text, at the bottom of the screen, in the margin, and in a pop-up window). The results suggest that learners prefer hypermedia annotation in the margin; however, no connections between participants' preferences and their vocabulary acquisition were found. This attempt serves as a good starting for exploring the relationship between learners' preferences and incidental vocabulary acquisition, although no clear connections have been found yet. Moreover, defining this relationship requires differences in learners' attitude to be identified first, which calls for further research.

#### 2.2.3.5 Factors that may affect research on incidental vocabulary acquisition

Acknowledging that not many agreements have been reached in studies of incidental vocabulary acquisition, this section deals with factors that may account for related inconsistencies. Researchers have mentioned many factors that may affect incidental vocabulary acquisition, such as language learners' learning strategies (Gu & Johnson, 1996; Nation, 2001; Schmitt & Meara, 1997) and its sub-area of metacognitive learning strategies (Chang & Li, 2009; Sheng, 2011); the amount of text surrounding a target word (Swanborn and De Glopper, 1999); motivation for learners to use glosses if target words are perceived as being related to their future careers

(Lenders, 2008) or freedom to choose reading materials according to their interests (Reynolds & Bai, 2013); genre of the text (Huang, Willson, & Eslami, 2012); and derivation of the word, which could cause difficulties in learning due to morphological complexity (Stock, 1976, as cited in Laufer, 1990).

Several researchers reported factors relating to word properties/features. Waring and Takaki (2003) found that some words are easier to acquire than others, although they do not offer a detailed explanation of why. Moreover, a study conducted by Nagy et al. (1987) identified factors that may affect incidental vocabulary; regression analysis revealed that conceptual difficulty was the only word property that significantly affected learning results. Other word properties involved in their research included the number of occurrences, the target word's length and part of speech, word familiarity, and item difficulty. On the other hand, Paribakht and Wesche (1997) found that part of speech affects incidental vocabulary learning. They reported that with an acquisition rate of 24%, nouns appear to be more easily acquired than verbs (which have a rate of 11%). In support of this finding, Laufer (1990) stated that among words of different classes, nouns can be acquired most easily, adverbs are the most difficult, and verbs and adjectives fall in between. Although this order was not obvious among high proficiency learners in Laufer's study, it was partially evidenced among both less proficient and proficient participants in Lin's (2010) study (which found that nouns and verbs are easier to acquire than adjectives).

However, it should be noted that the possible influences of the above-mentioned factors have not been fully understood. In addition, none of these factors seems to be of primary interest to researchers in the field of incidental vocabulary acquisition. Factors that have attracted much more attention include the frequency of encountering a word and effects of various reading comprehension tasks. Both of these factors are covered in great detail in the following section.

Even a cursory review of the literature on incidental vocabulary acquisition through reading indicates that the frequency of encountering target words has become

one of the most commonly examined variables in this field's growing body of literature. The frequency of encountering a target word, which refers to a target word's number of occurrences within the reading material, is also referred to as the frequency of exposure to target words (Reynolds & Wible, 2014) and the frequency of occurrences rate (Waring & Takaki, 2003).

Accepting that knowledge of a word needs to be built gradually through repeated exposure, it is very easy to speculate that the chances of acquiring a previously unknown word would increase when the number of times this word occurs in reading materials increases. Some researchers even further predict the existence of a certain critical point after which learners are more likely to gain new words incidentally through reading. Attempts have thus been made to identify the minimum or optimal frequency of encountering target words that learners need for acquisition. Waring and Nation's (2004) plausible conclusion suggests that previous studies agree that it takes approximately six to ten encounters with a word to learn it, although this conclusion is not fully supported by related empirical studies.

Many other studies have also yielded findings that related the frequency of encountering target words. Saragi, Nation, and Meister (1978) revealed correlations between word repetition and learning by employing 90 target words with a very large range of encounter frequency (i.e. from one to 209 times) and subsequently identified that a minimum of approximately ten repetitions are needed to learn a word. However, Jenkins et al. (1984) suggested that only about 25% of learners had learned a word after ten meetings. Horst et al. (1998) conducted research with the target appearing from two times to 17 times and suggested that eight times could be the critical point as new words encountered more than eight times would result in an obvious gain in target words. Rott (1999) compared words, which has been encountered two times, four times, and six times. Although higher scores were found on new words encountered six times, and no difference was found between words encountered two and four times, this result only is far less enough for interfering that six times of

encountering is required for acquiring a word. While in the study conducted by Waring and Takaki (2003), the results of comparing the effects of five bands of different number of occurrences, including once, four to five times, eight to ten times, 13 to 14 times and 15 to18 times suggested that a higher number of occurrences of target words, i.e. 15 times, only lead to a 40% chances of target words being learned. Hence, they made a farfetched speculation over the number of occurrences required for higher acquisition rate, which was 20 or even 30 times.

Apart from the inconsistent findings mentioned above, two shortcomings of studies mentioned above further weaken the value of previous research on the frequency issue. The first is a problem with the concept of the frequency of encountering target words. Reynolds and Wible (2014) pointed out that researchers almost always fail to describe how they calculated frequency in their studies. They conducted a study to demonstrate this problem explicitly and to show how different results can be when studies use varying definitions of the "same words", referring to how they count the inflectional and derived variants of target words. With the six studies selected under very strict criteria out of 25 published studies, these researchers found that varying definitions of word repetition not only produce a difference in reported word frequency but also affect acquisition results when reported results basing on the frequency bands of the target words – and therefore may trigger researchers to recommend different target word exposure thresholds. Their findings are hence less valuable in terms of being used as inferencing grounds for determining the number of occurrences needed to acquire a word through reading (Reynolds & Wible, 2014).

Secondly, another important issue related to frequency that needs to be addressed is the difference between the frequency of encountering target words and the frequency of noticing and consequently processing target words. No matter what terminology previous studies have used to refer to what is being counted in relation to a target word (e.g. occurrence, encounters, or exposures), frequency research has

focused on the times that target words appear in reading materials. However, it is not necessarily the case that learners notice all unknown words that appear in a text; they may ignore words, especially when doing so does not affect overall reading comprehension. In such circumstances, the act of noticing and processing new words (which is believed to be essential for vocabulary acquisition) does not simply equal the number of times new words are encountered.

Although no agreement seems to have been reached concerning how many times of encountering a word are required to acquire this word incidentally through reading, previous studies have inadvertently provided sufficient evidence that target words can be acquired through a very low frequency of encounter. In some studies, the number was as low as only one or two times (Nagy et al., 1985; Rott, 1999; Waring & Takaki, 2003). In addition, some researchers have also mentioned that the results of word acquisition would not be affected within a very low frequency of encountering the target words; for example, Qian (2003) and Hulstijn et al. (1996) suggested one to three times, and Rott (1999) noted two to four times.

The effects of reading comprehension tasks on incidental vocabulary learning are also frequently discussed in research related to incidental vocabulary acquisition. Tasks and exercises attached to reading materials were traditionally regarded as methods for evaluating language learning but barely discussed as facilitators of language learning. After Laufer and Hulstijn (2001) posited their task-induced involvement hypothesis, many researchers compared the effects of various tasks on incidental vocabulary learning. These tasks include retelling or summarising stories, filling in blanks, and making sentences with target words (Eckerth & Tavakoli, 2012; Ellis & He, 1999; Gai, 2003; Huang et al., 2012; Joe, 1998; Laufer, 2003; Lei, Wei, Ye, & Zhang, 2007; Li, 2008; Newton, 2013; Paribakht & Wesche, 1997; Wesche & Paribakht, 2000; Wu, Lang, & Duan, 2007). It has also been demonstrated that many tasks have positive effects on incidental vocabulary learning, such as reading comprehension, gap-filling, story retelling (e.g. Ellis & He, 1999; Joe, 1998; Laufer, 2003; Paribakht & Wesche, 1997;

Wesche & Paribakht, 2000; Gai, 2003).

The recent years have seen a small upsurge in comparing the effects of several types of task, perhaps driven by the proposition of the "task-induced involvement load hypothesis/involvement load hypothesis" (Laufer & Hulstijn, 2001). Both input and output tasks are often employed and discussed in research into incidental vocabulary learning. Generally speaking, tasks that only require receptive knowledge of target words should be considered as input tasks; examples include reading articles with glosses or a dictionary and answering questions relating to the content of reading materials (but not target words). In the latter case, target words may appear in the questions or as choice options in multiple-choice questions. However, completing the questions does not require productive knowledge of these words (i.e. they do not need to be written or spoken). In contrast, output tasks should require words to be produced to some extent (e.g. by filling in blanks, writing sentences with the words).

The results of studies investing effects of tasks show that different tasks would result in varying amounts of vocabulary acquisition, especially those that are not close to each other according to the *involvement index.* For example, the high involvement output tasks of writing sentences and compositions often outperform reading comprehension tasks in terms of incidental vocabulary learning (Huang et al., 2012; Lei et al., 2007; Wu, Lang, & Duan, 2007; Eckerth & Tavakoli, 2012). This finding is consistent with the involvement load hypothesis; however, differences between tasks become less obvious if the tasks have a similar involvement load. For example, no significant difference has been reported between "fill in the blanks" (Involvement Index 2) and "reading comprehension" tasks (Involvement Index 1) (Hulstijn & Laufer, 2001). Li (2008) also investigated the effects of fill-in-the-blank and composition tasks and reported no significant differences between them. Without more evidence, it is not clear whether this is due to a defect in the involvement load hypothesis itself or other factors involved in the research design that make the differences less significant.

### 2.2.3.6 Rationales for promoting incidental vocabulary acquisition

This section presents explanations provided by previous studies concerning the principles of incidental vocabulary acquisition. In this light, previous theories have been concerned with cognitive, psychological, and multimedia learning aspects. Theories often employed by researchers primarily include the input hypothesis (Krashen, 1985), the noticing hypothesis (Schmidt, 1995), the theory of levels of processing (Craik, & Lockhart 1972, Craik, & Tulving 1975) , as well as the dual coding theory (Paivio, 1986) and Mayer's cognitive theory of multimedia learning (CTML) with regards to multimedia annotation. In addition, the task-induced involvement theory more recently proposed by Laufer and Hulstijn (2001) is presented at the end of this section. Based on those theories and empirical studies, which support each other, the research design of this study is presented.

Krashen's input hypothesis is often considered the theoretical foundation of research into incidental vocabulary learning. This hypothesis asserts that learners can acquire some language points through comprehensible input (often shortened to "i+1"). When reading, unknown words (which reflect the "+1" part) are placed in a comprehensible context (which reflects "i") and can therefore be acquired by readers. It is clear that having an understandable context around target words is important for learners to successfully acquire unknown words from reading, as it helps them to guess the meaning of these words.

The importance of comprehensible input was also supported by the theory on levels of processing, which is based on the shallow and deep levels of the processing effect identified by Craik and Lockhart (1972). According to this theory, one end of the scale is shallow processing, or type I processing, which includes both structural and phonemic processing; the latter focuses on the surface physical features of words, such as the sound and orthography. At the other end of the scale is deep processing, also known as type II or semantic processing, which deals with meaning and requires more mental activity, thus helping learners to acquire a word (Craik & Tulving, 1975).

According to this theory, because the meaning of a word is needed to comprehend a text during reading, unknown words need to be processed at a relatively deep level. This involves a more meaningful analysis (using images, thinking, and associations) and leads to better recalling words acquired in the reading process.

Reading, therefore, theoretically serves as a good resource for incidental vocabulary acquisition and its effects on this field are worth investigating further. However, practical problems, such as the number of unknown words and noticing such words, need to be addressed when designing related research. Initially, Schmitt (2000) considered the ratio of known words as the most important factor determining the difficulty of reading a text. Studies, though not numerous and not specifically related to L2 Mandarin Chinese, have suggested that mastering 95% (the “i” part) of the running words in an article ensures comprehensible input and, consequently, the acquisition of the remaining 5% (the “+1” part) (Nation, 2001). A meta-analysis conducted by Huang et al. (2012) has also suggested that learners do not learn significantly more vocabulary when they read articles with up to 2% unknown words in comparison to articles with an unknown word ratio of 2-5%. Following those suggestions regarding the ratio of unknown words when reading articles, this research adopts 5% as the threshold of target words. More details can be found in Section 4.2.1, where the design of the reading material is discussed.

However, simply presenting unknown words in context does not lead to words being acquired automatically, as addressed by Schmidt's noticing hypothesis (1995), and readers must at least notice new words before they can correctly guess their meaning and then remember them. In the context of paper-based reading, ignoring unknown words is a strong possibility. This is supported by empirical studies such as Qian's (2003) research, which has shown that 43% of unknown words were ignored in this context. This is a reason to use techniques that increase text salience, such as highlighting, underlining, or glossing, to attract readers’ attention to words, as mentioned in Section 2.2.3.2.

In studies conducted by Bowles (2004) and Yanguas (2009), a substantial increase in the noticing of words was demonstrated when glosses were provided (more details concerning these two studies are presented in Section 2.3). However, the think-aloud tool employed in their studies may have provoked noticing in an "unfair" way, because participants had to notice the words in order to report their thinking of those glossed words. Therefore, it seems that no solution has been developed to ensure participants' noticing of unknown words. As such, it is perhaps more accurate and practical to adopt a tracking tool in the context of CALL, such as the one employed by Chun and Plass (1996), to automatically register study participants' clicking behaviour to a research log (Section 2.2.3.3 contains further details on this study). Although this does not attract readers' attention to unknown words when reading, a log file showing readers' each click provides a clear practical standard for deciding whether annotated words have been noticed or not. More details on the design of the online program and using of tracking tool are presented in Section 4.2.7.

Another point worth mentioning is that, according to Krashen's input hypothesis, comprehensible input alone is not sufficient for acquiring unknown words in an L2 incidentally, through reading. Krashen's input hypothesis was proposed based on his studies on mother language learners and immigrants who learned English in the US. Both kinds of learners were exposed to an environment in which the language they were learning was being used. In relation to L2 learning, a lack of sufficient exposure to a target language could be problematic for language learners.

The idea behind the input hypothesis and theory of level processing mentioned above is that unknown words would be guessed and processed on a semantic level through comprehensible input. Thereafter, learners may able to link the written form of a word with its meaning, thus achieving the goal of acquiring partial knowledge of a word. Although this may well explain what happened in traditional paper-based reading, it does explain the effects of different types of multimedia annotation.

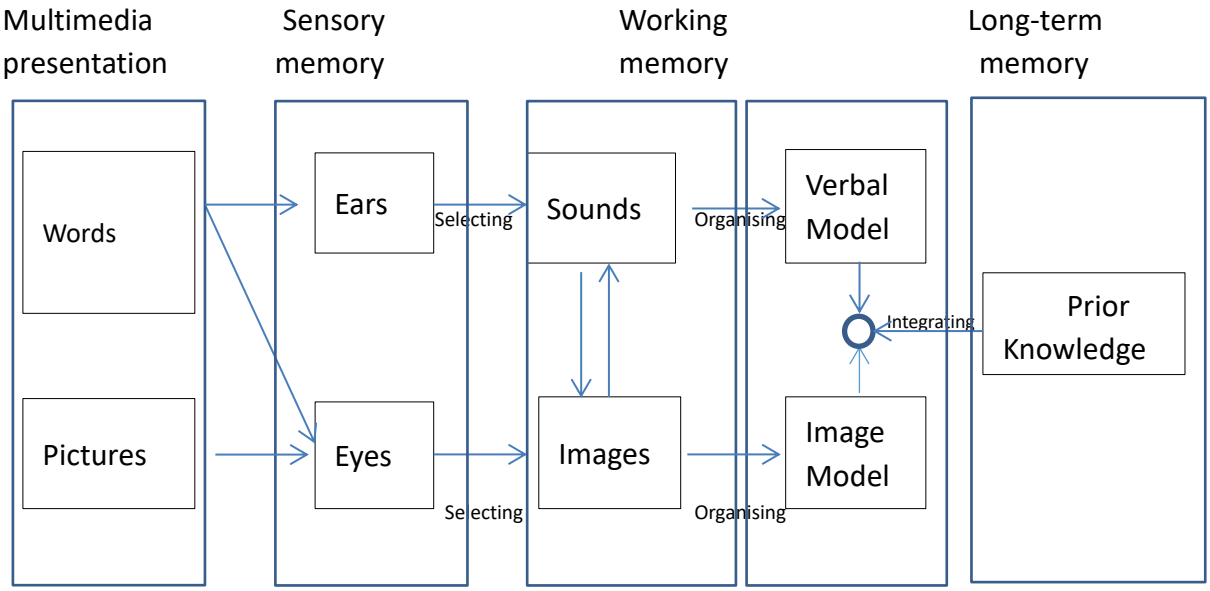
Turning to studies conducted in a CALL environment, both the dual coding theory

and the CTML were employed to explain the effects of different types of multimedia annotation. Several researchers, including Chun and Plass (1996) and Al-Seghayer (2001), have used the dual coding theory (Paivio, 1986) to explain the better learning results attained through multimodality annotations. The basic assumption of the dual coding theory is that information being processed in verbal and non-verbal channels is simultaneously enhanced and processed at a relatively high level. The kind of annotations examined by this theory use verbal and visual materials, such as text in combination with images. The theory explains that the combination of text with a picture annotation – as opposed to a text without a picture annotation or a picture only annotation – is more effective, as it requires a dual-channel processing of information, leading to better learning.

On the other hand, Yanguas (2009) employs the interrelated theory, namely Mayer's cognitive theory of multimedia learning (CTML), to account for the better learning results achieved with the text and picture annotation combination used in his test. The CTML centers on the idea that learners attempt to build meaningful connections between words and pictures, thus learning more deeply than when they are provided with words or pictures alone (Mayer, 2009). Taking three general cognitive science principles as theoretical foundations, namely the limited capacity assumption, and the active processing assumption, Mayer induced sensory modality channels (i.e. visual and aural) in addition to the original presentation modality channels (i.e. verbal and non-verbal). Once verbal and visual information has been selected, reorganized into coherent pictorial and verbal models, and integrated with each other and appropriate prior knowledge, meaningful learning occurs (for more details on the CTML, see Mayer, 1997, 2002; Mayer & Moreno, 2003; Mayer & Johnson, 2008; Moreno & Mayer, 1999). This process is demonstrated in Figure 2.1 below, which is a representation of this theory.

Figure 2.1 Cognitive theory of multimedia learning

(reprinted from Mayer, 2009)




Although no further explicit explanation was given by Yanguas, it is reasonable to believe that the text and picture annotation helped information better engage in the five cognitive processes, which are critical for the learning of information/knowledge to occur in a multimedia environment, more effectively helped with word learning. These processes are as follows: selecting relevant words; selecting relevant images; organizing selected words; organizing selected images; and integrating word- and image-based representations. However, it is also noticed by the researcher that the aural channel had actually been ignored, as the annotations did not include audio annotation for the target word. Therefore, there is a possibility of ensuring an increased effect of annotations if extra aural information is added. This is one of the main reasons that has inspired this study's attempt to investigate the effects of sound-related information involved in annotations. With the expectation to further explore the effect of audio annotation, an online program was developed in the second experiment to accommodate this type of annotation. More details on the design of this experiment can be found in Chapters 4 and 6.

The final theory that needs to be addressed in this section is Laufer and Hulstijn's (2001) task-induced involvement theory, which was also discussed in the previous section. By providing more elaborate distinctions on the levels of processing, Laufer and Hulstijn (2001) have made an attempt to explain the effectiveness of various tasks based on research into comprehension tasks. They introduced the construct "involvement load" into incidental vocabulary learning research by connecting the external notion of the task with the internal concept of level of processing. The involvement load concept combines three factors: need, search, and evaluation. Combinations of these factors and their degrees of prominence form the involvement index, which helps measure tasks and predict the results of incidental vocabulary learning. It is believed that tasks require a higher involvement load when target words are being processed; they thus result in better word retention than tasks with a lower involvement load. For example, Swain and Lapkin (1995) have argued that the better results of output tasks related to learning words are caused by the connection between mental processes and language production. Wittrock (1974) has also pointed out that productive tasks (such as making sentences with target words) may trigger deeper information processing than receptive tasks (for instance recognizing target words being given in sample sentences) and therefore lead to more effective vocabulary learning.

Although it should be acknowledged that Laufer and Hulstijn have made an important initial attempt to establish an operational measurement of various types of tasks, the involvement load hypothesis has yet to be entirely supported by the empirical literature. In addition, it is also difficult to determine the involvement load of all tasks according to this hypothesis. Consequently, this hypothesis was not considered when designing the reading comprehension questions for the first experiment, which could account for better learning results of certain words which appeared in the reading comprehension questions, discussed in Section 5.3.2. Having acknowledged the effects of different reading tasks on incidental vocabulary learning, a modification

of the reading material was subsequently made in the second experiment. It was decided that target words should not be used in reading comprehension questions, particularly not to ask participants to write them when answering the questions, and details are covered in Section 6.1.2.1.

To reiterate, the rationales mentioned above provide the theoretical foundation of this research in terms of designing the reading program. In addition, the CTML is the reason behind investigating the effects of sound information in different modalities in the second experiment.

# Chapter 3: Word knowledge and evaluating the results of incidental vocabulary acquisition

The previous chapter has presented an overview of incidental vocabulary learning, including its definition, the development of related research, and the remaining questions in the field. This chapter evaluates the results of incidental vocabulary reading focusing on a crucial concept, namely incremental knowledge of word, which refers to the incomplete knowledge of a word in incidental vocabulary learning. Two crucial questions, namely what types of word knowledge should be considered and how related word knowledge should be measured, have not been sufficiently researched and therefore remain unanswered. The inconsistency of types of word knowledge tested and the criteria applied in previous studies have hindered the generalization of conclusions concerning the effects of incidental vocabulary learning and consequently the application of the results to research into L2 incidental vocabulary learning.

Based on discussions on theoretical frameworks of word knowledge, as well as empirical studies conducted in this field, this research was designed following a more systematically standard in terms of evaluating word knowledge gain. Practically speaking, the types of vocabulary tests and partial knowledge sensitive marking criteria are central to this issue. It should also be noted that the unique nature of Mandarin Chinese challenges the current framework of word knowledge, as well as testing different types of word knowledge and marking partial knowledge of the word.

This chapter begins by discussing the word knowledge framework, which provides a theoretical foundation for understanding the incremental nature of word learning. Thereafter, previous studies conducted on incidental vocabulary acquisition are analysed with a special focus on identifying the word knowledge involved in the investigations and how researchers tested and measured such word knowledge. The following section then discusses some unique features of words in Mandarin Chinese,

in order to address issues that need to be considered, particularly when designing the partial knowledge sensitive criteria. Finally, this chapter ends with a summary of the literature reviewed in Chapters 2 and 3.

## 3.1 Word knowledge: What does knowing a word mean?

This section explores word knowledge that could be acquired incidentally by identifying every aspect of word knowledge within the word knowledge framework. One could argue that this is not necessary to concern other aspects of word knowledge, as incidental vocabulary learning only concerns word meaning. However, the idea that only a small amount of word knowledge (often refers to word meaning) can be learned incidentally through reading, is a common postulate in this field (Nation, 2001).

When Nagy et al. (1985) proposed the incidental vocabulary learning hypothesis, the incremental nature of incidental word learning was only represented by acquiring partial knowledge of word meaning (namely knowing the inaccurate meaning of a word). In many studies that followed, the small amount of knowledge in question was restricted exclusively to the receptive knowledge of a word's written form, namely the ability to recognize the written form of a word and connect it to word meanings (e.g. Chun & Plass, 1996; Laufer & Hill, 2000; Al-Seghayer, 2001). However, adopting this conventional view that is not backed by empirical studies, is problematic. In recent years, a few studies have challenged this viewpoint by demonstrating the possibility of learning other aspects of word knowledge incidentally through reading with examples including gaining productive knowledge of a word's written form (Bowles, 2004; Yanguas, 2009); receptive knowledge of Kana in Japanese (James, 2009); association (Horst et al., 1998); and collocation (Webb et al., 2013). In this case, further discussions regarding the word knowledge framework are required to understand what knowledge of a word can possibly be learned. Based on this theoretical foundation, the relationship between word knowledge and incidental learning can be researched thoroughly.

Word knowledge, which is defined as “the information about a word which is stored and interconnected in our mind” (Laufer, 2012, p. 1), has been investigated by many researchers. According to Meara (1996), the question “What does knowing a word mean?” was first addressed by Richards (1976) in his very influential paper on vocabulary acquisition; the impact of this question can still be seen in more recent works, such as Nation (1990, 2001). Researchers who attempted to answer this question usually adopted either a learner-centred or a word-centred approach The former provides a macro description of the features of all of the words that a learner has mastered (e.g. Meara, 1996), while the latter entails researchers trying to describe all of the aspects/types of knowledge of each individual word (e.g. Richards, 1976; Nation, 1990, 2001). Nation (2001) later proposed a word knowledge framework that accommodated ideas from both approaches.

Regarding the learner-centred approach, Meara (1996) suggested a learner-centred model that used two dimensions, namely size and organisation, to catalogue learners' lexical competencies. He developed this idea as a three-dimensional framework of word knowledge. This model simplified vocabulary knowledge by asking only three basic questions: how large is the learner's lexicon, how automatically is an item in the lexicon accessed, and how can we simply measure the richness of a lexical structure that links a word in the lexicon. The first dimension can be regarded as being equivalent to the breadth of vocabulary knowledge in Anderson and Freebody’s (1981) binary divisions of word knowledge, which simply refer to the number of words a person knows. The second dimension, automaticity, was referred to fluency in Daller, Milton, & Treffers-Daller (2007). It represents the adequacy of recognition/comprehension speed when reading or listening and the adequacy of retrieval/production speed when speaking or writing (Schmitt, 2010). However, this dimension has perhaps attracted the least research attention among the three dimensions. The last dimension, which has some similarities to the depth of vocabulary knowledge proposed by Anderson and Freebody (1981), is related to the quality of

vocabulary knowledge or how well learners master vocabulary. However, Meara's (1996) focus on this dimension, i.e. the link between words, relates to association more than to other aspects of word knowledge.

The establishment of such general categories of word knowledge was driven by the belief that it is impractical – perhaps almost impossible – to describe and test every aspect about a word that a learner needs to know in an L2. The idea of creating a framework of vocabulary knowledge through a word-centred approach is referred to by Meara as a "research cul-de-sac" (Meara, 1996, p. 4). His comments on the word-centred framework sound reasonable, and the idea of simplifying the framework into three dimensions appears attractive at first glance. However, without detailed instructions and measurements for these concepts, the framework is difficult to operationalise. In addition, it is obvious that the divisions are too ambiguous, and individual words must still be tested in language learning. As such, providing an in-depth illustration of vocabulary knowledge is perhaps more important than first realised.

On the other hand, a simple but well-accepted binary system breaks word knowledge down into receptive and productive knowledge, which is also known as passive and active knowledge. The importance of being aware of the distinction between receptive and productive knowledge has been well documented, as has the necessity to define each type of knowledge and assess it separately for the purposes of language teaching pedagogy (Crow & Quigley, 1985; Laufer, Elder, Hill, & Congdon, 2004; Nation, 2001; Schmidt, 1995). Generally speaking, receptive knowledge helps people to recognise a word when seeing or hearing it and implies that "we are able to comprehend the input and are able to perceive the form of the word and retrieve its meaning"; as such, it is closely related to reading and listening. In contrast, productive knowledge helps people to use a word in speech or writing and implies that "we can retrieve the appropriate spoken or written word form of the meaning that we want to express"; it is thus closely related to speaking and writing.

The association between word receptive knowledge and reading may account for the fact that previous studies on incidental vocabulary learning often ignore productive knowledge of a word in the context of reading. As to the relationship between the receptive and productive knowledge of words, these concepts have sometimes been regarded as different degrees on a continuum of word knowledge. For example, Melka (1997) suggested that learners' familiarity with words could be classified into four stages: imitation, comprehension, reproduction with assimilation, and production. In Meara's (1990) interpretation, however, reception seems to be separated from production as a qualitatively different system, which indicates that receptive and productive knowledge are completely different. Regardless of this disagreement on the relationship between receptive and productive knowledge, this binary system demonstrates researchers' attempts to more elaborately describe word knowledge from a language learner's perspective.

The word-centred approach, which attempts to establish a framework of word knowledge by identifying all aspects of knowledge of a word a learner should know, also merits further consideration. Richards (1976), who was the first to address word knowledge, proposed eight related assumptions. Although the theme he suggested seems more concerned with how classroom practice might be informed by thinking about the linguistics theories of that time (rather than building up a framework of word knowledge), he forced researchers to look more closely at individual words (Meara, 1996). Elaborating on Richards's list, Nation (1990) later proposed the "framework of word knowledge", which illustrates the complexity of lexical knowledge and uses a clear system to categorise vocabulary knowledge of a word. This system involves dividing this knowledge into eight subcategories, namely spoken form, written form, grammatical behaviour, collocational behaviour, frequency, register, conceptual meaning, and association). In 2001, Nation proposed a more complete and systematic summary of the aspects of knowledge that should be acquired to qualify as knowing a word completely. This often-cited framework is discussed fully in the next section.

Similar descriptions of word knowledge aspects have also been provided by other researchers. For example, Folse (2004) mentioned that knowledge of a word normally includes meaning (i.e. polysemy, denotation, and connotation), spelling and pronunciation, part of speech, frequency, usage, and collocation.

The often-cited framework of Nation (2001) must also be discussed in this context. Nation revised his initial framework of word knowledge (as mentioned above) by combing the word- and learner-centred approaches, distinctly different types of word knowledge, and receptive and productive word knowledge into one framework. In this word-knowledge framework, knowledge of a word is catalogued based on three dimensions: meaning, form, and usage. Each area has sub-areas, which are further divided into receptive and productive aspects (designated respectively as "R" and "P" in Table 3.1 below). However, as the current research is primarily interested in word form and meaning, only the word *form* and *meaning* sections dare later discussed.

Table 3.1 Word knowledge framework (reprinted from Nation, 2001, p. 27)

| | | | |
|---|---|---|---|
| Form | Spoken | R | What does the word sound like? |
| | | P | How is the word pronounced? |
| | Written | R | What does the word look like? |
| | | P | How is the word written and spelled? |
| | Word parts | R | What parts are recognisable in this word? |
| | | P | What word parts are needed to express meaning? |
| Meaning | Form and meaning | R | What meaning does this word form signal? |
| | | P | What word form can be used to express this meaning? |
| | Concept and referents | R | What is included in the concept? |
| | | P | What items can the concept refer to? |
| | Associations | R | What other words does this word make us think of? |
| | | P | What other words could we use instead of this one? |
| Use | Grammatical functions | R | In what patterns does the word occur? |
| | | P | In what patterns must we use this word? |
| | Collocations | R | What words or types of words occur with this word? |
| | | P | What words or types of words must we use with this word? |
| | Constraints on use | R | Where, when, and how often would we meet this word? |
| | | P | Where, when, and how often can we use this word? |

Note: "R" refers to receptive; "P" refers to productive.

Some questions of this framework need to be highlighted. Firstly, in addition to *spoken* and *written form*, a *word parts* sub-section indicates a learner's ability to master incomplete knowledge relating to word's form. However, as this information actually deals with more fragmentary knowledge (e.g. prefixes, suffixes, and even letters), It seems inappropriate to include it alongside the other two forms; as such, this sub-section should be downgraded in the framework. Such incomplete knowledge that relates to a certain aspect of word knowledge is defined as partial knowledge in this study. From this viewpoint, the *word parts* information included in Nation's framework is regarded as partial knowledge of a word's spoken and written forms.

Other researchers have provided similar descriptions of the partial knowledge of words. For example, Henricksen (1999) has stated that language learners' knowledge of a word could range from zero to partial to precise. This partial-to-precise dimension of vocabulary development, together with the depth of knowledge and receptivity to productive usability, form Henricksen's three dimensions of lexical competence. Schmitt (2000) has supported the partial-to-precise word knowledge development view. He has asserted that learners could master part of pronunciation or spelling knowledge when exposed to a new word. Moreover, after a few more exposures, these features will be consolidated and other meanings may be encountered. In essence, word learning is not an all or nothing process. To start acquiring a word, learners can normally obtain some aspects of knowledge of this word, as mentioned above. Furthermore, all knowledge acquired in relation to each aspect may not be complete: for example, one could only learn several letters of a word. In terms of words in Mandarin Chinese and considering the word form of a character, partial knowledge can be demonstrated by strokes or components and characters of a word. The features of this language are discussed later in Section 3.3.2 to help draft criteria that can be used in this study.

In addition, Nation's framework consists not only of different aspects of word knowledge, but also links two different aspects of word knowledge, as demonstrated

by the form and meaning knowledge types. These types suggest a further division of word knowledge, such as spoken form/meaning and written form/meaning. However, it should be noted that knowledge of a word's spoken and written forms does not seem to belong to any aspect of word knowledge within this framework. Some may argue that spoken and written forms can be automatically connected. Although this might be the case for languages with a shallow orthography (such as Finnish and Spanish), it might be difficult for deep orthography languages (such as English and Chinese). Moreover, in the case of Mandarin Chinese, an extra Pinyin system would be used together with the spoken and written form of the Chinese character, thus challenging the division of types of word knowledge involved in the framework. This question is also discussed in Section 3.3.2.

Overall, it should be admitted that Nation's framework is the most complete and systematic framework of word knowledge available so far in the vocabulary knowledge literature, regardless of the missing aspect of knowledge of a word's written and spoken forms and the incomplete description of partial knowledge. Knowing a word consists of not only knowing its meaning, but also acquiring various types of knowledge about the word. Consequently, this framework provides a good blueprint for word knowledge assessment that should be followed in vocabulary learning studies. Furthermore, it has therefore been adopted as the theoretical foundation of this research in terms of evaluating incremental word knowledge.

Theoretically speaking, in order to measure incremental word knowledge increased incidentally through reading, one should look at two main catalogues. One is different types of word knowledge including, for example, spoken form, written form, word meaning, and links between each of them. The other is partial knowledge of each type of word knowledge.

However, this becomes complicated when it comes to practical research. It seems that the first catalogue of word knowledge often connects to vocabulary tests, which will be discussed in the following section. In terms of word partial knowledge, no

indication is provided as to how each “part” should be counted. For example, it is not clear whether it should be counted by strokes or components in the context of measuring such knowledge of a word for Mandarin Chinese. Related issues are thus discussed in Section 3.3.2, and the detailed criteria used for marking partial knowledge are presented in Section 4.2.4.

Nevertheless, it is not necessary for an L2 learner to master all aspects of knowledge for a word to be considered as knowing it, as even native speakers cannot master the whole range of knowledge of every word. Therefore, as a starting point, knowledge relating to form and meaning, which is discussed relatively often in the field of incidental vocabulary learning, are covered in discussions in the following sections.

## 3.2 Testing word knowledge in the field of incidental vocabulary acquisition research

Now that is has been acknowledged that knowing a word involves mastering many types of word knowledge and partial knowledge for each of them, the next question is what word knowledge can be acquired incidentally through reading. Theoretically speaking, many (if not all) aspects of word knowledge can be acquired through reading. However, research on incidental vocabulary learning suggests otherwise. It is initially necessary to refer back to Nagy et al.’s (1985) study, as it is arguably the first piece of research into incidental vocabulary learning in English; more importantly, it served as the basis for these researchers’ incidental vocabulary learning hypothesis (for more details, see Chapter 1). It is noteworthy that the knowledge gain in their research refers to receptive knowledge relating to a word’s written form and meaning. Likewise, Chun and Plass (1996) stated that testing recognition, rather than production of words, is more in line with the reading comprehension tasks. Laufer and Goldstein (2004) asserted that the ability to establish the link between word form and meaning is the most important component of word knowledge. Richards (1976) even pointed out that vocabulary learning from natural reading only means understanding a word at the

moment of reading (with only recognition as the level of knowledge), which is far from knowing a whole word and memorising it over a long period. Similar perspectives can also be found in more recent research.

Empirical studies have provided evidence that it is not the case to believe that only acquire word meaning in incidental learning through reading, and the types of word knowledge can be acquired incidentally through reading has gone far beyond the original incidental vocabulary learning hypothesis. A few studies have begun to instead focus on other aspects of word knowledge. For example, in the latest study of Webb et al. (2013), the knowledge of words measured was collocation rather than word form and meaning. Their research proved that collocations can be learned incidentally through reading while listening to a graded reader, evidencing that the knowledge of words that can be acquired incidentally from reading refers not only to word meaning, but also to knowledge of the word's usage (for example, collocation in their case). Moreover, Horst et al. (1998) reported an increase of 16% word knowledge through incidental learning, as measured using an association test. In terms of knowledge relating to word form, productive knowledge can be acquired in incidental learning in addition to receptive knowledge, although to the best of my knowledge this was inadvertently proven by research that employed different types of vocabulary tests. Details of vocabulary tests and knowledge involved in previous studies are presented below.

This section explores what and how word knowledge were investigated in previous studies. It is noted that while types of word knowledge are often related to different vocabulary tests, partial knowledge is usually more related to the marking criteria used for tests. To begin, a selection of research is presented to demonstrate the types of word knowledge involved in this field. Thereafter some details related to the marking criteria are discussed.

As the main method of measuring knowledge gain from incidental vocabulary learning, an overview of the vocabulary tests used in previous studies is presented in

Table 3.2 in the following pages. The studies were selected based on a comprehensive consideration of the types of knowledge involved, the types of vocabulary tests, and the target language. The table provides detailed information on related studies including target languages, vocabulary tests used, and the word knowledge tested in the research. As mentioned earlier, within research into incidental vocabulary learning the emphasis on acquiring receptive knowledge of words incidentally in the context of reading seems to be inevitable – whether researchers acknowledge it or not in their studies. This point can be easily discovered from the table.

Table 3.2 Representative sample of studies on incidental vocabulary learning

| Study | Language | Vocabulary tests | Word knowledge being tested |
|---|---|---|---|
| Nagy, Herman, and Anderson (1985) | English | 1. Interview: providing word meaning<br>2. Multiple-choice questions: six options for each question, three difficulty levels (definitions for other words used as distractors; options repeated in different questions) | 1. Receptive knowledge of written word form<br>2. Receptive knowledge of written word form |
| Nagy, Anderson, and Herman (1987) | English | Multiple choice: five options, including "don't know" | Receptive knowledge of written word form |
| Qian (2003) | L1 = Japanese<br>L2 = Chinese | Multiple choice: three options, the last of which includes space for test takers to write the word's meaning | Receptive knowledge of written word form |
| Chun and Plass (1996) | L1 = English<br>L2 = German | 1. Study 1: Providing L1 equivalents *<br>2. Study 2: Providing L1 equivalents *<br>3. Study 3: Matching test (choosing the German word according to the information provided in a way that parallels the modality in which information was presented to users in the program) **<br>* Reported as a production test in the study<br>** Reported as a recognition test in the study | 1. Receptive knowledge of written word form<br>2. Receptive knowledge of written word form<br>3. Receptive knowledge of written word form |
| Hulstijn and Laufer (2001) | L1 = not mentioned (students from the Netherlands and Israel)<br>L2 = English | Write L1 equivalents or English explanation | Receptive knowledge of written word form |

| | | | |
|---|---|---|---|
| Laufer and Hill (2000) | L1 = not mentioned (students from Israel and Hong Kong)<br>L2 = English | Provide meaning of target words in L1 or L2 | Receptive knowledge of written word form |
| Al-Seghayer (2001) | L1 = various languages (i.e. Arabic, Japanese, Korean, Spanish, and Thai)<br>L2 = English | 1.Recognition test: multiple choice with four options of English equivalents<br>2.Production test**:** definition supply (in L2) | 1. Receptive knowledge of written word form<br>2. Receptive knowledge of written word form |
| Yeh and Wang (2003) | L1 = Chinese<br>L2 = English | 1. Word association questions<br>2. Multiple-choice questions on word meanings<br>3. Cloze test | 1. Association knowledge<br>2. Receptive knowledge of written word form<br>3. Multiple types of knowledge (e.g. productive knowledge of written word form, collocation) |
| Yanguas (2009) | L1 = English<br>L2 = Spanish | 1. Recognition test: multiple choice (choose the English equivalent of Spanish words)<br>2. Production test: provide Spanish (L2) equivalents to English (L1) words | 1. Receptive knowledge of written word form<br>2. Productive knowledge of written word form |
| Waring and Takaki (2003) | L1 = Japanese<br>L2 = English | 1. Word-form recognition test (circle any words recognised from the text)<br>2. Meaning (translation) test<br>3. Multiple-choice recognition test | 1. Knowledge of written word form<br>2. Receptive knowledge of written word form<br>3. Receptive knowledge of written word form |

However, previous studies have not provided clear conclusions regarding their stated objective of investigating the effects of incidental vocabulary learning by measuring or comparing participants' knowledge gain on previously unknown words. The reason is that these studies were plagued by design flaws, which means their results are not comparable. As no accurate understanding of the incremental nature of incidental vocabulary learning and corresponding measurements is available, vocabulary tests vary from study to study in many ways, including the number and language of options (in multiple-choice tests) and L1 or L2 use (in definition-supply tests). Whether recognition or production tests are utilised also differs.

As pointed out by Liang (2005) the design of these mainstream tests has not changed much and in the field of research on incidental vocabulary acquisition, vocabulary tests are mainly formulated using multiple-choice, translation and definition-supply questions. Multiple-choice questions that normally request test takers to choose the right meaning for a prompt (which can be an L1 word, picture, or video used in the study) may provide options in the L1 as in Bowles (2004) or in an L2 as in Chun and Plass (1996) and Al-Seghayer (2001). In contrast, translation questions require an L1 equivalent of the target words (e.g. Chun & Plass, 1996) and definition-supply questions require an L2 explanation of the target words (e.g. Al-Seghayer, 2001). As discussed in the previous section, irrespective of their format all of these vocabulary tests measure receptive knowledge relating to word form and meaning.

It is also discovered from reviewing the studies listed in Table 3.2 above that some studies attempted to adopt more than one type of vocabulary tests to increase the reliability of their results. For example, Waring and Takaki (2003) conducted research to investigate the amount of knowledge gain and rate of forgetting of incidental vocabulary learning, as well as the effect of the frequency of target word occurrence. Fifteen Japanese university students voluntarily participated in the research. After reading a text, they were given three types of vocabulary tests, including a word-form

recognition test, a meaning-translation test, and a multiple-choice recognition test.

However, it should be noted that the different types of tests employed in previous studies give rise to incidental vocabulary results that vary largely. For example, Chun and Plass (1996) suggested a 24.1% – 26.5% correct rate in Studies 1 and 2 based on a translation test (from L2 to L1), while in the Study 3 the learning outcome drastically increases to 77% with a matching test. In a study conducted by Waring and Takaki (2003), the results of an immediate posttest showed the highest mean score for the word-form recognition test of 15.3 out of 25 points (which translates to a 61.2% correct rate), followed by 10.6 points (42.4%) for the multiple-choice test and only 4.6 points (18.4%) for the translation test (L1 to L2). It has been widely acknowledged the difficulty of the translation test (L1 to L2), comparing with the multiple-choice question and matching test. It is therefore reasonable to accept a lower result from the translation test. When comparing the other two types of recognition tests, the immediate posttest demonstrated that the score of the multiple-choice test was more than two times (230%) higher than the score of the translation test; this ratio drastically increased to approximately seven times in the delayed posttest that was conducted three months later. Based on this finding, Waring and Takaki suggested that previous studies had overestimated knowledge gain by using multiple-choice tests. This is consistent with Hulstijn (1992), who considers multiple-choice tests to be better pedagogical classroom vocabulary activities than research tools.

However, it is also possible that such ease of the multiple-choice test was caused by, if not all, the different types of word knowledge, as well as partial knowledge tested. For example, from the perspective of the type of word knowledge being tested, the word-form recognition test adopted by Waring and Takaki (2003) asked participants to circle any words they recognised from the text. The word knowledge tested was thus only receptive knowledge of the written word form; it did not include the corresponding knowledge of word meaning. As such it is easy to understand why the scores for the word-recognition test were much higher than the scores for the other

two tests. The researchers also pointed out that it is important not to provide a context in the test, in order to reduce the possibility of test-takers guessing the meaning of the target words during the test.

Nonetheless, Waring and Takaki's conclusion about the multiple-choice test being easier than the translation test was doubtful, as they did not clearly explain the reason for the score difference. Labelling the multiple-choice test as a prompted meaning recognition test and the translation test as an unprompted form-meaning connection test implies that the researchers were aware of the difference between these types of tests. If partial knowledge of a word is taken into consideration (as mentioned in Section 3.1 and further discussed using empirical studies later in this chapter), the difference between a translation test (from L2 to L1) and a multiple-choice test (with the prompt in L2 and choices in L1) may be further understood: it may be caused by the different amount of partial knowledge concerning the link between words' written forms and meanings. Knowing an inaccurate meaning of a target word may make it impossible for learners to translate that word correctly in a translation test. However, when choices are provided, partial knowledge may be sufficient for the learner to select a correct answer. Unfortunately, researchers have not noted this until now.

Therefore, on top of the notorious difficulty associated with reliably constructing multiple-choice tests and the possibility of such tests being affected by random guessing, the types of word knowledge measured become a further issue for multiple-choice test design (which may account for the higher scores obtained using such tests and may assist investigation into incidental vocabulary learning). Although it is currently difficult to establish why higher scores are achieved on multiple-choice tests than on translation tests, it is reasonable to accept the position that translation tests are more difficult but also more reliable for measuring knowledge gain related to meaning and form obtained through incidental learning (Section 4.2.3 further discusses vocabulary test selection).

As noted above, it is clear that little receptive knowledge related to word form

and meaning in L2 can be acquired through reading. However, the plausible assumption that the small amount of knowledge gain through incidental learning in reading only relates to receptive knowledge of the written form of words is problematic, as proven by studies conducted by Bowles (2004) and Yanguas (2009). These studies, which both compare the effects of different types of annotation, demonstrate that productive knowledge can also be acquired through the process of incidental learning. Bowles (2004) conducted research with L2 Spanish learners (N=50) to investigate the effect of glosses in both CALL and paper-and-pen contexts using three measurements: a comprehension task, an immediate vocabulary posttest, and a delayed vocabulary posttest. Data collocated from the think-aloud protocol indicated benefits for increasing notice on the target words for the two groups that received annotated reading materials. Better reading comprehension and word knowledge gain results were also found in these two groups. More importantly, the production test (which asked participants to write L2 equivalents of L1 words) accidentally revealed an obvious productive knowledge gain by employing production tests.

Yanguas' (2009) study (L2=Spanish, N=94), which was a conceptual replication of Bowles' (2004) study with two additional multimedia annotations (namely pictorial and text + picture), confirmed a significant productive knowledge gain in both an immediate vocabulary posttest and a delayed vocabulary posttest administered three weeks later. As to the effect of different types of annotation, Yanguas pointed out that multimedia annotation helped with word recognition but not much with the word production task. The fact that the text + picture annotation group outperformed other groups in reading comprehension provided evidence for Mayer's CTML, which explains the enhancement of learning by using verbal and visual materials.

The findings of Bowles (2004) and Yanguas (2009) discussed above support (although not directly) the view that productive knowledge gain occurs during the process of incidental vocabulary learning. However, both researchers failed to demonstrate the amount of productive knowledge acquired by participants in their

studies, as it was not their original intention to measure productive knowledge gain. Another point that needs to be addressed is that the think-aloud protocol used to determine noticing in these studies may have increased knowledge gain, as the act of expressing their ideas may have drawn participants' attention to the words and glosses. In addition, Yanguas's suggestion to conduct more studies to assess the effects of different multimedia information based on the ideal theoretical framework indicated by the CTML helped to catalyse this study's comparison of visual (i.e. Pinyin) and auditory (i.e. audio) information (details concerning the CTML are presented in Section 2.2.3.6). Finally, it has to be borne in mind that since the aforementioned studies lacked a clear intention to investigate productive knowledge, related findings were just by-products of the different types of vocabulary tests that they used.

It may worth pointing out that both the vocabulary pretest and posttest that were involved in Yanguas's study used the same questions, that is, translation (L1 to L2) as the production test and multiple-choice question as the recognition test. In contrast, no vocabulary pretest was given in Chun and Plass (1996); participants instead had to identify whether they knew the target words when completing the vocabulary posttest. Although it is important to identify whether participants know target words before a test, repeatedly encountering these words in a pretest prior to reading the article may attract participants' attention and affect a study's results.

The last issue relating to testing and measuring types of word knowledge that needs to be addressed here is the relationship between the production test and the productive knowledge of words. It is noted that Yanguas (2009) mistakenly cited Chun and Plass (1996) to support his conclusion that "production of target vocabulary items not to be affected by the appearance of annotations" (Yanguas, 2009, p. 59). In actuality, in the study conducted by Chun and Plass, word knowledge tested by the production test (see Table 3.2) still reflected receptive knowledge of word form, and not productive knowledge of words as tested in Yanguas's study. This misunderstanding, which has also been pointed out by other researchers (e.g. Read, 2000; Schmitt, 1999),

was due to different understandings of the production test. In Yanguas's study, the production test was used specifically to measure learners' productive knowledge of words. On the other hand, tests that asked learners to provide some answers – such as by writing the meaning of target words in the L1, as in Chun and Plass (1996), or supplying definitions for target words, as in Al-Seghayer (2001) – were also reported as production tests. In Al-Seghayer (2001), the usage of the term "production test" cannot be considered incorrect, as the general understanding of the concept of production is not being violated; however, it might be more meaningful for the field of incidental vocabulary learning if the term provided detailed information about the test, as it did when used by Yanguas. Therefore, to clarify: in this study a production test refers to a test that requires the forms of words, not the word meaning, to be produced (i.e. the target word must be written according to its meaning), whereas a recognition test refers to a test that requires the written form of words to be recognised (i.e. by circling the words that appear in a text or writing the meaning of target words).

Turning to issues relating to testing and measuring partial knowledge, as mentioned earlier in this chapter, we should acknowledge that Nation made great progress in the development of a word knowledge framework (as highlighted in the previous section); however, generally speaking, the simple binary division of word knowledge (i.e. receptive and productive) does not really do justice to the complexity of developing one's word knowledge; many detailed taxonomies are needed to describe the whole process of mastering a word. In relation to relevant extensive reading and vocabulary growth in L2 studies, Nation (2001) also mentioned that "tests were not sensitive to small amounts of learning, did not adequately control text difficulty, and generally lacked careful control of the research design" (p. 155). Researchers have attempted to solve these problems by introducing the notions, for example, the strength of knowledge of word meaning, incremental knowledge, and partial knowledge of incidental vocabulary learning.

Laufer and Goldstein (2004) connected types of vocabulary tests with the strength

of knowledge of word meaning – or more accurately, the strength of the link between word form and meaning. They combined productive and receptive knowledge with recall and recognition to measure the degree of strength. In their theory, the form-meaning link in the mental lexicon can have four degrees of strength: active recall, passive recall, active recognition, and passive recognition. They assessed these degrees using four different tests in sequence, namely: translation (from to L1 to L2), translation (from L2 to L1), multiple choice (with the prompt in L1 and choice items in L2), and multiple choice (with the prompt in L2 and choice items in L1). Referring back to the word knowledge framework, active recall is related to the productive knowledge of a word (which requires producing the written form of an L2 word according to its meaning) and the passive recall relates to the receptive knowledge of a word (which requires providing the meaning according to the written form of an L2 word).

According to Laufer and Goldstein (2004), active recall was once regarded as the strongest of the above four degrees, while passive recognition was considered the weakest. However, it is important to acknowledge that they attempted to define the status of vocabulary knowledge in between “none” and “all”. Frustratingly, this degree of strength view has some weaknesses. Firstly, these initial two degrees do not necessarily indicate the strength of the link between word form and meaning. Secondly, the multiple-choice question used to evaluate the last two degrees, namely active and passive recognition, to my understanding tested the same connection between a word’s written form and meaning. With both the L1 and L2 forms of a target word provided in the question, does it matter the language in which the other options are given? Secondly, the four degrees do not cover the situation in which target words can be partially recognised. An example is a multiple-choice question that uses the prompt “中国” (it means China in English), with the following two sets of choices:

The first set of choices:

A. China B. the US C. the UK D. France

The second set of choices:

A. China B. good C. Sunday D. walk

Learners who know the second character (country) in the prompt can easily spot the right answer if they are provided with the second set of choices. Therefore, Laufer and Goldstein's (2004) attempt to evaluate the strength of knowledge word meaning is problematic.

In reality, partial knowledge in incidental vocabulary learning research usually relates to the meaning of words (which refers to subtle aspects of meaning and involves a word being vaguely comprehended before it is understood precisely). In the existing empirical literature, partial knowledge has been tested and measured in two ways. One relates to carefully designed vocabulary test; taking multiple-choice questions as an example, the choices can be designed to host both a general notion and a more precise meaning of a word (Swanborn & de Glopper, 1999). The other way relates to the marking criteria for questions; for example, in the translation and definition-supply questions shown in Table 3.2 (which ask test-takers to translate the tested word into L2 or explain its meaning in L1 or L2), partial word knowledge can be assigned scores based on criteria identified in different point scales.

Nagy et al. (1985) employed both of these methods: they included a four-point marking criteria scale in the interview and three difficulty levels in the multiple-choice test to measure partial knowledge of word meaning. In the interview, participants were required to provide the meaning of a particular word and their answers were assigned 0 – 3 points based on accuracy. Participants' acquisition of partial knowledge could be recorded at three different levels: at the first level, the answer showed minimal partial knowledge (1 point); at the second level, the answer displayed substantially correct knowledge, but some important components of scoring consistency were missing (2 points); at the third level, the answer was totally correct (3 points). In the multiple-choice test, questions for each word were distinguished to three different levels of difficulty. At the easiest level, the question could be answered correctly with only minimal knowledge of a word (e.g. the general semantic category or knowledge of

the part of speech); at the most difficult level, participants needed to distinguish the target word's meaning from meanings of some closely related words.

Based on the scores gathered using the partial knowledge-sensitive criteria in the interview, Nagy et al. (1985) reported a mean probability of an unknown word being learned (which was calculated by dividing the increased number of words known by the number of words originally not known). This mean probability ranged from 19% at the first level to 15% at the second level and 11% at the third (and most difficult) level. In the multiple-choice tests, the possibility of learning an unknown word was 15% for level three, 22% for level two, and 20% for level one. It is noted that the lowest percentage (11%) of participants obtained a fully correct answer in the interview at the third level; in other words, if the study employed criteria that were not sensitive to partial knowledge, the final result for incidental vocabulary learning would be 11% – which is much lower than the result obtained using partial knowledge-sensitive criteria (for example, 19% and 15% in this study). It is therefore obvious that marking criteria affect the results of studies on incidental vocabulary acquisition (especially in relation to the amount of word knowledge gain), although no significant difference was reported across the criteria levels int this study.

Although no evidence of the influence of partial knowledge-sensitive criteria was found within incidental vocabulary acquisition research, a study that demonstrated such influences was discovered within vocabulary acquisition research. Barcroft and Rott (2010) indicated a significant difference in knowledge gain of words when measured using partial knowledge-sensitive criteria. They examined partial word form learning in both L2 German and L2 Spanish by employing an L1 to L2 translation test. Instead of awarding points only for completely correct words, the researchers awarded 0.25 points for the correct production of partial words. The results indicated the production of approximately 49% more words when partial words were taken into consideration. It is therefore reasonable to believe that such criteria may significantly influence results of incidental vocabulary acquisition. By adopting such criteria, they

also suggested privileging for the word-initial position for both languages and a high percentage of one-letter fragments.

Many other researchers have also applied partial knowledge-sensitive criteria in their studies; however, the refinement of marking criteria has varied. For instance, a 0 – 2 point scale was used by Y. Zhu (2004), Qian (2003), and Granick (1997), whereas Wu and Xu (2009) awarded 0.5 points when the meaning of the word was close to the right answer. Reviewing these studies revealed that some researchers have not been aware of the importance and influence that partial knowledge-sensitive criteria may have. This can be demonstrated by comparing studies conducted by Qian (2003) and Zhu (2004).

Qian (2003) undertook research to test incidental vocabulary learning from reading with Mandarin learners from Japan. All of the participants were university students, although some had learned Mandarin for one year (with 4.5 contact hours per week) while others had studied it for two years (with 4.5 contact hours per week in the first year and three contact hours per week in the second year). It was found that learners' levels and incidental vocabulary learning might be positively correlated, although not at a significant level.

Zhu (2004) used reading material from the study conducted by Qian (2003) to test incidental vocabulary learning, but with students at different levels. He compared the results of his study with those of Qian (2003) and suggested that it is very likely that the higher level of students involved in his study caused the better results obtained on the vocabulary test. He was therefore confident in indicating that incidental learning is more effective for learners at an intermediate level or above. In addition, the higher a learner’s level, the more vocabulary knowledge he/she can obtain from reading. However, Zhu also mentioned the possibility of the results being affected by other causes – for example, the topic and length of the article and the criteria used for marking. Moreover, he suggested that the different topics used in his study’s materials could also be an important factor that affects the results. Although Zhu and Qian both

employed a scale of 0 – 2 points, instead of using the three levels utilised by Qian (namely 0, 1, and 2), Zhu divided the scale into five levels: 0, 0.5, 1, 1.5, and 2. Considering the incremental nature of incidental vocabulary learning, employing more elaborate criteria increases the chances that learners' knowledge gains can be captured later in the research.

It is also worth mentioning that in research that involves multiple tasks in a question (e.g. asking participants to provide meaning and Pinyin and to indicate whether they saw a word in the text), awarding partial points for completing some of the tasks does not qualify as measuring partial knowledge. The request to provide the meaning and Pinyin of target words measures different aspects of word knowledge, namely written form-meaning recognition and the production of Pinyin form; in contrast, the question about whether a word was seen in the article could be regarded as measuring partial knowledge of the character form of target words.

Apart from the vocabulary tests and criteria normally used for evaluating word knowledge gain incidentally through reading, researchers in this field have adopted another type of test: the vocabulary knowledge scale (VKS). The VKS was developed by Paribakht and Wesche (1993) to fulfil their need for a means to record partial understanding of target words. The five-level elicitation scale they devised combines self-reporting with some verifiable evidence of word knowledge in the form of a synonym, L1 translation, or sentence. The five levels are as follows:

I. I haven't seen this word.

II. I have seen this word before, but I don't know what it means.

III. I have seen this word before, and I think it means _____. (Synonym or translation)

IV. I know this word. It means _____. (Synonym or translation)

V. I can use this word in a sentence: _____.

Paribakht and Wesche themselves regard VKS as just a starting point for capturing word knowledge on different levels (including total unfamiliarity, recognition of the

written word/some idea of its meaning, and an ability to use a word with grammatical and semantic accuracy in a sentence). By analysing the design of the taxonomies, it is not difficult to find more than one type of framework/dimension of word knowledge underlying the design of the test. The five levels of the VKS attempt to demonstrate the stages at which a word being acquired partially is precisely mastered. The second level measures recognition of the written word form, while the following two levels assess receptive knowledge of the written word form (with the third level attempting to measure partial knowledge of word meaning). Finally, the fifth-level requirement for test-takers to write a sentence demonstrates many aspects of word knowledge, such as meaning, collocation, and restraint in usage (which indicates that the last level also measures word-centred judgment). It seems that test-takers are expected to choose the first or second level to indicate that they do not know a word or can only recognise it without connecting it to it meaning; alternatively, they may complete one or several of the last three levels to demonstrate different knowledge types and partial knowledge of a word.

A frequent criticism of the VKS is that the five-level criteria chosen to form it are questionable (Douglas, 2010). It seems that the five levels are in fact a mixed attempt to test both different knowledge types and partial knowledge of a word. The statements associated with levels 1 through 4 obviously focus on evaluating learners' knowledge of a target word's form and meaning at various levels, while the last level (which requires learners to compose a sentence with a target word) is actually interested in multiple types of word knowledge. Researchers have also criticised the five levels' limited coverage of word knowledge and the scale's lack of ability to reflect the change of status of people's knowledge of a word. In relation to the last point, no test has so far been able to meet this requirement – although the idea of measuring learners' knowledge of a word from a dynamic perspective sounds very attractive. Although the VKS has many shortcomings, it is a good attempt to assess both multiple types of word knowledge and partial knowledge in a single test. In addition, if

vocabulary learning is understood as not being an all-or-nothing situation, the levels or processes that lie between the two extremes remain unclear. Using the different levels in the VKS may reveal the status between knowing nothing and everything about a word and finally complete the overall picture of the process of learning a word.

Considering the shortcomings of the VKS, researchers have sometimes modified its five levels when adopting it; this has included, for example, Joe (1998); Zareva, Schwanenflugel, and Nikolova (2005); Fraser (1999); and Duan and Yan (2004). The VKS has also been adopted and modified by research conducted to investigate incidental vocabulary acquisition in L2 Mandarin Chinese. For example, Qian (2003) and Zhu (2004) used the following test in their studies:

Sample test use in Qian (2003):

下落 [target word, pronunciation = *xiàluò*, meaning = whereabouts]

A 没见过　　B 见过，但不知道意思　　C 意思是 _____

[Translation: A. I have not seen this word before; B. I have seen this word before, but I do not know its meaning. C. I know meaning of the word, it is_____]

Qian (2003) reported that utilising this type of test enabled her to identify that students tend to ignore some target words in the reading texts as many students chose the option A in the test. Each student noticed only 57% of the target words on average, and this rate has a positive correlation with students' language levels. As one of the main targets of Qian's research was to determine whether the students noticed unknown words in reading, it was reasonable to reduce the five levels of the VKS to these three. However, these levels may need to be redesigned when it comes to L2 Mandarin Chinese learning; in particular, more subtle scales may need to be considered – for example, knowing the meaning of one character in a word. This point is further discussed in the next section, where features of Mandarin Chinese words are presented.

A description of how the effect of incidental vocabulary acquisition was evaluated has been provided above, by analysing what and how different types of word

knowledge were tested and marked in previous studies relating to incidental vocabulary acquisition. To reiterate, the majority of the partial knowledge-sensitive designs and scoring schemes mentioned are related to receptive knowledge of word meaning, i.e. different levels of understanding a word's meaning. Moreover, the vocabulary tests and criteria applied affected the results of incidental vocabulary learning. More research on various types of word knowledge and the system of criteria applied to different types of partial knowledge is therefore required if our understanding of how incidental vocabulary learning occurs is to be improved. For many aspects of word knowledge (e.g. collocation and connotation), it may not easy to measure partial knowledge; however, it is surprising to see that productive word knowledge – which could easily be used to measure partial knowledge – is also missing from research on incidental vocabulary learning.

At this juncture it might be worth mentioning that results concerning partial knowledge gained incidentally through reading may help us in understanding how word knowledge can be built up gradually through incidental vocabulary learning. It has been widely accepted that word knowledge can – and needs to be – acquired gradually. Learners might thus be following certain patterns to establish their knowledge of individual words. However, until now researchers have not agreed extensively on these patterns in English, let alone in Chinese. While the incremental nature has not been completely understood within the field of incidental vocabulary learning, it provides a possibility to illustrate the process of building up knowledge of a word in a more detailed manner. As this study entails target words appearing at a low frequency, it is possible to describe the initial stage of building up knowledge of Mandarin Chinese words through incidental learning.

## 3.3 The unique nature of L2 Mandarin Chinese vocabulary acquisition

The previous section outlined what knowing a word means (from the perspective

of having different types of word knowledge and partial knowledge) and how previous studies have evaluated the results of incidental vocabulary learning; this section now covers some unique features of L2 Mandarin Chinese vocabulary learning. The reason for introducing these features is mainly that Mandarin Chinese is unique to incidental vocabulary acquisition research, as it is a non-alphabetic language and exists outside of the Indo-European language family (which predominates the current research in this field). Although the special orthography in this language (i.e. Chinese characters) has not been fully explored in previous studies, it may influence evaluation issues. As not many studies have investigated this language, it is not clear what types of word knowledge should be tested in the process of acquiring word incidentally through reading in L2 Mandarin Chinese or how partial knowledge should be measured in this process. This section therefore discusses related issues, from the Pinyin system and to the character form of words in Mandarin Chinese.

### 3.3.1 The Pinyin system for Mandarin Chinese

This section first briefly introduces the Pinyin system used in Mandarin Chinese, including its origin, basic structure, and importance to and benefits for L2 Mandarin Chinese learners. The challenges it presents to the current word knowledge framework are also outlined.

The Pinyin system, which is today used widely by both native Mandarin Chinese speakers and L2 Mandarin Chinese learners, adopts the Latin alphabet and tone marks to write Chinese. It was developed based on earlier forms of Romanisation in China and was revised several times after first being published by the Chinese government in 1958. In this system, each Chinese character can be represented by a Pinyin syllable that contains three basic parts: initial part, final part, and tone mark (which indicates the pitch of the whole syllable). The initial part is usually formed by one or two consonants, whereas the final part consists mainly of vowels (although some final parts end with one of two nasals, namely “n” or “ng”). The tone mark is written over one of

the letters in the final part (with the exception of the natural tone, which does not require a written tone mark in a Pinyin syllable).

Regardless of the initial intentions concerning replacing the character system with such a Romanised system in China, the Pinyin system has been actively utilised in adult education to help formerly illiterate individuals undertake self-study after a short period of Pinyin literacy instruction; it is also still used for educational purposes in schools in the mainland China. Within SLA, it serves as a useful tool for L2 learners of Mandarin Chinese by indicating the pronunciation of unfamiliar characters through a relatively shallow orthography. Moreover, it has been found that many L2 Mandarin Chinese learners, especially those with an alphabetic language as their L1, rely particularly on Pinyin in learning (Ma, 2016). Based on these circumstances, Pinyin instructions are provided at the beginning of textbooks and course curricula and delivered early on within Mandarin Chinese classes.

Given the inevitability of L2 Mandarin Chinese learners using Pinyin, one issue relating to word knowledge of Mandarin Chinese words then needs to be addressed: instead of dealing only with the spoken and written forms of words, L2 Mandarin Chinese learners need to deal with the extra Pinyin form. Word knowledge relating to word form in Mandarin Chinese therefore consists of spoken, written (i.e. character) and Pinyin forms, at least for L2 learners. It should be noted that the Pinyin form cannot simply be regarded as a written form for two reasons. The first is that it is not officially used in publications in China. The second reason, which is related to the first, is that Pinyin's main function is to help non-native Mandarin Chinese speakers with pronunciation, although this may not be relevant for learners at a beginner level. To these learners, Pinyin may be much more than only a sound information provider, especially if their L1 does not have characters or a similar written system. However, it is beyond the scope of this research to identify which form Pinyin should belong to; Pinyin is regarded here as sound-related written form, mainly aimed at providing sound-related information for Mandarin Chinese words/characters.

Several researchers have reported that apart from providing L2 learners with sound-related information for words, Pinyin is closely connected to – or even beneficial for – learning word meaning in Mandarin Chinese; this is especially true for learners whose L1 is alphabetic. More precisely, Pinyin may help learners to acquire receptive knowledge of word character form. Previous studies on the relationship between knowing sound-related information (Pinyin or spoken form) for a word and knowing its meaning suggests that L2 Mandarin Chinese learners, especially those with an alphabetic L1, strongly rely on the sound-related forms of words (Pinyin and spoken forms) in learning.

Everson (1998) conducted a study aimed specifically at examining this relationship. The 20 U.S. university students involved in the research were each shown 46 characters that had already been introduced in their Mandarin Chinese curriculum; they were then tested on both pronunciation and translation for each character. The analysis indicated that pronunciation ability was highly correlated with translation ability. More precisely, for any given word that was pronounced correctly, a 90.7% chance existed that the participant would correctly identify it; however, this dropped to only 12.0% when the word was pronounced incorrectly. The results suggest that if participants know how to pronounce a word, it is highly likely that they can recall its meaning as well (and vice versa).

Jiang (2003) found consistent results in her study of U.S. and Indonesian students. Her research involved 74 L2 Mandarin Chinese learners at the beginner's level from Japan, Korea, the US, and Indonesia. During the study, participants were asked to write both Pinyin and meaning according to character forms. The results suggest that knowing Pinyin and knowing meaning are significantly correlated for Indonesian and U.S. students, but not for Korean or Japanese students. Moreover, this relationship was not affected when the target character contained phonetic components, which suggests that learners with an alphabetic L1 tend to rely on the Pinyin in learning Mandarin Chinese. Jiang suggested that this finding was due to transferring learning

strategies from L1 to L2.

Although characters rather than words were involved in both of the above studies (relationships between character and word are discussed later in this chapter), the researchers suggested that acquiring sound-related knowledge of Mandarin Chinese words is beneficial (or perhaps even critical) for connecting the character form with word meaning. However, it should be noted that both of the studies were only able to identify the correlations between knowing the pronunciation or Pinyin with knowing word meaning, which is not sufficient to support their conclusions. Similarly, Service and Craik (1993) stated that the ability of correctly pronouncing a word is very crucial to learn it and Levenston (1979) pointed out that learners tend to ignore words which are hard to pronounce. However, reasons account for such phenomenon were not clearly explained. Instead, findings of research conducted in the fields of word recognition and phonological processing could account for why the L2 learners relied on the Pinyin or pronunciation of word in learning Mandarin Chinese.

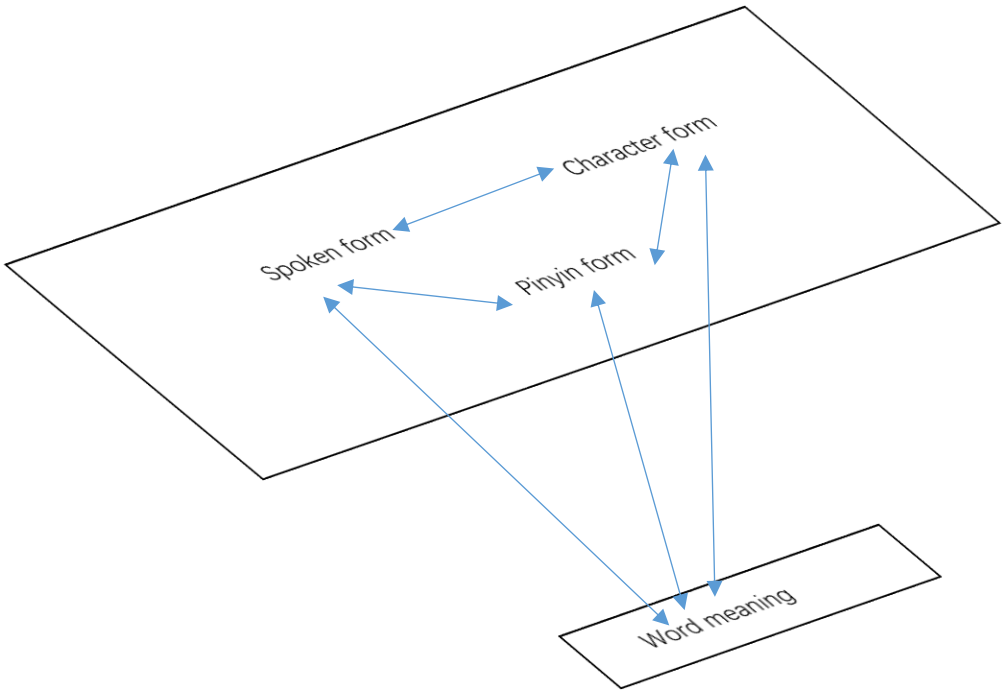
Some researchers have suggested that just as in English, the phonological process in Chinese is obligatory and plays a central role in visual word recognition (Tan & Peng, 1991; Koda, 1997; Tan & Perfetti, 1998; Everson, 1998). In L1 research, well-controlled experiments have repeatedly demonstrated that the phonetic information of a character is automatically activated in character recognition tasks as well as in reading (Perfetti, Zhang, & Berent, 1992; Perfetti & Zhang, 1995) – even before activation of word meaning (Jiang, 2007), but such information may not necessarily mediating the access to word meaning as the “phonology mediation” (a concept widely used in English word recognition since the 1970s) (Perfetti & Tan, 1998; Perfetti & Zhang, 1991). Nonetheless, irrespective of whether phonology mediates access to word meaning, the automatically activated nature of the phonological process suggests a chance to process sound or related information for unknown words through reading and thus to memorise these words to some extent. In this case, it is highly likely that sound-related information in annotations in the process of reading. This serves as another reason for

this research to identify the effect of sound-related information in the context of incidental vocabulary acquisition.

Now that the possibility of acquiring sound-related information through reading in L2 Mandarin Chinese learning is apparent, it is necessary to understand the challenges the Pinyin form presents to both the word knowledge framework and vocabulary testing. According to the above discussion, L2 Mandarin Chinese learners should have knowledge of three types of word form (spoken, written, and Pinyin) as well as be able to connect each form with another form or word meaning to fully master the knowledge related to a word's form and meaning.

The existence of Pinyin challenges the framework of word knowledge as proposed by Nation (2001) and discussed in Section 3.1 and makes it more complicated to explain word knowledge relating to both form and meaning in L2 Mandarin Chinese learning. Considering that the character, Pinyin, and spoken forms and links between word forms and meaning all need to incorporated into the framework's *Word form* section, it is clearer to illustrate the framework using a three-dimension chart rather than a simple two-dimension table, as shown below.

Figure 3.1 Word knowledge framework for L2 Mandarin Chinese
(forms and meaning only)



It should be noted that the two-way arrows in this figure indicate two things: firstly, they illustrate the links between the two elements being connected; secondly, they demonstrate that related word knowledge can be retrieved from both directions. Taking the link between the character and spoken forms as an example, if someone needs to take a note of an address or a person's name, the spoken form is connected with the written form. Moreover, it should be a two-way connection, as the reverse process will happen if someone needs to read information in a note out to others. Although the links between Pinyin and the other two forms cannot be easily demonstrated in daily life, examples can be found for L2 learners given that it is quite common for them to write several words or a sentence in Pinyin down based on what their teacher said (or vice versa, to read what they have written to the class). It might be worth mentioning here that the connection between a word's spoken and written forms is also not covered by Nation's word knowledge framework; only the connections between word form (both spoken and written) and meaning are covered.

Receptive and productive knowledge of a word are also important to consider in this regard. Taking knowledge of word character form as an example, both receptive and productive knowledge relate to knowledge of the character form and the link between character form and word meaning. If the connection between the character form and word meaning is established, the amount of knowledge concerning the character form seems to decide whether the word can be produced (which demonstrates that productive knowledge has been obtained) or must stay at the recognition level (which is a sign of only receptive knowledge). This study therefore tends to break the receptive and productive knowledge to "knowledge on word form and links between word form and meaning" when discussing and analysing the types of word knowledge that have been tested using various vocabulary tests in later chapters. Regardless of whether it is receptive or productive, learners should obtain word knowledge related to forms and the links shown in the figure above gradually by repeatedly encountering a word. Partial knowledge of each form and link in this figure

should thus be considered as an important element of word knowledge.

Moreover, the Pinyin form also complicates the design of vocabulary tests. It is perhaps worth mentioning the vocabulary test used in the HSK[3] – although as a counterexample, given that the format did not serve the test's original purpose. A sample question is given below (Hanban, 2011a):

| | | | |
|---|---|---|---|
| 21. |  | māo<br>猫<br>[Translation: cat] | |

(Sample paper code: H11004, p. 6)

This test requires test-takers to decide whether the word given (in character and Pinyin) matches the picture on the left by marking a tick or cross in the blank. As this test is part of the *Reading Section* of the test paper, it is supposed to test recognition of formal Chinese written forms that are characters rather than Pinyin. However, providing the Pinyin above the words blurs the testing objective, and the test consequently fails to fulfil its original objective. Instead of testing recognition of Chinese words to the level of understanding their meaning, the focus could be on recognising a word's Pinyin. In addition, with the "true or false" format, participants stand a high chance of guessing the answer correctly without knowing a word's real meaning. Moreover, if pictures must form part of the questions, words with abstract meanings cannot be easily tested. Due to the lack of awareness of the changes in word knowledge brought by the Pinyin form, this test does not serve its original purpose as a

[3] HSK, which stands for "汉语水平考试 (Hànyǔ Shuǐpíng Kǎoshì) [ Mandarin Chinese Proficiency Test]", is China's only standardised international Chinese proficiency test for non-Chinese native speakers. This test is administered by Hanban, a non-governmental organisation affiliated with the Ministry of Education of the People's Republic of China. The HSK was reformed in 2007, with its difficulty reduced to meet the requirements for the language test for non-Chinese native speakers. The number of test levels was reduced from 11 to six and Pinyin was added to corresponding Chinese characters throughout the test for levels one and two.

Mandarin Chinese reading test.

### 3.3.2 The character form of Mandarin Chinese words

Now that issues related to Pinyin in L2 Mandarin Chinese vocabulary learning have been discussed, in this section we move on to issues related to the special orthography of Mandarin Chinese (i.e. the character form) to further understand what constitutes partial knowledge of Mandarin Chinese words, factors that might influence Chinese character recognition and the necessity to provide Pinyin or audio annotation to assist L2 Mandarin Chinese learning. It should be borne in mind that, as mentioned in the literature section, few relevant studies have dealt with Mandarin Chinese as the target language in the field of incidental vocabulary learning. Studies concerning statistical research on Mandarin Chinese word and character, and character recognition were hence briefly introduced to help identify features of words that might affect this study. This section starts with topics related to some features of words before moving on to some features of characters, which serve as important units of words.

Mandarin Chinese has been through a process of vocabulary bisyllabilization, and most of its words consist of two characters, instead of only one character, in the modern Mandarin Chinese. Evidence of two-character words being the majority in modern Mandarin Chinese can be found in censuses conducted using dictionaries and vocabulary lists. For example, Zhou (2004) undertook a census of Chinese words in the *Modern Chinese Dictionary* (1996 version) and found that among the 58,481 items included, 67.63% (or 39,548 items) were two-character words; this rate may increase if only the most frequently used words are considered. A similar result was also reported by Wan (2012). According to Wan's census conducted using *The List of Frequently Used Words in Modern Mandarin Chinese* (published in 2008), the 56,008 words in the list can be classified as follows: 3181 are formed of single character; 40,351 are formed of two characters; 6459 are formed of three characters; 5855 are formed of four characters; and 162 are formed of five characters. Knowing that Mandarin Chinese

words are comprised of different numbers of characters, it is reasonable to assume that partial word knowledge can be measured on a character basis. However, as single-character words do exist, identifying partial knowledge should go deeper beyond the character level instead of stopping there. Details concerning the basic structures of Chinese characters and factors that affect character recognition are therefore presented in the following paragraphs.

Thanks to interests in the characters of Mandarin Chinese, which can be dated back to the early 1920s, a large number of studies on character recognition were found. Researchers have focused on elaborated and smaller units of characters, namely strokes and components. Before details of these studies are presented, a brief introduction to the basic units and structures of Chinese characters is needed.

A Chinese character can be made up of several components – or more precisely, by several strokes. The strokes have a function that is similar to the function of letters in English words, while the components (which are formed by several strokes) are similar to certain combinations of multiple letters in English words. Some components of characters, which are known as phonetic or meaning components, may carry information related to pronunciation or word meaning. According to Wu et al.'s (2005) census based on the *Chinese national standard code for information interchange*[4]*,* the average number of strokes in Chinese characters is 10.62 (although individual characters may contain one to 30 strokes). According to Zhou and Chen (1998), the number of components for each character ranges from one to 13.

[4] Also known as *GB2312*, this code contains a key official character set of the People's Republic of China that is used for simplified Chinese characters. GB abbreviates guójiā biāozhǔn (国家标准), which means national standard. The 6763 characters listed in this set cover 99.99% of the characters of contemporary usage.

Unlike the special structure of English words (in which letters are always arranged one after another), strokes and components can be placed in various places in Chinese characters. For example, strokes can cross and be attached to or detached from each other, and a component can be put next to, above, below, or around another component. In addition, as opposed to written English, in which there is a space between each word, there is no such space between words or characters in written Chinese.

Another point that needs to be addressed here is the two standard character sets of the contemporary Chinese written language, namely simplified Chinese characters and traditional Chinese characters. The simplified characters were created by decreasing the number of strokes and simplifying the forms of a sizeable proportion of traditional Chinese characters, but not by creating new strokes or structure of characters. This set of characters has been promoted since the 1950s in an attempt to increase literacy. However, it should be noted that many Chinese characters were left untouched. Wang (2004) points out that only 31.9%, that is 1,116 of the most common 3,500 character in the List of Frequently Used Characters in Modern Chinese, published by the Ministry of Education of the People's Republic of China in 1988, underwent simplification. Currently, simplified Chinese characters are officially used in the People's Republic of China and Singapore, while the traditional Chinese characters are mainly used in Hong Kong, Macau, Taiwan and some overseas Chinese communities. In the curriculum of Mandarin Chinese at UK universities, learning simplified Chinese characters is required and it is only at SOAS, so far as I understand, that the traditional form is also required, but for Chinese major students only.

With this background information on characters in mind, we now move on to topics relating to Chinese character recognition. Related studies have been conducted with samples of both Chinese native speakers (e.g. Zhang & Feng, 1990) and L2 Mandarin learners (e.g. Hayes, 1987; Jiang, 2004a; Jiang, 2004b; You, 2003). Although it is still difficult to describe the whole procedure of recognising a character and it is

not even clear whether processing Chinese characters starts with the whole character or components of the character, empirical studies have suggested several factors that might influence character recognition, including the frequency of characters, the number of strokes, the structure of characters, the position and frequency of components, and the number of components and special components (Chen, & Huang, 19999; Luo, Chen, & Peng, 2007; Luo et al., 2010; Gao, Zhong, & Zeng, 1995; Peng & Wang, 1997; Wan, 2003; Zhang & Feng, 1992; Zhang, 2002).

Luo et al. (2010) conducted an experiment with a sample of 27 Chinese university students to investigate the effect of Chinese whole character recognition over component processing by comparing three types of stimuli: real characters, pseudo characters and component characters. Both the Reicher-Wheeler task, which presents the stimuli before the target components, and a component judgement task, which presents the target radical before different types of stimuli, were adopted. The results of their experiments suggested that the reaction time for left-right structure characters was shorter than those for the top-bottom characters, and the spatial position of the components had significant influence component recognition, indicating structure effect and position effect in Chinese character recognition.

In another study conducted by Luo, Chen and Peng (2007) to investigate the relationship between whole characters and the components primed by the whole character, and components and stroke. A special task was adopted which required the participants to decide whether the strokes provided constituted a real Chinese character. Rather than using tasks that are often employed in character recognition research, e.g. the character/component naming task and the real character/pseudo character judgement task, this stroke-form-character task tests the production of a character rather than recognition of it. As mentioned in the previous sections in this chapter, with different types of word knowledge being tested, different tasks might lead to different interpretations of the results of studies. Although this point has attracted the attention of some researchers, such as Luo et al. (2010), it has often been

ignored in relevant studies. Therefore, this point needs to be borne in mind when interpreting the results of previous research into Chinese character recognition.

Turning to the stroke number effect, although some researchers believe that a stroke number effect does not exist, given that characters are processed as whole units in the human brain (Gao et al., 1995), empirical studies have repeatedly shown that stroke number might affect learning results. For example, stroke number affected the character writing tasks but not the recognition tasks in a study conducted by X. Jiang (2000). Another example is a study by Xiao (2002), which was conducted with a sample of 34 beginner's level Chinese learners. The participants performed better in both recognition and production tests when the characters had fewer than six strokes (as opposed to seven to 11 strokes or more than 12 strokes). In the recognition test, no significant difference was found between characters with seven to 11 strokes and characters with more than 12 strokes. Peng and Wang (1997) took this a step further by taking a character's number of components into account and demonstrated that each component's number of strokes and a character's number of components affected character recognition (in terms of processing time).

However, there are still debates about the role of strokes in the cognition of Chinese characters. Some researchers believe that the strokes are not good measuring indexes of the features of Chinese characters. Hayes (1987) inferred that the number of strokes might not be the reason that learning is difficult and proposed instead using the "complexity of character" concept to explain learning difficulty. He asserted that complexity was caused *inter alia* by combined strokes and difficult structures. However, this point of view is not supported by empirical studies, because there is no clear index indicating how this complexity should be counted. Researchers have so far only been able to demonstrate the influence of a single feature of Chinese characters. Apart from the stroke number effect, Zhang, Wang, Zhang and Zhang (2001) found stroke repetition effect and suggested that high-repeated stroke Chinese characters were recognised more quickly and with a lower error rate than the low-repeated-stroke

Chinese characters; You (2003) found that the left-right structure was the most difficult for Chinese learners, compared with single, top-bottom and surrounding structures; Yu (1998) suggested a position effect because it is easier to recognise components in the left part of a left-right structure character, as well as the top part of a top-bottom structure character.

It was initially hoped that studies on Chinese character recognition would help to identify factors which might affect recognition and possibly incidental learning of words in Mandarin Chinese, and consequently should be considered when choosing target words for this study. Although it can be seen from the brief review of Chinese character recognition that factors relating to stroke, component and structure of characters might affect the process of character recognition, most of the research was conducted with Chinese native speakers focusing on recognising characters that were previously known. It is not yet known whether these factors will work for L2 Mandarin Chinese learners, so further empirical evidence is required to answer this question. In addition, in most studies the focus was on single characters rather than words, which are the target of this investigation. Even if character is the basic unit of word, the procedure of recognising a two-character word is very unlikely to be simply equal to that of recognising two single characters together. This point, however, has not been adequately researched. Furthermore, without universal agreement about which factors contribute to the difficulty of a character, it was not clear which factors could be considered to help with target word selection. Therefore, factors related to Chinese character recognition were not considered in the process of choosing target words for this research. However, such factors might have helped to reduce the difficulty of learning certain words in this study and thus affected the participants' learning results, as discussed in Chapter 4. Thus, with the data collected in this research, further analysis was developed to identify factors that might affect incidental vocabulary learning in Mandarin Chinese, as described in Chapter 8.

At last, the deep orthographic system of Mandarin Chinese needs to be addressed

in this chapter to further explain the importance of providing Pinyin and audio annotations in this study. The very deep orthographic system means that Chinese characters do not impart phonological information directly. However, Chinese native speakers can sometimes guess a character's pronunciation fully or partially correctly through its phonetic component (also called an idea-phonetic or idea-sound character), which provides sound information in pictophonetic characters. However, for L2 learners (especially those at a beginner's level), it seems to be very difficult for the following reasons. Firstly, in order to extract a character's phonological information, logographic readers must have the awareness that some characters are made up of two parts, namely a semantic radical and a phonetic component (the latter of which shares its pronunciation with the character to some extent). However, this kind of phonological awareness is missing for non-logographic language users.

Secondly, although modern Chinese has a large number of pictophonetic characters, those characters' numerous phonetic components do not seem to make them easy to pronounce. The number of pictophonetic characters increased throughout the very long evolution of Chinese characters (which as far as we know spans over 3000 years) until this type of character eventually came to dominate the contemporary Chinese writing system. Li and Kang (1993) conducted a census of the number of pictographic characters based on the *Table of General Standard Chinese Characters*.[5] The results suggest that a high proportion (namely 80.5% or 5631 out of 7000) of the characters on the list are pictographic. This rate is 65.76% in the first 2500 most commonly used characters (Zhang, 2007) and 58.3% in the 1000 most frequently used characters in the Modern Chinese Frequency Dictionary (1986) (Pan, 2004). However, to obtain a character's phonological code, learners must first recognise the

[5] *The Table of General Standard Chinese Characters* (通用规范汉字表; Tōngyòng Guīfàn Hànzì Biǎo) is a standard list of 8105 simplified (and unchanged) Chinese characters. The list was issued in late 2013 by the State Council of the People's Republic of China.

phonetic component and then be able to retrieve its sound. As such, learners need to know many of the 1325 phonetic components (Li & Kang, 1993) before they can predict the sound of a multi-component character; however, it is difficult for learners at the beginner's level.

In addition, the evolution of the Chinese language and the Chinese government's continuous efforts related to language standardisation in the 20th century (which have included simplifying the characters and promoting Putonghua, the official spoken language used in Mainland China) has drastically impacted the phonological consistency between a character and its phonetic component. Phonological consistency ranges from 57.83% (Zhang, 2007) to 66% (Li & Kang, 1993). Zhang (2007) has also reported that only 19.60% (490) of the 2500 most commonly used characters completely share the same pronunciation with their phonetic components. It is therefore generally believed that providing phonetic information for words is necessary for L2 Mandarin Chinese learners to acquire vocabulary.

## 3.4 Summary of the literature review

Chapters 2 and 3 have reviewed research on incidental vocabulary learning. Chapter 2 provided an overall review of the background, development, and main gaps in this field. Chapter 3 then analysed previous studies from the perspective of evaluating incidental vocabulary acquisition from the types of word knowledge and related partial. It also pointed out the challenges posed by the special features of L2 Mandarin Chinese, in terms of testing different types of word knowledge and marking partial knowledge of the Pinyin and character forms.

In general, it can clearly be seen from the literature review that research into L2 incidental vocabulary acquisition has increased drastically in the past 20 years and that many efforts have been made to compare the effects of different types of annotations on incidental word gain through reading. Researchers in the field of incidental vocabulary acquisition through reading have made many achievements, but many

questions still remain unanswered.

The main conclusions that can be drawn from these two chapters are as follows:

1. The target language of previous studies has mainly been English. Research on other L2s (e.g. Spanish, French, German, and Japanese) is less common, and studies on L2 Chinese Mandarin learning can rarely be found in this field.

2. It is not just the meaning of words that can be learned incidentally from reading. However, until recently empirical studies have provided little evidence of learning other types of word knowledge in this manner. Most researchers seem to admit that the target of incidental vocabulary learning is only to acquire receptive knowledge of word form, or more precisely, knowledge of word form and links between word form and meaning. In addition, related research mainly focuses on the written form (i.e. spelling) of target words; phonetic knowledge is not a priority.

3. Text modification/enhancement, which usually refers to glosses or annotations, is often adopted by researchers to improve the efficiency of incidental vocabulary learning. In the context of CALL, the effect of text + picture annotation (among a limited number of annotation types) achieves the best results, possibly due to the two channels involved when processing information of words. It is noted that many researchers have adopted audio annotation in the CALL context, although they rarely address its effect in their analyses.

4. To evaluate the effect of incidental vocabulary learning, it is important to understand what and how word knowledge is tested and marked. Most tests adopt a basic format, such as translation or multiple-choice questions. The marking criteria are sometimes sensitive to partial knowledge of word meaning but not to other types of partial knowledge (e.g. productive knowledge of word form). These issues have greatly influenced the evaluation of results of incidental vocabulary research.

5. Several factors seem to affect incidental vocabulary learning through reading, including frequency of encountering unknown words, tasks and exercises after reading, and issues related to word features (e.g. part of speech and conceptual difficulty).

However, it appears that none of these factors have been fully investigated and that no universal agreement has been achieved concerning their effects on incidental vocabulary acquisition. Moreover, it is highly possible that many factors are yet to be discovered. Taking Mandarin Chinese as an example, no factors relating to this language's features have been included in previous studies. Factors suggested by censuses of this language and research on Chinese character recognition could be worth considering (e.g. a word's number of characters, a character's number of strokes, components, complexity of strokes, and the effect of phonetic component on Chinese character recognition).

The literature review has provided an understanding of research related to L2 incidental vocabulary learning and questions that remain unanswered in this field. Against this backdrop, the present study is motivated by the incidental vocabulary learning hypothesis and related studies. It attempts to conduct research on L2 Mandarin Chinese, with a special focus on the effect of annotations that provide sound and sound-related information for words. The study's research questions and some major methodological concerns are presented in the next chapter.

# Chapter 4: The present study

The primary objectives of this chapter are to present an overview of this study, how it was conducted and major concerns on methodology issues. The research questions are outlined first, with each question followed by a brief explanation of why it was chosen and how it was investigated. Thereafter, some general information on issues relating to methodology is provided. Both experiments of this study use the incidental research approach and reflect several features of that methodology (including in relation to the design of annotated reading materials, vocabulary posttests, partial knowledge-sensitive criteria, and the questionnaire). Major concerns regarding design issues are also outlined in this chapter, together with the main differences caused by adopting an online testing program in the second experiment. However, further details regarding methodology are provided separately in Chapters 5 and 6, where the study's two experiments are presented.

## 4.1 Research questions

Given the lack of incidental vocabulary acquisition research on L2 Mandarin Chinese and the effect of sound and related information, this study attempts to investigate the effect of annotation with sound-related information in the context of reading in L2 Mandarin Chinese using two types of sound-related annotation (namely Pinyin and audio). To this end, the following three research questions and have been developed:

**Research question 1:** What effects does providing annotation with information related to word sound have on incidental vocabulary learning in L2 Mandarin Chinese?

This research question consists of two related sub-questions:

a) Do learners learn sound-related knowledge for previously unknown Chinese words as well as word meaning and character form when reading an annotated article?

b) How much knowledge do learners gain incidentally with the help of Pinyin

annotations under the partial knowledge-sensitive criteria?

**Research question 2:** Do different types of annotations (i.e. text + Pinyin, text + audio, text + Pinyin + audio) affect the results of incidental vocabulary learning?

**Research question 3:** What are learners' attitudes to different types of information (namely meaning, audio, and Pinyin) provided in an annotation?

Two experiments were conducted to answer these research questions: the first one, which involved a pen-and-paper environment, compared the effects of text-only and text + Pinyin annotations; the second one, which used an online environment, compared text + Pinyin, text + audio, and text + Pinyin + audio annotations. Both experiments adopted the incidental design by keeping the vocabulary posttest a secret from participants (who were at a beginner's level in L2 Mandarin Chinese learning). Each experiment involved asking participants to read article(s) with different types of annotations and then complete a reading comprehension exercise; thereafter, they were given an unexpected vocabulary posttest. At the end of each experiment, participants were asked to complete a questionnaire to collect information on items such as language background and attitudes towards different annotations.

The primary objective of the first experiment was to answer the first research question and identify the possibility of acquiring knowledge related to word sound incidentally through reading in L2 Mandarin Chinese, as this possibility has not been identified in previous incidental vocabulary learning research. An additional objective was to determine how much knowledge gain can be expected if such a possibility does exist. Sound-related information was provided by the Pinyin annotation in this experiment. A total of 25 participants were involved, broken into a control group (which had no access to the Pinyin annotation for the target words) and a treatment group (which had access to the Pinyin annotation). The two groups' scores on the vocabulary posttest were compared quantitatively to measure the differences in knowledge gain.

Some assumptions were made in relation to the first experiment. Firstly,

participants in the treatment group were expected to gain some sound-related knowledge concerning the target words, along with knowledge related to word meaning and character form. Secondly, although not expected to be high, knowledge gain in the Pinyin form in this experiment was anticipated to be both receptive and productive for the treatment group (which had access to the Pinyin annotation). Moreover, the treatment group was expected to gain more knowledge concerning word character form than the control group, due to participants' dependency on sound-related information in the treatment group to gain other types of word knowledge (as mentioned in Section 3.3.2). Finally, on a related note, it was anticipated that the treatment group may gain more knowledge on the Pinyin form than on the character form.

It should also be highlighted that the vocabulary tests should be marked using criteria sensitive to word partial knowledge, to meet the requirements related to the incremental nature of incidental vocabulary acquisition through reading. If not, the understanding of the amount of word knowledge gained may be affected. However, apart from word meaning, criteria sensitive to partial knowledge of other types of word knowledge (e.g. written word and spoken forms) were not found in previous studies. To prove the usefulness of such criteria, the results of the vocabulary tests marked using such criteria were therefore compared to those marked using criteria not sensitive to partial knowledge.

In contrast to the first experiment which explored the effect of Pinyin annotation, the second experiment then included audio annotation in a researcher-developed online reading program to answer the second research question. It involved 48 participants reading three articles with three types of annotations, namely text + Pinyin annotation, text + audio annotation, and text + Pinyin + audio annotation. Referring back to Mayer's CMTL (as mentioned in Section 2.2.3.5), it was assumed that the text + Pinyin + audio annotation (with sound-related word knowledge provided in both visual and audio modalities) would assist incidental vocabulary acquisition more

than the text + Pinyin and the text + audio annotations involved in this experiment.

To support the audio annotation used the research, the researcher first created an online reading program. This program hosted the reading comprehension exercise and various types of annotations, in addition to vocabulary tests and a questionnaire. It also featured an embedded tracking tool that recorded participants' clicks on each annotation in the text. A brief introduction to the design of this online program is presented in the last section of this chapter to assist other researchers. With three articles and 15 target words, the second experiment employed a "crossover research design" to counterbalance article order effects, combinations of words and varying types of word knowledge, and learner and word differences. A crossover research design originally refers to a longitudinal study in which subjects engage in a sequence of interventions such that each one participates in all conditions, as adopted by Fisher et al. (2012). However, as discussed in Section 2.2.3.3, this design could have revealed the existence of a vocabulary posttest and thus harmed the study's incidental nature; as such it had to be modified. More details of the design are provided in Section 6.1.

It should be noted that the design of the second experiment is different from the design of the first experiment in many ways. The major modifications include reducing the frequency of encountering target words (to decrease the frequency effect) and changing the types of vocabulary tests used. Details concerning vocabulary test selection are provided later in this chapter, whereas the reasons for changing the test are presented in Chapter 5 (where findings from the first experiment are further discussed).

In contrast, the third research question was designed to determine participants' attitudes towards different types of annotations in reading (especially audio and Pinyin annotations, as they both provide knowledge related to word sound – which have rarely been researched in this field). Data related to attitude were collected through a questionnaire attached to the reading exercise in both experiments. This questionnaire asked for participants' attitudes towards the usefulness of different types of

information (in terms of helping them to remember unknown words encountered while reading) using a five-point Likert scale.

## 4.2 Major methodological concerns

The issues presented in this section are mainly relevant to the instruments employed in this research. They relate inter alia to the design of the reading comprehension exercises; the selection of target words, vocabulary tests, and partial knowledge-sensitive criteria; and the online reading program used in the second experiment.

### 4.2.1 Materials for the reading comprehension exercise

Considering the beginner's level of the participants involved in both of this study's experiments, it is impossible to use authentic reading materials. This is because mastery of the most frequently used 2500[6] characters is required to understand 97.97% of authentic reading materials written in Chinese (Su, 1992). As the participants also did not seem to know at least 500 to 600 Chinese characters (which was their expected level), appropriate reading material had to be developed.

To this end, five articles were selected from the book *Elementary Chinese Reading and Writing Course (Two)* (Liu, 2007), which was designed for learners who comprehend approximately 800 to 1000 Chinese characters. The articles all have different genres, but their topics relate to daily life and can be easily understood. All of the articles were then modified firstly by eliminating potentially unknown grammar

[6] *"现代汉语常用字表 （xiàndài hànyǔ chángyòngzì biǎo）[List of Frequently Used Characters in Modern Chinese]" was* released in 1988 by the National Commission on Language and Script Work and the State Education Commission of the People's Republic of China. This list contains the most frequent characters in the first band (2500 characters) and second band (1000 characters), which cover 99.48% of Chinese written materials.

points. Any sentence containing a grammar point not covered in the participants' textbook was removed or replaced by a more easily understandable sentence. Only one article was used in the first experiment while three were utilised in the second, mainly due to time constraints when the experiments were conducted.

Three additional issues need to be addressed in relation to designing the reading materials, namely the length of the article, the ratio of unknown words, and the target words' frequency of occurrence (see also Section 2.2.3.5 for more details concerning for the last two issues). These issues are further discussed below.

The first issue relates to article length. A decision was taken to cut the length of the articles to around 220 characters, based on a comprehensive consideration of the length of reading materials used both in the tests for the HSK Level 3 (which is the level this study's participants should achieve according to their teachers and the course syllabus) and the students' textbook. The reading comprehension section of the HSK Level 3 utilises only very short passages of approximately 100 characters. However, the participants' textbook indicated that they were able to and familiar with reading longer materials. For example, in the textbook *Colloquial Chinese*[7] (which the participants from SOAS[8] used), text length increases from 152 characters in the first reading text (which appears in the Lesson 3) to 315 characters in the last text (which is found in Lesson 14) and 66.67% (or eight out of 12) of the texts were not longer than 224 words. In addition, the first experiment was conducted during a revision lesson in the classroom and only about 15 minutes could be assigned to the study's exercise. The class teacher indicated that an article of approximately 220 characters should be the right length and that the participants should be able to complete the reading

[7] *Colloquial Chinese* is used for both Chinese major and non-Chinese major students at SOAS. It contains 20 lessons, each of which includes a list of required characters.

[8] School of Oriental and African Studies, this is also the university that the researcher is affiliated with.

comprehension exercises, vocabulary tests, and questionnaire within the given time. This length was also used in the second experiment.

The second reading material issue concerns the ratio of unknown words. According to the input hypothesis, comprehensive input is crucial for learners to acquire unknown words incidentally through reading. To ensure adequate comprehension, Laufer (1992) and Nation (2001) assert that the ratio of unknown words should be no more than 5% (as mentioned in Section 2.2.3.6). One may argue that comprehensive input needs to be addressed in natural reading, as it is important for learners to guess unknown words and consequently acquire them through reading. However, as providing annotations in reading materials eliminates the need for learners to guess the meaning of unknown words, this ratio becomes meaningless in such research. It might be reasonable to believe that annotation makes it possible to involved more unknown words in reading material. However, an upper limit should exist, as the appearance of too many target words could interrupt the reading process and in an extreme case make an article difficult to understand. As researchers in this field have not provided an updated ratio, this study adopts the 5% ratio mentioned above in designing its reading material.

It should be noted that in texts written in Chinese, no space is included between words and all characters are arranged in exactly the same way in terms of format. To compare the words in the articles with the vocabulary lists mentioned above, the first step was therefore to identify all of the words in each article. This was done using the software “Institute of Computing Technology, Chinese Lexical Analysis System” (ICTCLAS), which is an automatic segmentation tool that segments sequential text into character strings based on word units. It was created by the Chinese Academy of Science’s Institute of Computing Technology, which is the leading computing technology institute in China. The ICTCLAS currently supports many different standards for word segmentation, including those of the Chinese State Language Commission, the Taiwan Academia Sinica, and Beijing University.

The third issue related to the design of the study's reading materials concerns the frequency of the target words. The first experiment followed the suggestion that the critical point for acquiring words incidentally through reading is approximately eight times (see Section 2.2.3.5). In that experiment, the target words appeared at different times, but always fewer than eight times in an article. However, the results suggested that some of the target words were better acquired than others – and one reason could be the additional times they appeared (see Section 5.3 for more discussion). As such, the frequency of encountering each target words was reduced to one or two times in the second experiment.

In addition to the steps described above, the modified articles were also given to some experienced teachers (including those of the participants) to see whether the topics, vocabulary, and sentence structures were appropriate for the participants.

### 4.2.2 Selection of target words

Empirical studies previously conducted within the field of incidental vocabulary acquisition do not offer any clear guidance for selecting target words in the context of L2 Mandarin Chinese. The few studies that mention factors related to word features (as discussed in Section 2.2.3.5) and features of Mandarin Chinese words and characters (as presented in Section 3.3.2) were indeed all taken into consideration. However, it was difficult to include all these factors into a single study. As such, this research attempts to start from controlling the word length by focusing on the two-character words used in this study. The main reason for choosing the two-character words is that this type of word is dominant in modern Mandarin Chinese, as explained in Section 3.3.2.

As this study focused on sound-related knowledge of a word, the phonetic components were also considered when selecting the target words. In both experiments, three types of target words are used: 1) words with a phonetic component in their first character, 2) words with a phonetic component in their second

character, 3) words with a phonetic component in both of their characters. One loan word with a pronunciation that is very similar in Mandarin Chinese and English was also included.

In relation to the part of speech, since the first experiment involved just five target words, only nouns and verbs were selected. In the second experiment, the 15 target words included nouns, verbs, adjectives, and adverbs.

When selecting the target words, the most important criterion was that they should be unknown to the participants. All the words that appeared in the articles were therefore checked against the vocabulary lists in the participants' textbooks. In addition, considering the existence of single-character words (as mentioned in Section 3.3.2), each of the target words' characters were also checked against the vocabulary lists to ensure participants were not familiar with them. Confirmation was also sought from the participants' teachers that these words would not normally be known by the learners. Lastly, tests or questions concerning pre-knowledge of these words were also involved in both experiments as a final control (more details can be found in the methods sections in Chapters 5 and 6).

### 4.2.3 Testing different types of word knowledge

As extensively discussed in Section 3.2, the design of a vocabulary test is closely related to the types of word knowledge being tested and may influence researchers' evaluations of the effect of incidental vocabulary acquisition. This section thus discusses three essential issues considered when this study was designed: 1) the types of word knowledge that needed to be tested; 2) the options for vocabulary tests, and 3) the advantages and disadvantage of each test option.

The first issue concerns the types of word knowledge tested in this study, as many options existed. Based on earlier discussions about the word knowledge framework (Section 2.1.4) and the special features of learning words in Mandarin Chinese (Section 3.3), the word knowledge pertaining to form and meaning that can be tested in L2

Mandarin Chinese learning includes spoken, character forms, and Pinyin forms; links between each form; and links between each form and meaning. Considering the possible reliance on sound-related information in learning character-related word knowledge mentioned in Section 3.3, this study also included word knowledge related to the character form. It would be ideal to test all of the types of word knowledge listed above; however, it is not practical to involve all of them in a single study. As such, efficiency (in terms of ability to test multiple types of word knowledge) became a very important consideration when choosing the vocabulary tests for this study.

The second issue pertains to vocabulary test options. Vocabulary tests continue to be routinely used in second language teaching with a variety of assessment purposes and are quite simple in some respects. It is only a matter of selecting a suitable number of target words and assessing them through an established test format, such as yes/no questions,[9] multiple-choice questions, gap-filling, and translation (Read, 2000; Liang, 2005). Previous studies have also employed the VKS (e.g. Qian, 2003; Zhu, 2004), matching tests (Chun & Plass, 1996), definition-supply questions (in L2) (Al-Seghayer, 2001), association tests, and cloze test (Yeh & Wang, 2003) (see Section 3.2 for more details). As this study focuses on word form and meaning, association, cloze, gap-filling, and VKS tests (which all require other types of word knowledge to complete) and definition-supply questions in L2 (which require other L2 words to be produced) are excluded from further discussion. Moreover, only translation-type questions (including L2 meaning to L1 Pinyin or character, character to Pinyin, and Pinyin to character) are suitable for testing production of the Pinyin and character forms. As such, the issue was narrowed down to determining the appropriate test for recognising the three forms of

[9] Meara and his colleagues, first in London (Meara & Buxton, 1987) and then at Swansea University (Meara, 1992; Meara & Milton, 2005), have taken the lead in developing yes/no tests for L2 learners and making them available for practical use as placement tests or as general measures of vocabulary size or competence in a language. However, this type of question has not been used in tests related to incidental vocabulary learning. The closest could be the word-form recognition test in Waring and Takaki (2003).

a word and links between forms or between form and meaning (the spoken form is only included in the second experiment, which provides audio annotation for the target word). The remaining choices were yes/no questions, translation from L1 to L2 and matching test.

Turning to the advantages and disadvantage of each test option, it is noted that the word-form recognition test adopted by Waring and Takaki (2003) is the only test identified in this field that focuses on word-form recognition, although it does not include a word's form-meaning link. It measures the same type of word knowledge as the first level of the VKS ("I have seen this word before") and the yes/no questions often used as placement tests or general measures of vocabulary size or competence (Anderson & Freebody, 1983; Meara & Buxton, 1987; Meara & Milton, 2005). The word-form recognition test should be valued for incidental vocabulary learning, as it measures a type of incremental word knowledge often neglected in this field. However, it has the weakness of overestimating people's vocabulary knowledge, as results may possibly be affected by guessing – even when non-words[10] are added. In addition, it only tests one aspect of word knowledge, which does not seem to be very efficient compared to other types of test, such as translation from L2 to L1, which tests both a word's form and the link between form and meaning.

Among all of the types of tests listed above, multiple-choice and translation tests have achieved the widest acceptance. In previous research on incidental vocabulary acquisition, multiple-choice questions have often been presented by providing a prompt word together with four or five choices. This prompt word can be delivered in either L1 or L2, as can the choice items. However, in translation tests, an equivalent (or definition) in L1 or L2 is required for an L2 word. Considering the related discussion in Section 3.2 about multiple-choice tests being affected by guessing and serving as

[10] Anderson and Freebody (1983) introduced an important innovation by including a substantial proportion of non-words among the items to reduce the impact of guessing.

better language learning tools than testing tools, the translation test (L2 to L1) becomes a more reliable choice for this study. Nonetheless, the weakness of the translation test is comparatively difficult. Without choice options, it becomes difficult for participants to produce a correct answer if they have only acquired partial knowledge of word meaning. In addition, considering that Mandarin Chinese is the target language, both the character and Pinyin forms need to be tested; as such, ten questions are required for five target words – which does not seem very efficient. More importantly, it was difficult to test only the link between the character and Pinyin forms, without asking participants to produce one of the forms.

Considering the efficiency of vocabulary tests, this research thus followed Chun and Plass's (1996) design by employing a matching test as the recognition test in the first experiment. This option was chosen as it made it possible to test multiple types of word knowledge at one time, including knowledge of Pinyin and character forms (recognition only), the link between Pinyin and character forms, the link between Pinyin form and word meaning, and the link between character form and word meaning. It should be noted that instead of matching meaning and written form only (the format usually employed when English or other alphabetic languages are tested), Pinyin was also added to the test. The result was that participants had to match the Pinyin, word meaning in English, and character form of the target words. However, as revealed by the first experiment, some problems were experienced with this test when the L2 became Mandarin Chinese; a detailed discussion and the format of the matching test are presented in Chapter 5. The recognition test was consequently changed to a translation test in the second experiment, which meant that another often discussed factor that may impact language testing results, namely context, was also taken into consideration.

The long-standing topic of research and development in second language vocabulary assessment is closely associated with reading comprehension or communication ability. According to this viewpoint, vocabulary should be tested in

certain contexts rather than through decontextualised tests. This point of view is clearly favoured and reflected in some of the comprehensive English language tests most widely taken by second language users, including the Test of English as a Foreign Language (TOEFL) and Graduate Record Examination (GRE), and in vocabulary size tests such as the new Vocabulary Size Test (Nation & Gu, 2007). However, previous studies related to incidental vocabulary acquisition, though not many, present another opinion. For example, Waring and Takaki (2003) mentioned that their subjects might have been able to guess the meaning of words if context had been used. In the present study, for example, the translation question can be provided in a richer context. An example is given below:

Translate the underscored word.

请给我一双耐克。 _____

(Qǐng gěi wǒ yìshuāng Nàikè。) [Please give me a pair of Nike (shoes).]

The giving sentence may provide other clues could help the test-taker to answer the question. For example, the word "一双 (yìshuāng) [a pair of, measure word, often used with hand, foot, shoes in Chinese]" might already suggest several possible answers and help test-takers to recall the meaning of the word when encountered in reading the text. As such, instead of recognising the character forms of the target words, the real source of right answers could be the context of the sentence or knowledge of collocation between the word "一双 (yìshuāng) [a pair of, measure word]" and the target word. If the sentence is taken from the reading material, the test may become easier. All of the tests in this study were therefore provided without any context.

Another point that needs to be addressed is the difference between tests in pen-and-paper and online environments (even with the same format). A meaning-based character form production test can be used as an example. In a pen-and-paper environment, the word knowledge being tested is character form (in handwriting) and the link between this form and word meaning. In the online

environment, the word knowledge being tested can be kept the same as in a paper-and-pen test if a handwriting board is used; however, different word knowledge will be tested if Pinyin input is employed. In the second circumstance, the word knowledge tested becomes the link between word meaning and Pinyin form, Pinyin form (production), the link between Pinyin and character forms, and character form (recognition). Considering the feature of learning in CALL, the second experiment employed Pinyin input in the character production test. For clarity, the test was subsequently referred to as a Pinyin input assisted meaning-based character form production test. This choice was made for various reasons. First (and most importantly), as discussed in Section 2.2.3.3, a computer-assisted learning program should not just recreate what can be done in a pen-and-paper environment; its features should be utilised to assist learning (Zheng, 2014). Second (and on a more practical level), Pinyin input has been widely adopted by native Mandarin Chinese speakers and using it only requires software to be installed. In contrast, handwriting boards are not very commonly used in either language learning or daily life and require extra equipment to be purchased and prepared for each computer used in an experiment.

Finally, a question about how to test participants' pre-knowledge of the target words also exists. In previous studies, this knowledge was tested with a pretest before participants read articles (e.g., Yanguas, 2009; Qian, 2003) or assessed in conjunction with the posttest after the reading exercise (e.g. Chun & Plass, 1996). The major concern about testing before reading is that it might attract participants' attention to these unknown words and thus hurt the incidental nature of learning. In addition, extra encounters with the target words may affect participants' learning results, for example in relation to the forms of the words, although they may not know the meanings during the pretest. As such, the first experiment tested pre-knowledge of the target words after the participants completed all of the reading comprehension exercises and the vocabulary posttest. However, as the results of the first experiment suggested problems with this design, a proper pretest had to be employed in the second

experiment (for more details see Section 5.3).

In the same vein, the production and recognition tests used in this study had to be given in a strict order. The production test, which involved L1 to L2 translation, was given first to reduce extra exposure to the target words in the recognition test.

### 4.2.4 Measuring partial knowledge gain

Given the incremental nature of incidental vocabulary learning through reading and the partial knowledge of different types of word knowledge that learners can acquire in such process, it becomes very important to use appropriate criteria that are sensitive to partial knowledge in measuring knowledge gained incidentally through reading. As mentioned in Section 3.2, only a few researchers have employed such criteria (e.g. Nagy et al., 1985; Qian, 2003; Granick, 1997; Y. Zhu, 2004; Wu & Xu, 2009). Generally speaking, they have adopted different marking scales, including three-point scales (0, 1, 2 or 0, 0.5, 1) and five-point scales (0, 0.5, 1, 1.5, 2). However, many other studies have simply employed criteria that are not sensitive to word partial knowledge, in which only completely correct answers to questions are awarded points (with partially correct answers not receiving any points). However, it is also noted that all partial knowledge-sensitive criteria adopted in previous studies have focused on measuring word meaning – which means the more accurate the meaning given for a tested word, the higher the score. In this study, a 0.5-point scale can be easily adopted by following previous researchers. Participants were awarded one point for a fully correct answer concerning word meaning and 0.5 for a partially correct answer. For example, the answers “silk robe” for “丝绸 (sīchóu) [silk]” and “training” for “培养 (péiyǎng) [cultivate]” were both awarded 0.5 points. It needs to be noted that the same criteria were not applicable in the first experiment, as the matching test did not require learners to produce word meanings.

However, as the production of word form has been ignored in this field, no such criteria can be found for marking word form in related L2 English literature – not to

mention for measuring the forms of a Mandarin Chinese word. As this study tested both Pinyin and character forms, the researcher developed related criteria taking the features of both the Pinyin and character forms of Mandarin Chinese words into account (see Section 3.3.2 for more details).

With respect to the production of the Pinyin form of the target words, partial knowledge can be assessed in two ways: by determining whether each letter and the tone mark were produced correctly or by determining whether each part of the Pinyin syllable was correctly produced (i.e. the initial part, the final part, and the tone part). With consultant to the participants' answers to related tests, the second standard was adopted in this study, as participants rarely produced correct letters. In the tests that required the Pinyin of words to be produced, one point was awarded for a fully correct answer (correct initial part, final part, and tone in either character of a two-character target word) and partial points were awarded for partially correct answers (e.g. 1/6 of a point was given for the correct initial part or tone for one character).

With respect to the handwritten production of Chinese character forms, knowledge of strokes and components is more likely to add value to the partial knowledge of a Chinese word's character form. Considering the wide variation in the number of strokes in one character (see Section 3.3.2) and, more importantly, that the participants provided almost no discrete strokes, it was more practical to regard a character's components as the basic unit for marking partial knowledge. Points were thus awarded according to the number of components participants successfully write.

Components of a character were separated according to the GB.13000.1 Character Set for Information Processing-Specification On Chinese Character Components (The National Language Committee, 1997)[11] (see Appendix D for the

[11] GB standards are the Chinese national standards issued by the Standardization Administration of China (SAC). GB stands for "国标"（Pinyin: guó biāo; Meaning: national standard）. This character set contains 393 sets of components including 560 basic components to form Chinese characters.

components of each target word). The component appeared repeatedly in a single word were repeatedly counted. Each right component written correctly was awarded one point. However, participants could also be awarded 0.5 points if they only made some small mistakes in the components (e.g. missed or added one or two strokes or mixed similar components). In addition, if the components were put in the wrong position, for instance if the character "絲（sī）[silk]" was written as "糸幺 " with each component correctly written but in the wrong position, 0.5 was deducted from the participant's total score. It needs to be borne in mind that these criteria only applied to the character production test in the first experiment, in which participants could produce incomplete characters (partial characters could not be produced in the online production vocabulary test, which used Pinyin input). Full details of the target words' components can be seen in Appendix D.

### 4.2.5 Collecting data through the questionnaire

The questionnaire, which was delivered to participants as the last part of both experiment, aimed at collecting demographic information (e.g. age, contact email, L1, and other language levels) in addition to information about learning style (i.e. visual, aural, kinaesthetic, or mixed) and how participants completed the online exercise. As part of the formative evaluation of the experiment, participants were also asked to provide comments and feedback on the whole reading exercise. The complete questionnaire can be found in Appendix C.

The most important issue that needs to be addressed here is the participants' language background. Learners with knowledge of Japanese Kanji or Korean Hanja find Mandarin Chinese easier than learners who only know Western languages. In addition, as reported by Qian (2003) these learners may rely less or not at all on phonetic information when learning new words (see Section 3.3.1). It is therefore important to separate these learners from the others. In order to collect related information, the questionnaire included a question that asked about the participant's language learning

experience. The answers were mainly used for selecting samples (i.e. to help decide whether a participant should be excluded from the study).

Another major objective of the questionnaire was to collect data on participants' attitudes towards different types of annotation – that is, to determine how useful they think certain types of annotation would be in terms of helping them learn words while reading. A five-scale Likert question was included in the questionnaire to this end. Participants were required to mark the degree of usefulness of the information provided in the annotation using a five-point scale in which 1 = not useful and 5 = very useful.

It should be noted that five types of annotation were involved in the attitude question in the first experiment, namely meaning of words, Pinyin, audio, and emphases phonetic component or meaning component of a word by marking them, as mentioned in the previous chapter. However, the last three did not appear in the first experiment. Participants' attitudes about audio annotation were solicited to enable comparison with the second experiment. As for the last two included in the question, they appeared because they were initially going to part of the second experiment. However, as participants had low interest and involving too many types of annotation in the second experiment created difficulties, these two types were not included in the annotations and consequently removed from the attitude question in the second experiment.

### 4.2.6 Data analysis procedure

The last issue that needs to be addressed in this chapter relates to data collection and the selection of statistical tests for the study. Quantitative analyses were utilised in both the first and second experiments. At the data coding stage, the participants' answers to the vocabulary tests were transformed to scores for later analysis using the statistical software SPSS 21. To increase reliability, the vocabulary tests were also scored by a second scorer using criteria sensitive to partial knowledge.

The scores of the vocabulary tests were compared across annotation type using either the ANOVA test or its non-parametric counterpart test (if the data were not normally distributed and the groups had heterogeneous variances). The distributions of the samples were tested utilising both the Shapiro-Wilk and Kolmogorov-Smirnov tests; the results of the former were used to determine the distribution if conflicts were found between the two tests, due to its higher accuracy according to Field (2013). The homogeneity of variance was tested via Levene's test. Details of the statistical tests chosen for each experiment are presented in Chapters 5 and 6.

One thing about the statistical analysis is worthy of note. In many cases, the study required non-parametric analysis. The logic behind non-parametric tests is incredibly elegant. The lowest score is ranked as 1, within higher scores being assigned higher ranks. With the dependent variable measured at the interval level, the analysis is then carried out utilising the ranks rather than the real data. This process is an ingenious way of breaking parametric assumptions (Field, 2013) and enabled a more technical method to be used when analysing this study's non-parametric data. However, as they neglect many information-related details, non-parametric tests are sometimes believed to be less powerful than their parametric counterparts and more likely to cause a "type II error" (which entails falsely accepting the null hypothesis). In this study, such an error would imply that no differences exist between the effects of different types of annotation in terms of acquiring varying types of word knowledge. As a result, scores for each part of the vocabulary posttest had to be provided in an in-depth analysis to help understand the results of the statistical tests.

Before moving to the details of the first experiment, the final issue that needs to be addressed is related to data triangulation to ensure the reliability and validity of this research. The first and the third research questions were answered by both experiments in this study. With two experiments conducted with different participants and carried out at a different time, it is safe to say that the findings on sound-related word knowledge being learned incidentally in reading, as well as that of participants'

attitude to various types of annotation, are reliable. As regards the second research question, great care was taken of methodological issues relating to designing the online reading program. For example, given that word frequency and noticing of the word are essential to incidental vocabulary learning, a tracking tool was embedded in the program and the log file data was considered to determine relevant questions. In addition, relating to the frequency issue discussed in Section 2.2.3.5, participants were asked in the questionnaire about the time they looked at the target words. In this case, data from both the questionnaire and the log file was employed to ensure accuracy. Beyond that, apart from consulting participants' teachers on target word selection, the reading material was, word by word, analysed against students' textbook to determine the target words.

### 4.2.7 Issues related to the design of the online program

As audio annotation was added to the reading materials, the whole reading and testing program had to be presented in an online environment (which can easily host such annotation). This section identifies three main concerns about developing such an online program for the second experiment: embedding a tracking tool within the program, difficulties in delivering various types of tests and questions, and embedding the Google Pinyin input within the program. Thereafter the software and platforms used to design the online program are mentioned.

The first issue concerned embedding a tracking tool within the online program developed for this study, which was done in accordance with previous studies (Chun & Plass, 1996; Bowles, 2004; Yanguas, 2009). This tracking tool produced a log file that registered every click participants made when completing the online reading program. The data revealed much interesting information, such as participants' IP addresses, the time they started the reading exercise, how long they spent reading each page, and which annotation they clicked. However, the main catalyst for employing the tracking tool was the discussion on the noticing issue in Section 3.2, including the studies of

Qian (2003) and Bowles (2004), and the discussion relating to the frequency issue in Section 2.2.3.5. It is crucial to ensure that participants read the annotation and did not ignore the target words and to determine how many times they actually looked at each annotation.

The tracking tool was selected among five statistics service providers, namely Google Analytics, Extreme Tracking, Piwik, and two Chinese companies (Baidu Analytics and CNZZ) using the following criteria: 1) easily accessible online; 2) quick response, which means providing the tracking data instantly or within a short time; 3) adding tracking code to the website automatically, which is required mainly to reduce the designer's workload; 4) easily readable tracking data; 5) providing detailed tracking data (e.g. IP address and accessing time); and 6) availability of a free service. No provider met all six criteria, but Baidu Analytics was chosen as the main tracking tool mainly due to the third and fourth criteria, while the CNZZ and Extreme Tracking were also included for additional information.

The second concern related to the difficulty of presenting all types of tests and questions in the second experiment. The list included the vocabulary test (which required extra spaces for detailed information to be given and using audio clips as prompts), reading comprehension questions (which entailed both multiple-choice and open questions), and a five-point Likert question to record participants' attitudes towards various types of annotation. For example, in the VKS test involved in the second experiment's pretest, if a participant indicated that he/she partially knew a word, a space was required for him/her to provide details of the known parts, However, in many online systems (e.g. the Google Drive), such space cannot be added to questions. The second experiment also adopted an audio recognition test that required participants to produce word meaning according to an audio clip, which was challenging for many online testing and survey systems. Considering the large range of questions types used in this study and the need for customisable system, the China-based online survey system Sojump.com was employed.

The third concern relates to the Pinyin input assisted meaning-based character form production test. The Pinyin input required to complete this test could be achieved using an input that is either installed in a computer directly or embedded online. The second option was adopted, and the Google Pinyin Input was embedded in the study's online program. This option was selected mainly due to the fact that different input methods may provide different options for the word needed (or the same options in a different order) according to the Pinyin typed in – which may affect participants' choices. More importantly, most input methods include a recommended function that collects users' typing information (e.g. choice of words) for later use. If the same Pinyin form was typed in, the previously chosen word could possibly appear at the top of the list of choices, which may affect the choice of later users. As the participants in this study came in small groups, the same computers were repeatedly used (more details are provided in Section 6.1, which deals with the study's procedures). To exclude this influence, Google Pinyin web input was ultimately chosen. Even if the program also collects users' choices to improve its service, it is highly unlikely that a small number of clicks will change the order of choices that appear in the recommendation bar and hence affect participants' answers.

To exclude the possible influence caused by different input tools, participants were thus instructed to use the Google Pinyin Input to complete the posttest (rather than using the default inputs installed in the computer). To this end an instruction page was added before the vocabulary test section to instruct participants to click a given link to activate the Google Pinyin Input. This page also contained information on how to switch between Chinese input (for the character production input) and English input (for the Pinyin production test), brief instructions for using Pinyin input, and a question that asked participants to type "你好, (nǐhǎo) [hello]" as a trial. Finally, the software and platform used to host the web pages were briefly introduced for the reference of future researchers.

The whole test was edited using the web design software Adobe Dreamweaver 8

and then uploaded to a specifically created website. Considering the server speed needed for opening the audio files embedded in the annotation and the charges associated with online hosting services for the program, 000webhost.com was selected as the host.

# Chapter 5: The first experiment: Pen-and-paper environment

This chapter presents the details of the study's first experiment conducted in a pen-and-paper environment, comparing the effects of the text-only and text + Pinyin annotations. As explained in the previous chapter, the primary objective of this experiment was to answer research question 1, which explores the effects of Pinyin annotation on incidental vocabulary acquisition through reading in a paper-and-pen environment. In other words, the goal is to investigate whether sound-related word knowledge can be acquired incidentally through reading in L2 Mandarin Chinese with Pinyin annotation provided, and how much of each type of word knowledge learners can obtain with the Pinyin annotation provided in this process. Participants' attitudes towards different types of annotations were also examined. Moreover, as previous studies have not investigated the effects of partial knowledge-sensitive criteria and such criteria have not been used to measure knowledge relating to word form, it was important to test whether such criteria affect the results of incidental vocabulary acquisition in this experiment.

This chapter starts with an introduction to the research methods, including the participants; instruments and procedures; and data collection, coding, and analysis. The results are presented thereafter, followed by a discussion of the questions found relating to research design. The chapter ends by outlining the experiment's implications.

## 5.1 Methods

With the previous chapter's discussion of methodological concerns in mind, this section provides detailed information on participants; instruments and procedures; and data collection, coding and analysing.

### 5.1.1 Participants

A total of 41 first-year students majoring in Chinese at SOAS took part in this experiment. Most of these students had English as their native language, and as the others had been accepted by SOAS for academic purposes, it was assumed that they had a high level of English and would not have any problems in understanding English annotations or instructions. Most of the participants had no Chinese learning experience before starting their Mandarin Chinese course. Those who met the following criteria were excluded from the final analysis: 1) they did not successfully complete all the sections of the exercise, which included reading comprehension, the vocabulary tests and the questionnaire;2) they already knew 10% or more of the target words; and 3) they had previous knowledge of a writing system that has similarities to Chinese (mainly Japanese Kanji or Korean Hanja). The participants' answers to questions related to pre-knowledge of the target words and language learning experiences in the questionnaire revealed whether they fulfilled criteria 2 and 3. Very strictly, the data reported in this thesis is based only on participants who chose the answer "I didn't know this word before reading this passage" for all the five target words and those who showed less than 10% knowledge of the target words under the partial knowledge sensitive criteria. In other words, none of the participants involved in the final analysis would have known a single character included in the target words, no matter whether he/she was in the control group or the treatment group.

Based on the above criteria, 25 participants were left in the final pool; 11 were in the control group (nine English, one German, and one Spanish) and 14 were in the treatment group (nine English and one each from Germany, the Netherlands, Russia, Poland, and Italy). Three participants were removed because they either had Japanese learning experience or lived in a Chinese family, which could offer them an advantage in learning Mandarin Chinese over other participants. Two participants were also excluded because they did not properly complete the questions on previous knowledge

of the target words and 11 samples were excluded because the participant showed more than 10% of the correct answers in the question dealing with their pre-knowledge of the target words.

The experiment was conducted towards the end of the university's second term, by which time the participants should have covered 14 lessons in their *Colloquial Chinese* textbook and 564 characters. Although the majority of Mandarin Chinese beginner's university syllabuses teach simplified characters, SOAS Chinese major students are required to master traditional Chinese characters in order to build a better foundation in later character recognition and production, as well as to deeply understand the Chinese writing system. Therefore, the participants had received instruction on traditional Chinese characters from the beginning of the course and had only begun being instructed on the simplified characters some weeks before this test. As such, they were more familiar with the traditional character form. The whole test – including the article, reading comprehension exercise and instruction – was therefore delivered in traditional Mandarin Chinese characters.

## 5.1.2 Instruments

This experiment's instruments included an annotated article with reading comprehension questions, vocabulary posttests, and a questionnaire. Detailed information on these items is presented below, apart from on the questionnaire (as the most important questions in the questionnaire were discussed in the previous chapter).

### 5.1.2.1 Annotated reading material

The annotated reading material was prepared in accordance with the discussion of reading material in Section 4.2.1 and the discussion of target word selection in Section 4.2.2. Considering the limited time allocated for the test in the participants' Mandarin Chinese class, only one article, "This silk is the payment for the pair of shoes",

was used in the pen-and-paper environment at this stage. This article was chosen mainly due to its similarity in style to articles that appear in the participants' textbook, as well as its shopping-related topic (which can be easily understood). The article involved 196 characters, including title and punctuation. The reading material used in the experiment is included in Appendix A.

At this experiment, annotations comprising each target word's character form, Pinyin, and meaning were presented as marginal glosses to the reading material for the treatment group, while the Pinyin was removed from the annotation for the control group. As shown in Table 5.1 below, the article contains five target words. The target words were selected using the criteria mentioned in Section 4.1.2; four were nouns and one was a verb. Among them, "耐克(Nàikè) [Nike]" is a loan word, "價值(jiàzhí) [value]" has a phonetic component in both of its characters, "改變(gǎibiàn) [change]" has no phonetic component, "絲綢(sīchóu) [silk]" and "判斷(pànduàn) [judge]" have a phonetic component for one character.

Table 5.1 Target words with Pinyin and English equivalent

| Target word | Pinyin | English equivalent |
|---|---|---|
| 絲綢 | (sīchóu) | [silk] |
| 價值 | (jiàzhí) | [value] |
| 改變 | (gǎibiàn) | [change] |
| 耐克 | (Nàikè) | [Nike] |
| 判斷 | (pànduàn) | [judge] |

#### 5.1.2.2 Vocabulary tests

Two types of vocabulary tests were involved in the first experiment. The L1 to L2 translation test (which served as the production test) preceded the matching test (which was employed as a recognition test). As discussed in Section 4.2.3, for the sake of efficiency and being able to assess many types of word knowledge in a single test,

this study adopted the matching test used by Chun and Plass (1996). It was modified by adding Pinyin forms of the target words to the options. Concerning types of word knowledge, the character form and Pinyin forms of the target words, the links between these forms with word meaning, and the link between character from and Pinyin form were involved. The test asked participants to group the target word's Pinyin, English meaning, and character form, as shown below:

Please group the pinyin, character form and English meaning.

| | | |
|---|---|---|
| 1. sīchóu | 6. 價 值 | 11. Nike |
| 2. value | 7. jiàzhí | 12. 判 斷 |
| 3. change | 8. nàikè | 13. 耐 克 |
| 4. gǎibiàn | 9. 絲 綢 | 14. silk |
| 5. 改 變 | 10. pànduàn | 15. judge |

（You don't need to copy the word; you can use the number.）

Group 1. _____

Group 2. _____

Group 3. _____

In contrast, the production test required participants to translate English words into both Pinyin and character forms. The types of word knowledge tested include character form, Pinyin form, and links between these two forms and word meaning. To measure partial knowledge acquired incidentally through reading, participants were encouraged to record their understanding of parts of the vocabulary even if they were unable to provide the whole word in characters or Pinyin. The instructions explained that participants could provide parts of a word, such as one of its characters, components of characters, or a part of the Pinyin.

In relation to testing participants' pre-knowledge of the target words, the strict target word selection process mentioned in Section 4.1.2 was taken into consideration, together with the goal of avoiding extra pre-exposure to the target words that was discussed in Section 3.2. As a result, the first experiment used a method adopted by some researchers (e.g. Chun & Plass, 1996; Hulstijn & Laufer, 2001; James, 2009) that

entailed asking participants to answer questions about their pre-knowledge of the target words after completing the reading comprehension and vocabulary posttest (rather than including a pretest). And such question was thus involved in the questionnaire attached to the end of this experiment. In order to explore multiple types of word knowledge (including word partial knowledge) and not just assess participants' previous knowledge of word meaning, the question was provided in the following format:

A. I did not know this word before reading this passage.

B. I know the pronunciation of this word

C. I know both characters in this word

D. I know only one character or part(s) of the character in this word

| Target word | Your choice of answer | If you choose B, C or D, complete this part. | | | |
|---|---|---|---|---|---|
| | | Part(s) which you know | Meaning of the part(s) | Pinyin of the part(s) | Other information you know about this word |
| 絲綢 | | | | | |
| 價值 | | | | | |
| 改變 | | | | | |
| 耐克 | | | | | |
| 判斷 | | | | | |

The participants had to choose the option that best described their knowledge of a word and were asked to provide detailed information if they selected B, C, or D.

### 5.1.3 Procedure

This reading test was administered in a revision class (which is a type of class that SOAS teachers conduct at the end of the second term to help students prepare for forthcoming examinations). It was used as a revision tool and delivered with the help of the teacher at the beginning of the period. The participants were randomly assigned

to the control group or the treatment group when their teacher gave them one of the two versions of the reading materials. The control group received the article with a text-only annotation that provided word meaning in English and the target word written in character form, while the treatment group had a text + Pinyin annotation that provided the Pinyin of the target words as extra information.

Students were asked to complete a reading comprehension activity in class and not told anything about the vocabulary test. They were requested to complete the exercises in order and not to jump to the next section without completing the previous one. Related notes, such as “please do not look at the original text in the previous page” and “please do not go back to this part again after you finish it” were also printed in bold next to the reading article.

In the vocabulary test, the production test came before the recognition test in order to reduce the influence of extra exposures to the target words in the recognition test. After participants completed the reading comprehension and production test, they handed this part in. The second part, which included the recognition test, was then distributed. The students were again asked to complete the exercises in a specified order. The production test was on the back of the article; however, the teacher told the students not to jump the next part before completing the first one. All of the participants had to hand the first part in before they could complete the recognition test. The students were given 15 minutes to complete the reading comprehension and vocabulary test and another 5 minutes for the questionnaire.

### 5.1.4 Data collection, coding, and analysis

This section presents details concerning data collection, coding, and analysis according to the three purposes of this experiment mentioned at the start of this chapter. These purposes are as follows: testing the partial knowledge-sensitive criteria by comparing vocabulary posttest scores using different criteria, comparing these scores between the control and treatment groups to identify the effects of the Pinyin

annotation, and comparing participants' attitudes towards different types of annotations).

In terms of achieving the first objective, scores for the Pinyin and character production tests in the first experiment were involved in the analysing procedure as participants provided detailed partial knowledge. Their answers were first marked using criteria that are sensitive to partial knowledge (see Section 4.2.4) and thereafter using criteria that are not. The one-way ANOVA was employed to test the differences between the two sets of data, after testing the distribution of the data and the homogeneity of variance. The independent variables in this statistical tests were the types of criteria and the dependent variables were the scores for the Pinyin and character production tests.

In order to test the effect of the Pinyin annotation in the first experiment, scores of the vocabulary posttests were collected separately for the control and treatment groups. Considering that Pinyin annotation may have different influences on the acquisition of various types of word knowledge, scores for each type of knowledge tested were also separated for later comparison. The independent variables were thus the annotation types (namely text-only and text + Pinyin) and the dependent variables were the scores on the vocabulary tests (including the total scores for the posttest and the scores for each type of word knowledge tested). In order to conduct further statistical tests, the distributions of the scores for each group were first analysed using the one-sample Shapiro-Wilk test, while the homogeneity of variance was assessed via Levene's test. Considering this experiment's between-subject design, the non-normally distributed data were analysed using the "Mann-Whitney U test". An in-depth analysis of the percentage of correct answers in each test was also undertaken to help to understand the statistical results. These analyses also demonstrated the amount of word knowledge (including partial knowledge) gained.

Regarding participants' attitudes towards various types of annotation, participants' answers to the five-point Likert-scale question were collected and compared across

annotation types. The distributions of the participants' marks for each type of annotation were first analysed using the one-sample Shapiro-Wilk test, while the homogeneity of variance was assessed via Levene's test. Corresponding statistical tests were decided accordingly.

Apart from the data collected through the vocabulary tests, answers from the questionnaire relating to participants' language learning backgrounds and the number of times they looked at each annotation were also analysed. This supplementary information was helpful inter alia for the selection of participants.

In relation to participants' attitudes towards different types of annotations, their responses to the five-point Likert-scale question were collected without separating the control and treatment groups. No significant differences were noted between the two groups on the various types of annotations being asked about, even the Pinyin annotation. The independent variable was therefore the type of annotation, while the dependent variables were participants' marks. As the data were non-normally distributed, a non-parametric test was required to carry out the statistical analysis. Considering that more than two types of annotations were being compared and given the within subject design (which entailed all participants indicating their attitudes towards all five types of annotations), a *Friedman's two-way analysis of variance* (related samples) with Wilcoxon signed-rank tests as the post hoc tests was determined to be appropriate.

One issue is noteworthy here. As pointed out by Field (2013), increasing the number of post hoc tests inflates the Type II error rate (which means the null hypothesis is falsely rejected). In this study, this would result in the different types of annotations having the same effects on incidental vocabulary learning. In other words, it is possible that the differences between the five types of annotations would be reported as significant where they should not be with a critical value of 0.05. The Bonferroni correction was applied to avoid this situation: instead of using 0.05 as the critical value for the significance for each test, 0.005 was utilised. This adjusted figure

was determined by dividing 0.05 by 10, the number of tests conducted in this experiment (given that all five types of annotations needed to be compared with each other in the post hoc tests).

## 5.2 Results

In line with the research design explained earlier, this section presents results related to all of the questions that need to be answered (i.e. Research question 1 and 3). As the partial knowledge-sensitive criteria may affect incidental vocabulary acquisition, they are tested first. Since the results of testing the partial knowledge sensitive criteria suggest a significant difference when using varying criteria, such criteria then included when comparing the vocabulary test scores across the control and treatment groups. An in-depth analysis of the answers to each type of vocabulary test (i.e. matching, Pinyin production, and character production) is therefore undertaken to demonstrate the amount of incidental word gain through reading in L2 Mandarin Chinese. The section then ends with a brief examination of participants' attitudes towards different types of information within the annotations.

### 5.2.1 The effect of the partial knowledge-sensitive criteria

The recognition test was formulated as a matching test, which meant participants did not need to provide word meanings. Partial knowledge of word meaning was therefore not measured, and only the answers to the Pinyin and character production tests were used to examine the effect of partial knowledge-sensitive criteria. This section begins by utilising the one-way ANOVA test to compare the scores of the meaning-based Pinyin form production test (as calculated using both criteria); the same procedure is then repeated for the meaning-based character form production test. The results of the statistical analysis for each test are followed by the percentage of correct test answers, to demonstrate the increase in knowledge gain achieved by adopting the partial knowledge-sensitive criteria.

Firstly, for the meaning-based Pinyin form production test, only the percentage of correct answers provided by the treatment group was examined, because as mentioned above, the control group provided hardly any correct answers to this test and was thus excluded from the analysis. The one-way ANOVA results indicate a significant difference between the scores marked using different criteria for the meaning-based Pinyin form production test ($F(6.12, 58.5) = 6.12$, $p = 0.03$).

Specifically, among the 14 participants in the treatment group who produced the Pinyin form for each of the five target words, only ten answers were completely correct. Using criteria not sensitive to partial knowledge, the percentage of correct answers is 14.29% (10 / (14 x 5) = 14.29%, where 14 = the number of participants, 5 = the number of target words). Under partial knowledge-sensitive criteria, however, each Pinyin syllable is divided into three parts (i.e. initial, final, and tone), which means every two-character target word has six parts in total. All correct answers (161) for this type of small part of a word were recorded. Using this type of criteria, the percentage drastically increased to 38.33% (161 / (14 x 5 x 6) = 38.33%, where 161 = number of correct answers to part of the Pinyin syllable, 14 = the number of participants, 5 = the number of target words, 6 = the number of parts in each Pinyin syllable).

As to the meaning-based character form production test, the one-way ANOVA results indicate a significant difference between the scores marked using different criteria ($F(5.33, 36.58) = 5.33$, $p = 0.01$). In particular, utilising criteria not sensitive to partial knowledge, only a completely correct character form for each two-character word was awarded a point. As such, correct answer percentages were 7.14% for the treatment group and 4.29% for the control group. Using the partial knowledge-sensitive criteria, which divided the five target words into 33 small components (see Section 4.2.4 and Appendix D), caused these figures to increase to 18.07% for the treatment group and 11.26% for the control group.

The above results reveal that the partial knowledge-sensitive criteria apparently had a significant effect on the results of the production tests in this experiment. The

percentage of correct answers under such criteria nearly tripled the percentage attained using criteria not sensitive to word partial knowledge.

### 5.2.2 The effect of Pinyin annotation

This section compares the results of vocabulary posttests for the control and treatment groups to ascertain whether providing Pinyin in the annotations affected incidental vocabulary acquisition in L2 Mandarin Chinese by helping participants to gain sound-related and other types of knowledge. The scores for the recognition and production tests are compared to determine the learning results of different types of word knowledge, including the link between Pinyin form and word meaning, the link between the Pinyin and character forms, the link between character form and word meaning, and knowledge on the Pinyin and character forms of the target words.

Among all of the tests conducted, the scores of the control and treatment groups differed only in the meaning-based Pinyin form production test. According to the Mann-Whitney test, the treatment group's scores for that test (M*dn* = 1.50) were significantly higher those of the control group (M*dn* = 0.00, U = 28.00, z = 2.897, p < 0.02, r = - 0.58). In other words, when Pinyin was provided in an annotation, participants acquired significantly more knowledge related to the Pinyin form.

With respect to the other tests, the one-way ANOVA results indicate no significant differences in the control and treatment groups' scores for the meaning-based character form production (handwriting) test (F (8.18, 582.31) = 0.323, p = 0.575), character form recognition test (F (3.68, 25.68) = 3.29, p = 0.083), Pinyin form recognition test (F (3.14, 32.86) = 2.2, p = 0.152), or character-based Pinyin form production test (F (2.86, 25.136) = 2.864, p = 0.119). This suggests that the groups gained a similar amount of word knowledge in these other vocabulary posttests. In other words, the results of the statistical analysis suggest that the Pinyin annotation was not more helpful for acquiring the word knowledge tested in this experiment.

### 5.2.3 Amount of word knowledge gain

This section identifies knowledge gain of the target words by examining the percentages of correct answers attained in different vocabulary posttests, including 1) the meaning-based Pinyin form production test, 2) the meaning-based character form production test, and 3) the matching recognition test. It should be noted that the amount of knowledge gain was calculated using the partial knowledge-sensitive criteria, apart from for the matching test (where partial knowledge cannot be demonstrated in this experiment).

**1) The meaning-based Pinyin form production test**

As discussed in Chapters 3 and 4, the word knowledge needed to complete the meaning-based Pinyin form production test includes Pinyin form and the link between Pinyin form and word meaning. The overall amount of related word knowledge gain can be demonstrated by the percentage of correct answers provided concerning a whole target word's Pinyin form. Elaborate information on knowledge gain related to each part of the Pinyin form for both of a target word's characters can also be presented given that the criteria employed are sensitive to partial knowledge of the Pinyin form. In the data collection stage, it was discovered that participants in the control group produced almost no correct answers in this test. As such, only results for the treatment group are included in the analysis. Table 5.2 in the next page presents the control group's correct answer percentages in relation to each part of the Pinyin syllable for both individual characters and the complete target words.

Table 5.2 reveals that on average participants can produce 38.33% of the Pinyin form correctly in this test, which indicates a mean percentage of knowledge gain in Pinyin form (productive) of 38.33%; of that gain, 40% – 50% is in the initial and final parts with just over 24% in the tone part. As it attained the lowest percentage of correct answers in all of the Pinyin syllables produced, tone – or more precisely, tone mark – is the most difficult part of the Pinyin form to learn incidentally through reading

in L2 Mandarin Chinese. It is also noted that production of the Pinyin forms for both "耐克 (Nàikè) [Nike]" (51.19%) and "絲綢 (sīchóu) [silk]" (61.9%) outperformed other words (which ranged from 22.62% to 30.95%) in this test. Moreover, the correct answer percentages for almost all parts of the Pinyin forms of these two words are higher than for other target words. These results suggest that the Pinyin forms of these two words may be easier to obtain than the Pinyin forms of other target words.

Table 5.2 Percentage of correct answers for the Pinyin form production test

| Target word | Correct answers for the Pinyin of the first character (%) | | | Correct answers for the Pinyin of the second character (%) | | | Correct answers for Pinyin of the whole target words (%) |
|---|---|---|---|---|---|---|---|
| | Initial | Final | Tone | Initial | Final | Tone | |
| 絲綢(sīchóu) [silk] | 57.14 | 57.14 | 35.71 | 57.14 | 64.29 | 35.71 | 51.19 |
| 價值(jiàzhí) [value] | 28.57 | 28.57 | 21.43 | 28.57 | 28.57 | 0 | 22.62 |
| 改變(gǎibiàn) [change] | 35.71 | 35.71 | 7.14 | 28.57 | 28.57 | 14.29 | 25 |
| 耐克(Nàikè) [Nike] | 71.43 | 64.29 | 35.71 | 71.43 | 78.57 | 50 | 61.9 |
| 判斷(pànduàn) [judge] | 57.14 | 35.71 | 21.43 | 28.57 | 21.43 | 21.43 | 30.95 |
| Mean | 50 | 44.28 | 24.28 | 42.86 | 44.29 | 24.29 | 38.33 |

**2) The meaning-based character form production test**

As discussed in Chapters 3 and 4, the word knowledge needed to complete this meaning-based character production (handwriting) test are word character form and the link between word character form and meaning. The overall amount of related word knowledge gain can be demonstrated by the percentage of correct answers attained for the character forms of whole target words. Elaborate information on knowledge gain in each character of the target words can also be presented given that the criteria employed are sensitive to partial knowledge of Mandarin Chinese characters. Table 5.3 below indicates the percentages of knowledge gain among the treatment and control groups for both individual characters and the complete target

words.

Table 5.3 Percentage of correct answers in the meaning-based character form production test

| Group | Correct answers for the first character (%) | Correct answers for the second character (%) | Correct answers for the whole word (%) |
|---|---|---|---|
| **絲綢 (sīchóu) [silk]** | | | |
| Treatment | 14.88 | 18.45 | 33.33 |
| Control | 16.07 | 13.69 | 29.76 |
| **價值 (jiàzhí) [value]** | | | |
| Treatment | 4.76 | 9.52 | 14.29 |
| Control | 3.57 | 3.57 | 7.14 |
| **改變 (gǎibiàn) [change]** | | | |
| Treatment | 4.76 | 3.57 | 8.33 |
| Control | 2.38 | 4.76 | 7.14 |
| **耐克 (Nàikè) [Nike]** | | | |
| Treatment | 10 | 27.86 | 37.86 |
| Control | 2.14 | 10 | 12.14 |
| **判斷 (pànduàn) [judge]** | | | |
| Treatment | 2.14 | 5 | 7.14 |
| Control | 1.79 | 2.86 | 4.64 |
| **Overall results of the meaning-based character form production test** | | | |
| Treatment | 6.25 | 9.93 | 16.18 |
| Control | 5.29 | 8.98 | 14.27 |

The above table reveals that on average the participants in the treatment group produced 16.18% of the target words' character forms correctly, with correct ratios of 6.25% for the first characters and 9.93% for the second characters. In contrast, the control group achieved correct ratios of 14.27% for the whole target words, 5.29% for the first characters, and 8.98% for the second characters. The treatment group thus achieved slightly higher percentages of correct answers than the control group on almost all items compared, except for two characters: "絲" in "絲綢 (sīchóu) [silk]" and "變" in "改變 (gǎibiàn) [change]".

It is also noted that results for both "絲綢 (sīchóu) [silk]" (33.33% for the treatment group and 29.76% for the control group) and "耐克 (Nàikè) [Nike]" (37.86% for the treatment group and 12.14% for the control group) were higher than other

words, with the results ranging from 7.14% to 14.29% for the treatment group and from 4.64% to 7.14% for the control group in the posttest. The higher scores for both characters in those two target words contributed to these results.

**3) The matching recognition test**

To complete the matching recognition test, the word knowledge needed includes links between the target words' Pinyin and character forms, Pinyin form and word meaning, character form and word meaning, as well as Pinyin form and character forms. The overall amount of related word knowledge gain is demonstrated by the percentage of correct matches made between each of these pairs. Elaborate information on knowledge gain in each target word is also presented for the treatment and control groups. Table 5.4 below shows the overall and detailed percentage of correct matching made by the participants in this test by both groups.

Table 5.4 Percentage of correct matching: Pinyin form, character form, and word meaning

| Group | Correct matching of Pinyin and character forms (%) | Correct matching of Pinyin form and meaning (%) | Correct matching of meaning and character form (%) |
|---|---|---|---|
| **絲綢 (sīchóu) [silk]** | | | |
| Treatment | 100.00 | 100.00 | 100.00 |
| Control | 63.64 | 72.73 | 90.91 |
| **價值 (jiàzhí) [value]** | | | |
| Treatment | 85.63 | 78.48 | 78.48 |
| Control | 63.63 | 27.27 | 45.45 |
| **改變 (gǎibiàn) [change]** | | | |
| Treatment | 57.14 | 71.43 | 78.57 |
| Control | 54.54 | 45.45 | 36.36 |
| **耐克 (Nàikè) [Nike]** | | | |
| Treatment | 92.86 | 100.00 | 92.86 |
| Control | 36.36 | 81.81 | 45.45 |
| **判斷 (pànduàn) [judge]** | | | |
| Treatment | 64.29 | 71.43 | 85.72 |
| Control | 45.45 | 45.45 | 54.54 |
| **Overall results of the matching test for all target words** | | | |
| Treatment | 80.00 | 90.00 | 87.14 |
| Control | 52.73 | 65.45 | 65.45 |

The above table indicates that on average the participants correctly matched the target words' Pinyin and character forms 80% of the time; the correct ratio was 90% for matching Pinyin form and word meaning 87.14% for character form and word meaning. For the control group, the respective ratios were 52.73%, 65.45%, and 65.54%. The treatment group clearly outperformed the control group in each type of matching within this test, given that it achieved higher results in every item analysed. The participants' answers also reveal that 50% (7 participants) of the treatment group matched all five words completely correctly, which is much higher than in the control group – where only 9.09% (1 participant) had fully correct answers.

It is not surprising that the treatment group achieved much higher scores than the control group when matching the Pinyin with the target words' meaning or character form, as participants in this group had access to Pinyin annotation. However, it should be noted that the control group's correct percentages were unexpectedly also very high (i.e. all above 50%). It was initially assumed that the participants in this group might only be able to obtain a small amount of sound-related knowledge – or even no related knowledge at all – as they did not have access to the Pinyin annotation. Reasons that may account for this situation are discussed in detail in Section 5.3.

A final issue merits being addressed here. The in-depth analyses above reveal that the treatment group obtained higher scores than the control group for all types of vocabulary posttests. The same is true for every part of the Pinyin and character forms, with the exception of the two characters mentioned earlier (namely "絲" in "絲綢 (sīchóu) [silk]" and "變" in "改變 (gǎibiàn) [change]"). These findings indicate that Pinyin annotation may have a positive effect on incidental vocabulary acquisition in L2 Mandarin Chinese, in terms of obtaining not only sound-related knowledge for a word but also knowledge related to character form.

### 5.2.4 Participants' attitudes towards different types of annotations

This section presents results relating to participants' attitudes towards the five

types of annotations, namely word meaning, Pinyin, audio (i.e. the spoken form of target words), phonetic components of characters, and meaning components of characters. These attitudes were solicited using five-point Likert-scale questions, where 1 = not useful and 5 = very useful. The means of the choices made by the participants (with standard deviation) are presented in Table 5.5 below.

Table 5.5 Mean scores for the participants' views on different types of annotations

| Type of annotation | Mean | Std Deviation |
|---|---|---|
| Meaning | 4.32 | 0.894 |
| Pinyin | 3.86 | 0.941 |
| Audio (word read by a native speaker) | 3.32 | 0.995 |
| Phonetic components of characters | 2.77 | 0.922 |
| Meaning components of characters | 3.32 | 1.129 |

A *Friedman's two-way analysis of variance test* (related samples) revealed significant differences among the five types of annotations ($\chi 2$ (4) = 24.011, p = 0.000). A post-hoc analysis with Wilcoxon signed-rank test was conducted with a Bonferroni correction, which resulted in a significance level set at $p < 0.005$ (see Section 5.1.4 for more explanation). A significant difference was seen between the following pairs of annotations: word meaning and phonetic components of characters (Z = -3.821, p = 0.00), word meaning and meaning components of characters (Z = -2.977, p = 0.003), and Pinyin and phonetic components of characters (Z = -2.937, p = 0.003).

The last difference listed above (i.e. Pinyin vs. phonetic components of characters) is noteworthy, as both annotations involved can provide sound-related information for the target words. As the means are 3.86 for the Pinyin annotation and 2.77 for the phonetic components of characters, it is clear that the participants regarded the latter type of annotation as much less helpful in terms of learning unknown words from reading. However, when interpreting the results, it should be borne in mind that they now do not seem logically consistent. For example, if the attitudes towards the annotation that entailed marking the phonetic components of characters were significantly different than for the Pinyin annotation but not

significantly different than for the audio annotation, why did the statistical tests not identify a significant difference between the Pinyin and audio annotations? An error therefore seems to have been made in the statistical analysis. It is beyond the scope of this research to explore the issue further, but the above results should be interpreted with caution.

## 5.3 Discussion

The discussion in this section focuses mainly on important phenomena observed in this experiment, with the goal of improving the design of the second experiment. Other issues, including types and amount of partial knowledge gain and the use of the partial knowledge-sensitive criteria, are discussed in Chapter 8 (which presents a general discussion on the findings of both the first and second experiments).

Two phenomena need to be addressed here in particular. Firstly, as suggested by the results of the first experiment (as presented in the previous section), the control group unexpectedly obtained relatively high scores when matching both target words' Pinyin and character forms and Pinyin form with word meaning. Secondly, some target words seem to be easier than others in this experiment, in terms of acquiring both different types of word knowledge and partial knowledge of words. Reasons that may account for these two phenomena include the design of the matching test and question used to collect information on participants' pre-knowledge of the target words, the frequency of target word encounters, and the use of the target words in the reading comprehension question. These two phenomena are discussed in detail in the following sections.

### 5.3.1 Possible effects of the vocabulary tests adopted in this experiment

This section mainly deals with the following two questions: Does the matching test work for evaluating the results of incidental vocabulary acquisition in L2 Mandarin Chinese? How should the pre-knowledge of target words be tested?

As mentioned in Section 4.1, the original assumption for the control group was that without extra sound-related information provided in annotations, participants might obtain little or no knowledge related to the sound of words. In this study, such knowledge refers to the links between both the Pinyin and character forms and Pinyin form and word meaning. Relatively low scores were thus expected in the meaning-based Pinyin form production test, as well as for matching Pinyin with both word character form and word meaning in the matching test. The assumption for the meaning-based Pinyin form production test is confirmed, as the control group produced almost no correct answers in this test. However, the results were not as expected for the matching test, as the control group had correct percentages of 52.73% for matching Pinyin and character forms and 65.45% for matching Pinyin form with word meaning.

One may conjecture that the above high scores were due to a weakness of the matching test that allows participants to guess answers correctly. However, this explanation does not seem to cope with such high correct ratios, although it is possible that the participants did randomly make some correct guesses. It is therefore reasonable to believe that the participants obtained some clues from the test question that made guessing easier. One reason could be the phonetic components involved in the target words. As discussed in Section 3.3.2, the phonetic components of Chinese characters carry sound-related information for a word, although it may be difficult for beginning learners to discover. In this test, several target words contain a phonetic component in one or both of their characters; for instance, “周 (zhōu) [week]” is a phonetic component of the second character in “絲綢 (sīchóu) [silk]”. Other examples of phonetic components can be seen in “價值 (jiàzhí) [value]” (which contains “賈 (jiǎ) [introduce]” and “直 (zhí) [straight]”, both of which function as phonetic components) and “判斷 (pànduàn) [judge]” (which contains only “半 (bàn) [half]” as a phonetic component of the first character). In this case, these phonetic components might affect the results of the matching test.

However, this inference above contradicts the information collected on participants' pre-knowledge of the target words. According to sample selection criteria mentioned in Section 5.1.1, only participants with no more than 10% pre-knowledge of the target words were kept in the final pool. This situation thus leads to conjecture that participants' pre-knowledge answers were inaccurate. Evidence of this was found in the feedback question on the questionnaire, where two participants mentioned that the format of the question related to word pre-knowledge was difficult to understand – which indicates that this question did not fully serve its intended purpose. If this is true, the question concerning pre-knowledge of the target words needs to be improved.

Alternatively, if the participants did correctly answer the question about their pre-knowledge of the target words, the control group's high scores may have been caused by the matching test itself (given that the participants' teacher indicated that they should not have encountered more than three of the phonetic components mentioned above in their textbook or class). Apart from the possibly known phonetic components, the similarity of the Pinyin form and English meaning of both words ("絲綢 (sīchóu) [silk]" and "耐克(Nàikè)[Nike]") may also provide clues that helped participants in guessing. The very high percentage of correct answers in matching the Pinyin form and meaning of these words (respectively 72.73% and 81.81) supports this view. Strictly speaking, the matching test thus serves better as a facilitator for vocabulary learning than as an evaluation tool for word knowledge gain in L2 Mandarin Chinese. As such, another vocabulary test was needed as the recognition test for the second experiment. In light of the discussions in Chapters 3 and 4, a translation test – the only remaining option – was adopted.

A point is also worth mentioning regarding the question used for collecting information about participants' pre-knowledge of the target words. As noted before, participants whose responses were more than 10% correct were excluded from the final analysis. Four actually achieved a very high percentage of correct answers here (ranging from 50% to 70%) and one of them reported having some learning experience

in Korean. It is possible that their answers were affected by the reading exercise and vocabulary test and did not reflect pre-knowledge of the target words. If this was the case, the question related to pre-knowledge would be better placed before the reading material. Moreover, since 11 participants were removed from the final pool as explained in Section 5.1, the experiment's sample size of 25 was relatively small.

It is also reasonable to observe that many participants reported having some knowledge of parts of the target words, as the component is the basic unit that forms Mandarin Chinese character (as discussed in Section 3.3.2). The question is then if not excluding such knowledge (even if limited) from the vocabulary posttest would lead to results that suggest a slightly higher amount of knowledge gain. The scores for pre-knowledge of the target words should thus be taken off from that for the responding posttest scores.

### 5.3.2 The potential ease of certain words

The in-depth analysis carried out in Section 5.2.3 to indicate the amount of word knowledge gain showed higher scores for "耐克 (Nàikè) [Nike]" and "絲綢 (sīchóu) [silk]" than for the other target words in almost every test, for both the treatment and control groups. One may conjecture that these words may simply be easier for the participants; however, considering the discussion of influential factors in Section 2.2.3.5, the explanation could also be connected to two other issues: frequency of encountering the target words and the high task-induced involvement loads in the reading comprehension questions. This study follows the suggestion made in previous studies to limit the number of occurrences of both words no more than eight times; however, instead of appearing two or three times (as the other words did), "耐克 (Nàikè) [Nike]" appeared five times and "絲綢 (sīchóu) [silk]" appeared seven times. They appeared in both the reading article and the reading comprehension questions. The failure to stipulate whether the open questions in the reading comprehension exercises should be answered in English or Mandarin Chinese provides participants the

chance to use the target words in the reading comprehension questions. According to the task-induced involvement theory discussed in Sections 2.2.3.5 and 2.2.3.6, the need to use these words to answer the questions (and thus search for and evaluate them) may cause high task-induced involvement and consequently lead to a better learning of these words. As both words were required for answering the reading comprehension questions, their results from the vocabulary test may not reflect incidental vocabulary acquisition from the reading only – but instead from both reading and completing the reading comprehension exercises. To avoid these problems, the frequency of encountering both words should be reduced and they should not be used in the reading comprehension questions.

However, the possibility of the two words being easier than the others does also exist, and factors relating to word difficulty – although not frequently mentioned in previous studies (see Section 2.2.3.5) – could have contributed to the results. Unfortunately, the related factors investigated in this field, namely part of speech and conceptual difficulty (Laufer, 1990; Lin, 2010; Nagy et al., 1987; Paribakht & Wesche, 1997), do not seem to answer this question. This is because both of these two-character nouns are conceptually easy to understand, which is a feature that three of the experiment's other target words share. Without clear guidance on this issue, it was not possible to make any improvements to target word selection. However, the issue is revisited in following chapters, as the second experiment suggests that it may affect the study's overall results.

A last point worth mentioning here is the question about participants' attitudes towards the annotations. The results indicate that they regard the Pinyin (M = 3.86, SD = 0.941) and audio (M = 3.32, SD = 0.995) annotations as helpful for remembering words in reading but marked phonetic components as less helpful (M = 2.77, SD = 0.922). The last type of annotation was consequently not included in the second experiment and removed from the attitude question in the questionnaire. Moreover, as this study focuses primarily on annotations that provide sound-related information,

the annotation that presented marked meaning components was also excluded from the second experiment.

## 5.4 Summary of findings and implications for the second experiment

### 5.4.1 Observations from the first experiment

This experiment investigated the effect of a sound-related annotation, namely Pinyin, on incidental vocabulary acquisition in L2 Mandarin Chinese by comparing control and treatment group scores for various vocabulary posttests. Based on the above discussion, the following preliminary observations can be drawn concerning the effect that providing a Pinyin annotation has on incidental vocabulary.

Firstly, in relation to the first sub-question of research question 1, this experiment's results clearly demonstrate that sound-related knowledge for a word can be acquired incidentally through reading in the context of L2 Mandarin Chinese learning. These results indicate the possibility that the following types of sound-related word knowledge can be acquired: Pinyin form, link between Pinyin and character form, and link between Pinyin form and word meaning. Moreover, the results of both the meaning-based character form production and matching tests provide evidence that knowledge related to word character form can be acquired incidentally in the process of reading in L2 Mandarin Chinese (which is a field that has been rarely researched). In addition, the knowledge gain in both the character and Pinyin forms of the target words exceeded the receptive level and reached the productive level.

Secondly, the only significant difference found between the control and treatment groups was in the meaning-based Pinyin form production test, which indicates that a Pinyin annotation is essential for learners to obtain sound-related knowledge at a productive level. More importantly, although the test scores were not significantly different between the two groups, the treatment group's overall higher scores indicate

that Pinyin annotation may have positive effects on obtaining both other types of sound-related word knowledge (e.g. links between Pinyin and character form) and knowledge not related to word sound (e.g. character form and the link between character form and word meaning).

Thirdly, in relation to the second sub-question of research question 1, the experiment demonstrated the following word knowledge gain: 38.33% in the meaning-based Pinyin form production test and 16.18% and 14.27% respectively for the treatment and control groups in the meaning-based character form production test (by handwriting). In the matching recognition test, word knowledge gained related to links between Pinyin and character form, Pinyin form and word meaning, and character form and word meaning; here the scores were respectively 80%, 90%, and 87.14% for the treatment group and 52.73%, 65.45%, and 65.45% for the control group.

Finally, a significant difference was also found between the scores of the vocabulary posttest marked using criteria that are or are not sensitive to partial word knowledge. This suggests the possibility that the amount of word knowledge gain was underestimated in previous studies.

### 5.4.2 Important implications for the second experiment

To explore the possibility of acquiring sound-related word knowledge incidentally through reading in the context of L2 Mandarin Chinese learning, this study's second experiment involved adding audio annotation to investigate the effects of learners being exposed real word sounds. As such it was very important to make some major improvements to the research design based on problems found in the first experiment.

Three problems were identified and related research design need to be improved for the second experiment. Firstly, although the guessing of the recognition (i.e. matching) test could be reduced by involving some distractors, this option does not seem to fit the requirements of this study on incidental vocabulary acquisition in L2 Mandarin Chinese. Using the transliterated word "耐克 (Nàikè) [Nike]" greatly reduced

the difficulty of the matching questions in this experiment. In addition, participants may be able to guess answers correctly with partial knowledge of a word (e.g. its phonetic or meaning components), which could make it difficult to identify the type of knowledge being tested.

Secondly, considering the possibility of pre-knowledge not being measured properly, it is difficult to support the idea that people do not have knowledge of certain words and then consequently learn them through reading. Therefore, a better pretest was required. In addition, Deducting the pre-knowledge test scores from the posttest scores in the second experiment would enable incidental vocabulary acquisition to be evaluated more accurately through knowledge gain (as represented by the word score differences between the pre- and posttests, rather than simply the posttest scores).

Finally, the finding that the scores for "耐克 (Nàikè) [Nike]" and "絲綢 (sīchóu) [silk]" were much higher than those for the other target words in both the matching and production tests may have been influenced by two factors: the high frequency of occurrence of these words and the high task-induced involvement attached to them in the reading comprehension questions. To address the first issue, the numbers of occurrences of both words were reduced in the second experiment and no target word appeared more than three times in the articles. The second issue was handled by not involving any target words in the reading comprehension tasks. Moreover, only English answers are required for the reading comprehension questions, to avoid extra usage of these words and high task-induced involvement.

# Chapter 6: The second experiment: The CALL environment

This experiment adopted a quantitative research approach by utilizing unexpected vocabulary tests attached to a reading comprehension task in an online reading environment in order to compare annotations containing multimodality sound information, namely text and Pinyin, text and audio, and text with Pinyin and audio annotations (in order to answer the second research question mentioned in Section 4.1). With the objective of probing more deeply into the processing mechanism with the Mayer CTML model, audio annotation was utilized considering the positive effect on incidental vocabulary learning it could have, based on the earlier discussion on CTML Section 2.2.3.6. However, in a normal paper-based context, it is very difficult to include such audio annotations. Therefore, an online program was first developed by the researcher to solve this problem. It was also hoped that such an online language learning program could provide some initial empirical evidence on incidental vocabulary learning in L2 Mandarin Chinese, as discussed in Section 2.2.3.3, thus contributing to a deep understanding of research in this field. As mentioned in Section 2.2.3.6, incidental word learning starts from an important, yet measurable notion, namely noticing. A tracking tool was therefore embedded into the online program to determine whether participants noticed the target words or not (for more details about the online program design please refer back to section 4.2.7 and are thus not repeated here).

Based on the discussion of the methodological concerns in Chapter 4 and the limitations of the first experiment discussed in Section 5.4, improvements on the research design made for this experiment are presented in relevant sections in this chapter. To recapitulate, the major improvements made were as follows: replacing the matching test with a translation test; adding a proper pretest to test participants' pre-knowledge of target words; reducing the frequency of occurrence of both "耐克 (Nàikè) [Nike]" and "絲綢 (sīchóu) [silk]" in the article and removing them from the

reading comprehension questions; and deleting the annotation types not included in this experiment (namely phonetic and meaning components for characters) from the attitude question in the questionnaire.

Details of the related changes made to the vocabulary pre- and post-tests, the crossover research design, and the researcher-designed online program are presented in the first part of this chapter. Thereafter, the experiment's results are presented; however, they are not discussed until the next chapter, where related results from the first experiment are also taken into consideration. Although no statistically significant difference was found between the three types of annotation, there is a tendency for the text with Pinyin and audio annotation to be more helpful than the other two types of annotations. In addition, the in-depth analysis of vocabulary test scores in Section 6.2.2 reveals the amount of word knowledge gained and helps clarify the results of the statistical analyses in relation to the first research question. Furthermore, participants' attitudes towards the different types of annotation seem to be consistent with the finding of the first experiment.

## 6.1 Methods

This section outlines the methods used to conduct the second experiment, including in relation to participants; instruments (i.e. the reading comprehension exercise, vocabulary tests, questionnaire, and online program); procedures; and data collection, coding, and analysis. Information related to these issues that was presented in Chapters 4 and 5 is not repeated here.

### 6.1.1 Participants

A total of 52 proficiency-matched students from both SOAS and the University of Nanjing (China) were recruited as participants for this experiment. The 28 SOAS students had nearly completed the Special Chinese II course at the time of participation. The main reason for involving them was that by the time of the

experiment they had learned 14 lessons, which suggests a mastery of the same amount of 546 characters from their textbook (*Colloquial Chinese*) as the participants from the first experiment. However, unlike the participants in the first experiment, these non-Chinese major students were not required to master traditional Chinese characters and had received instruction in simplified Chinese characters from the beginning of their course. Therefore, all the characters in this experiment were simplified Chinese characters.

In the first experiment, 16 of the 41 participants were excluded from the final analysis for various reasons; 68.75% (11 participants) were excluded due to having too much pre-knowledge of the target words, leaving a relatively small sample (see Section 5.1.1). This situation might happen again because it is almost impossible to predict what students might have been exposed to outside class. If this is the case in future, a sample size that is much smaller than 28 could affect both the reliability and value of the results of this experiment. Therefore, more participants would be needed to increase the sample size. The test was conducted in early May, after the SOAS participants had completed their learning of the 14 lessons. This was during the university examination period in the UK. Therefore, it was very difficult to find a relatively large group of university students in the UK to take part in the research at that time. Thus international students studying at universities in China were taken into consideration. At last 24 proficiency-matched international students from the University of Nanjing were recruited. At the time of the study, they had had one term of intensive lower-beginner Chinese classes and, according to their teachers, should have been able to achieve HSK Level 3, which suggests a mastery of approximately 623[12] Chinese characters. The large overlap in the characters that the participants from the two universities had to master through their coursework thus suggested that

[12] The number of characters covered by the HSK Level 3 is 623, according to my calculation based on the vocabulary lists for HSK Level 1 to Level 3 (Hanban, 2011b).

they had achieved a similar level of Mandarin Chinese. Apart from that, to reduce any possible influences caused by the difference between the proficiency of students from SOAS and those from the University of Nanjing, a crossover research design was used, which will be explained in great detail in Section 6.1.2.3.

In addition, they were told that the test would take approximately an hour and they could complete it at their own pace. The data collected by the tracking tool mentioned in Section 4.2.7 revealed that the students took 65.22 minutes on average to complete all the tasks (See Section 7.7 for detailed information on time spent on each part of the test).

Most participants from SOAS have English as their native language. As in the first experiment, it was assumed that the others speak fluent English (as they are studying at SOAS) and would not have difficulties in reading the instructions and annotations in the test. The participants from the University of Nanjing were told that fluent English was required to complete the online exercise before they volunteered to participate. Data from Japanese and Korean students were later excluded during the data coding stage, as the similarity between Chinese and their native languages (in either character form or word pronunciation) may benefit them when learning Chinese vocabulary.

A final pool of 48 participants remained after participants were eliminated for one of the following reasons: 1) scoring more than three out of the total 15 in the pretest (1 participant); 2) failing to answer all of the sections in the posttest (one participant); or 3) having knowledge of Japanese Kanji or Korean Hanja (two participants).

As the participants had varied learning environments, it was important to ensure that the differences did not come into play when comparing the effects of annotation types. A “crossover research design” was thus required to counterbalance any possible influence, as detailed in Section 6.1.2 below. In actuality, a comparison of the scores for the participants from both universities does not suggest any significant difference. However, as the study’s objective was not to investigate differences between the groups, the issue is not further discussed.

## 6.1.2 Instruments

### 6.1.2.1 Reading material and target words

The three articles employed in the second experiment were written in simplified Chinese characters (which are now officially used in mainland China), rather than in traditional characters as in the first experiment. It was hoped that adopting simplified Chinese characters would enable more participants (such as those from the University of Nanjing) to take part in the study.

This experiment utilised three articles (see Appendix A) mainly due to test length considerations. Table 6.1 provides details of the articles, including title, length (by number of characters), and target words. A total of 15 target words (five from each article) were selected from the reading materials. To recapitulate, the main target word selection criteria were as follows: (1) the word is formed from two characters, (2) it is not included in the vocabulary list of HSK Level 3 or below, and (3) neither of the word's characters is included in the participants' textbooks (for more details, see section 4.2.2). The 15 target words selected included seven nouns, four verbs, two adjectives, and two adverbs.

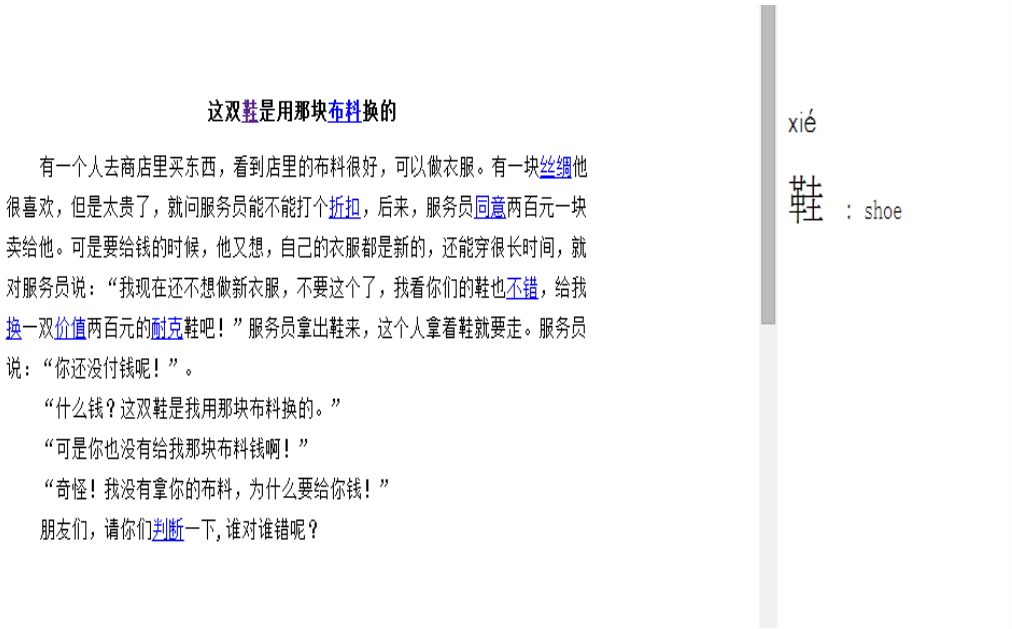
Table 6.1 Overview of articles used in the second experiment with target word

| Article title | Length (number of characters) | Target words (characters, Pinyin, English equivalent) |
|---|---|---|
| "This silk is payment for the pair of shoes" | 236 | 1 丝绸(sīchóu) [silk], 2 价值(jiàzhí) [value],<br>3 耐克 (Nàikè) [Nike], 4 判断(pànduàn) [judge],<br>5 折扣(zhékòu) [discount] |
| "Napping" | 212 | 6 减缓(jiǎnhuǎn)[slowdown], 7 眩晕(xuànyūn) [dizziness],<br>8 尽量(jìnliàng) [to the best of one's ability],<br>9 症状(zhèngzhuàng) [symptom], 10 稍微(shāowéi) [slightly] |
| "Stories of life" | 208 | 11 庄稼(zhuāngjia) [crops],12 辛苦(xīnkǔ) [to work hard; laborious], 13 培养(péiyǎng) [cultivate],<br>14 保持(bǎochí) [to keep; to maintain],<br>15 懒惰(lǎnduò) [lazy] |

### 6.1.2.2 Article format (online)

The characters in the article are in the Song font, which is often used on Chinese websites and in Chinese books and newspapers. Character size is set to five for Chinese characters (a little smaller than size 11 for English words), but participants can change this by zooming or out. As shown in Figure 6.1, the article and comprehension questions are presented on the left side of the webpage (taking up roughly 75% of the space) and annotations and instructions relating to the exercise are presented in a narrow column on the right.

Figure 6.1 Screenshot of article A with text + Pinyin annotation




All of the target words and distractors are underlined and written in blue to attract participants' attention. When used on the Internet, this format used normally suggests that a reader can open another page by clicking on the link. It was possible that readers would not pay any attention to such links or bother to click them. If this would have happened, the program would have failed the initial and most important goal, namely to compare the effects of different kinds of annotations. An instruction on the right

side of the page therefore also stated: “Please click the underlined word to see annotation. In this part, you'll see the word with Pinyin and English meaning.” Clicking the word made the annotation appear in a column on the right of the screen. The character size in the annotations is slightly larger than in the article itself.

#### 6.1.2.3 Crossover research design

The research design also needs to be addressed here as it affects the arrangement of the articles and annotations on the webpages. To reduce the influences possibly caused by learner and word differences (as suggested by the first experiment), the combination of words and varying types of word knowledge, and sequence (given the use of three articles), a modified crossover research design was employed. A crossover design originally refers to a longitudinal study in which participants engage in a sequence of interventions, such that each one individual participates in all conditions. As discussed in Section 2.2.3.3, Fisher et al. (2012) adopted such a design in their research. As mentioned earlier, this design may have affected the incidental nature of their study, especially for the second and third tests – which hence failed to meet the requirements for research on incidental vocabulary acquisition as the vocabulary posttest was not unexpected. To ensure that the current experiment qualifies as incidental vocabulary acquisition research, the design was modified by presenting the vocabulary posttest after the reading comprehension exercises for all the three articles.

Adopting such a design meant participants encountered the articles in different orders but the annotation types always in the same order, namely text + Pinyin, text + audio, and text + Pinyin + audio (which was last as it provides richer information than the other two types of annotations). The article combinations (with types of annotation) are shown in Table 6.2. This arrangement meant every participant had the chance to access all three types of annotations and read all three articles, as well as to encounter all 15 target words. It thus eliminated any possible impacts on learning results caused by learner ability and word differences. As the articles were arranged in

different orders, the chance association between specific words and annotation types could be avoided.

Table 6.2 Combinations of articles and annotations

| | Sequence of article and annotation type | | |
|---|---|---|---|
| 1 | Article1-text + Pinyin | Article 2-text +audio | Article 3- text + Pinyin + audio |
| 2 | Article2-text + Pinyin | Article 3- text +audio | Article 1- text + Pinyin + audio |
| 3 | Article3- text + Pinyin | Article 1- text +audio | Article 2- text + Pinyin + audio |

Finally, as suggested by the results of the first experiment, none of the target words appeared in the reading comprehension questions (which included both multiple-choice and open questions). An instruction attached to the open question, which required participants to summarise the article's main idea, specified that the answer should be written in English to ensure that the participants did not use the target words.

### 6.1.3 Vocabulary tests

As discussed in Section 5.3.1, the results of the first experiment suggest that a proper pretest is required to measure participants' pre-knowledge of the target words. Moreover, concerning the format of the recognition test, the matching test should be replaced with a translation test. However, it should be noted that given the online testing environment of the second experiment, the word knowledge assessed by the Pinyin input assisted meaning-based character form production test under the same format of that used in the pen-and-paper environment was changed. This section presents detailed information on related tests.

#### 6.1.3.1 Vocabulary pretest

The pretest that constituted the first part of the online program was designed to obtain information concerning participants' pre-knowledge of the target words. However, its design raised the issue related to extra exposures to the target words that was discussed in Sections 4.2.3 and 5.3.1. To decrease possible influences, distractors were included in the pretest to reduce participants' attention to the target words. All 15 target words and 16 distractors were tested randomly and did not follow the same order as they appeared in the articles. To avoid drawing participants' particular attention to these words, at the beginning of the test it was explained to them that they would encounter words in the vocabulary pretest that they may not have learned before – but they did not need to worry about them, as the primary target of the vocabulary pretest was to assess their vocabulary and language level and they would have access to annotations for these words during the reading itself. The participants were also encouraged to complete the pretest quickly without spending too much time trying to guess the meanings of unknown words.

A modified VKS test was adopted to measure participants' pre-knowledge of the target words, as the different levels in such a test's questions make it possible to understand the status of participants' various types of word knowledge and partial knowledge of those words. Each question was formulated with a prompt and three levels related to word mastery. Test-takers had to choose the expression that best described their knowledge of a word.

The VKS used by Paribakht and Wesche (1993) and Qian (2003) (see Section 3.2) was modified in three ways for the current experiment: first, apart from providing meaning for word character form, Pinyin form is must also be provided by the participants; second, the scale related to making a sentence with a word, which requires word knowledge beyond form and meaning, was removed; and third, in addition to using word character form as a prompt, the spoken form of the target words was also utilised as a second type of prompt.

The first part of the vocabulary pretest, which employed a word's written character form as the prompt, was used to decide whether the participant could recognise the written form of a word, connect its Pinyin and character forms, and produce its Pinyin version. For example:

Prompt word (written in character form)

1) I have never seen this word before;

2) I have seen this word before, but I don't remember its meaning or sound;

3) I know the meaning/Pinyin of the whole word or part of the word_____.

The second part of the test had the same format, but the prompt was instead a sound clip. The aim was to assess participants' knowledge related to the sound of words, which connects spoken form with meaning. For example:

Prompt word (audio clip)

1) I have never heard this word before;

2) I have heard this word before, but I don't remember its meaning or sound;

3) I know the meaning of the whole word or part of the word_____.

In both formats, participants were required to provide detailed information of the word in the given space if they chose the last response. The program default was set in a way that participants were not allowed to submit their answers if they did not provide required information on partial knowledge. An instruction was also included to encourage participants to add as much information as they could in the blank, even part of the word's meaning or Pinyin.

#### 6.1.3.2 Vocabulary posttest

The vocabulary posttest, which was designed to evaluate participants' knowledge of all 15 target words, came after reading. The participants needed to read all three articles and complete the reading comprehension tests before they could access the

vocabulary tests. To obtain more information on incremental knowledge from the participants' answers, the test included instructions that encouraged the participants to include as much information as they could about the word in each question (e.g. even part of the word's Pinyin).

Both recognition and production tests were employed to form the four sequential parts of the posttest, namely 1) a meaning-based Pinyin form production test, 2) a Pinyin input assisted meaning-based character form production test, 3) a Pinyin form recognition test, 4a) a character form recognition test, and 4b) a character-based Pinyin production test. To avoid any influence that might arise from extra encounters with the target words in the recognition test, the production tests were given first. In the first two parts, participants were asked to respectively provide a word's Pinyin or character form from its English meaning; in part 3 they were then asked to write the English meanings of words according to sound files. Thereafter, in part 4 participants were asked to provide the Pinyin and English meaning based on words written in characters (see Appendix B).

Although all of the above-mentioned tests were conducted in a translation test format, the types of word knowledge tested differed from what was assessed both in previous studies and the current study's first experiment (which was conducted in a pen-and-paper environment). The Pinyin input assisted meaning-based character production test was used in the context of the online test. Along with the development of computer science, the Pinyin input has been widely used as a tool for writing Chinese characters into a computer. Compared with other types of input methods, the Chinese Pinyin input method requires almost no extra learning as long as Pinyin is mastered by users. The Pinyin input method is hence widely used for Chinese inputting and was therefore also adopted for this study. Although it was still possible to test handwriting in a character production environment, it is more reasonable to adopt a test that meets the production procedure normally involved when a computer is used rather than simply adapting what can be done in a pen-and-paper environment to the

CALL context.

Table 6.3 below presents the word knowledge that each vocabulary posttest in this experiment assessed. Decisions were based on the discussions of word knowledge and vocabulary tests in Chapters 3 and 4.

Table 6.3 Word knowledge assessed by each vocabulary test

| Vocabulary test | Word knowledge needed to complete the test |
|---|---|
| 1) Meaning-based Pinyin production test | Pinyin form (productive level)<br>Link between Pinyin form and word meaning |
| 2) Pinyin input assisted meaning-based character form production test | Pinyin form (productive level)<br>Link between Pinyin form and word meaning<br>Link between Pinyin and character forms<br>Character form (receptive level) |
| 3) Spoken form recognition test | Spoken form (receptive level)<br>Link between spoken form and word meaning |
| 4a) Meaning-based character form recognition test | Character form (receptive level)<br>Link between character form and word meaning |
| 4b) Character-based Pinyin production test | Pinyin form (productive level)<br>Character form (receptive level)<br>Link between Pinyin and character forms |

### 6.1.4 The piloting of the online program

As noted previously, an online program was designed to administer the above-mentioned vocabulary tests to the participants in the second experiment (see C Section 4.2.7 for the details of the program). After the online program was established, it was piloted to determine whether it served the purpose it was designed to, namely providing reading materials with different types of annotations; collecting answers to vocabulary tests and a questionnaire; and recording participants' clicks during the whole process. This piloting entailed sending a link to the online test to different groups of people to test the whole reading program itself. Three Mandarin Chinese teachers were first asked to check the articles and language used in the program. After a few typing errors were fixed and misleading instructions were rewritten, the link was

sent to seven L2 Mandarin Chinese learners who were asked to complete the whole test online as test-takers in the pilot phases. Some of them reported technical problems, such as sometimes not being able to open the sound files and test pages. While the online program did not seem very stable, it was difficult for me to solve the problems the learners encountered or find a remedy in time (particularly as the learners completed the online exercises by themselves in different locations and countries). As a result, the participants actually formally involved in the second experiment had to be asked to come to a classroom where laptops and Internet access had been set up and checked before the test.

### 6.1.5 Experiment procedures

A short introduction to the online program was given by the researcher at the beginning of Mandarin Chinese classes in both universities to recruit participants. Students who wished to participate had to book a time slot and then arrange to come in small groups accordingly. Considering the technical issues reported during the piloting phases, this experiment was finally conducted in a classroom with computers set up for the participants, with the researcher present throughout to both immediately solve any technical or connectivity problems encountered by the participants and answer any operational questions.

Upon arrival, participants were briefly introduced to the online reading program. This included how to enlarge the text, which web browser can be used, how to enable the sound files in the annotation, and how to use the Google Pinyin Input. At this stage, participants were also instructed as to how to provide their consent by clicking a link. Consent information was linked to the program's starting page, where participants were informed that their data would be used for research purposes and that they could withdraw at any time. The format of the test and what participants were expected to do were also briefly explained. The students were told that they would complete some reading comprehension exercises at their level to help test an online reading program;

they were also informed that they could expect both a vocabulary pretest for language level evaluation purposes at the beginning and a questionnaire relating to the exercises at the end. However, to ensure the vocabulary posttest was completely unexpected, it was not mentioned at this juncture. In addition, they were told that the test would take approximately about an hour and they can complete it at their own pace. The data collected by the tracking tool mentioned in Section 4.2.7 revealed that the students took 65.22 minutes on average to complete all the test (See Section 7.7 for detailed information on time spent on each part of the test).

When completing the reading comprehension exercises and vocabulary tests, the participants could only complete the sections in sequence. To prevent them from going back to a previous page in the test, only a "next" button was embedded in the test. Later pages also included notes informing participants not to return to previous pages using the browser's "back" button because they could lose their answers by doing so. This prevented participants from going back to previous pages when completing the vocabulary posttest. As to the annotations, participants had to click a link and to reveal an annotation in the margin. Audio clips that were included in the annotation were played automatically when the participants clicked the link; participants could re-click the link if they wished to hear the clip again.

Once they completed all of the online exercises (including the reading comprehension tasks, vocabulary tests, and the questionnaire), participants were shown a concluding page that thanked them for their participation and asked them not to reveal the details of the program to their classmates. They were also told that they receive their scores upon request.

### 6.1.6 Data collection, coding, and analysis procedures

This section outlines the data collection, coding, and analysis procedures connected with each of the experiment's three purposes, namely comparing the effects of three types of annotations on incidental vocabulary acquisition in L2

Mandarin Chinese, identifying the amount of word knowledge gain in this experiment, and comparing participants' attitudes to different annotations.

The vocabulary test scores were first organised according to the three types of annotations in this study and coded using the partial knowledge-sensitive criteria. To achieve the experiment's first objective, which relates to the differences between annotation types, total scores for all the vocabulary posttests and scores for each vocabulary posttest were compared to identify possible differences in acquiring varying types of knowledge. The independent variable was the annotation type (namely text + Pinyin, text + audio, or text + Pinyin + audio), while the dependent variables were the vocabulary test scores.

In most cases this experiment's data were not normally distributed and thus required a non-parametric statistical test for comparing scores among the annotation groups. Considering that more than two annotation types were being compared and that a crossover research design (which also qualified as a type of within subject design as all of the participants encountered all of the annotations) was being employed, a *Friedman's two-way analysis of variance* (related samples) was selected. Where a post hoc test was needed, Wilcoxon signed-rank tests were employed with the Bonferroni correction applied (as explained in Section 5.1.4).

In relation to the attitude question, unlike in the previous experiment, only three post hoc tests were required at this stage: Test 1 compared word scores from a) text + Pinyin annotation and b) text + audio annotation; Test 2 compared word scores from a) text + Pinyin annotation and b) text + Pinyin + audio annotation; and Test 3 compared words scores from a) text + audio annotation and b) text + Pinyin + audio annotation. Instead of utilising 0.05 as the critical value for the significance of each test, 0.0167 (i.e. 0.05 divided by three, the number of tests conducted) was used.

Regarding the amount of knowledge gain (which relates to the second sub-question of research question 1 mentioned in Section 4.1), the mean scores of the vocabulary tests and percentage of correct answers were determined with descriptive

statistics. To be noted, participants' knowledge gain in this experiment was measured by the differences between the vocabulary pre- and posttests scores (as calculated by subtracting the former from the latter). This method was suggested by the findings of the first experiment, as discussed in Section 5.1.3. The results for the second experiment presented in this chapter thus reflect correct answer percentages that were calculated based on such score differences (although no points were deducted from the final score for the production test, as no production test was used when testing participants' pre-knowledge of the target words). The score differences were then divided by the maximum scores (i.e. 15 points, or one point maximum for each target word) for corresponding tests to yield acquisition rates that indicate the overall amount of word knowledge gain demonstrated by that test.

It should be noted that the criteria that are sensitive to partial knowledge employed in this study could be used to mark the tests that require word meaning, Pinyin form, and character form to be produced. However, the in-depth analyses of the amount of word knowledge gain only present details of partial knowledge gain in the Pinyin form of the target words in two related tests, namely the meaning-based and character-based Pinyin form production tests. Details for the recognition tests (which asked participants to provide word meaning based on a target word's character or spoken form) are excluded due to the very small number of partially correct answers provided by participants. The raw data showed that only 26 answers among three participants qualified for being awarded half a point (the two recognition questions required a total of 1440 answers from the 48 participants for all 15 target words). As it is not reliable to draw conclusions from such limited data, no related in-depth analysis was conducted. While the test requires participants to produce word character forms using the Pinyin input on the computer; as partial character forms cannot be produced in this process (see Section 4.2.3), related data were not analysed in this section.

With respect to the third research question, attitudes towards different types of annotations, participants' responses to the five-point Likert-scale question were

collected and compared across the three relevant annotation types. The independent variable was the type of annotation and the independent variables were participants' marks. As the data were non-normally distributed, a non-parametric test was needed to conduct the statistical analysis. Given that more than two types of annotations were being compared and that a within subject design was being utilised (i.e. all participants indicated their attitudes to all three types of annotations), a *Friedman's two-way analysis of variance* (related samples) with Wilcoxon signed-rank tests as the post hoc tests was deemed appropriate to use. Bonferroni correction was applied (significance level of $p < 0.0167$).

In addition, answers from the questionnaire relating to participants' language learning backgrounds and the number of times they looked at each annotation were also analysed. This supplementary information was helpful inter alia for the selection of participants. And the data from the log file were collected to confirm the number of times the participants looked at each annotation.

## 6.2 Results of the second experiment

This section presents the results for all three of this study's research questions. Comparisons between the effects of the three types of annotations involved in the second experiment are provided first, followed by the amount of word knowledge gain and the participants' attitudes towards different types of information used in the annotations.

### 6.2.1 The effects of different sound-related annotations

Research question 1 considered whether incidental knowledge learning is affected by various types of annotations, namely text + Pinyin, text + audio, and text + Pinyin + audio. The participants' scores on target words in the vocabulary tests were gathered, analysed according to the different types of annotations provided, and compared using statistical tests chosen according to the distribution and homogeneity of related data.

The total scores achieved through each type of annotation were compared first; thereafter scores for each part of the vocabulary posttest were compared to investigate the learning results in a more detailed manner (i.e. by looking at different types of word knowledge). However, as discussed in Section 4.2.6 the results may be affected by the non-parametric statistical tests (that is, the results may falsely indicate that no differences exist between the groups). An in-depth analysis across the three annotation types was thus conducted to compare the vocabulary posttest scores and help explain the results of the statistical tests.

In accordance with the statistical methods mentioned in Section 6.1.6, the data collected for each annotation group were tested using the Shapiro-Wilk test (for the status of normal distribution) and Levene's test (for the status of homogeneity). The results suggested the application of *Friedman's two-way analysis of variance* (related samples) to compare the results of all of the vocabulary tests, as mentioned above. The results of the Friedman's test suggested that no significant difference exists among the three types of annotations (i.e. text + Pinyin, text + audio, and text + Pinyin + audio) in the scores for the vocabulary posttest ($\chi2$ (2) = 3.884, p = 0.143), the spoken form recognition test ($\chi2$ (2) = 4.467, p = 0.107), or the character-based Pinyin form production test ($\chi2$ (2) = 28.00, p = 0.242). The same is true for the scores of both the meaning-based Pinyin form production test ($\chi2$ (2) = 5.545, p = 0.062) and the character recognition test ($\chi2$ (2) = 5.695, p = 0.058), although these two tests indicated a marginal effect of annotation type on knowledge gain as the P-value is close to the critical value of 0.05.

The only significance found at this stage was in the scores for the Pinyin input assisted meaning-based character production test ($\chi2$ (2) = 6.091, p = 0.048). However, the results of the post hoc analysis conducting using the Wilcoxon signed-rank tests with a Bonferroni correction (and significance level of $p < 0.0167$) suggested that no significant differences existed between text + Pinyin annotation and text + audio annotation (Z = - 0.095, p = 0.925), Pinyin annotation and text + Pinyin + audio

annotation (Z = - 0.625, p = 0.532), or text + audio annotation and text + Pinyin + audio annotation (Z = - 1.612, p = 0.107).

To summarise, no statistically significant differences were found between the scores of each part of the posttest across the three annotation types. This means that participants who received various types of annotations obtained a similar amount of word knowledge in the five types of vocabulary posttests. In other words, the statistical tests indicate that type of sound information provided in the annotations seems to be irrelevant to the results of incidental vocabulary learning in L2 Mandarin Chinese. However, the in-depth descriptive analyses of the results suggest otherwise.

Each participant's mean scores in each part of the posttest across the different annotation types are presented in Table 6.4 below, with the highest score for each test in bold and the second highest underscored. It should be recalled that the highest possible mean score for each annotation type in each test was five points.

Table 6.4 Mean participant scores for the vocabulary posttests across the three annotation types

| Vocabulary posttest | Text + Pinyin | Text + audio | Text + Pinyin + audio |
|---|---|---|---|
| Meaning-based Pinyin form production (part1) | 0.25 | <u>0.32</u> | **0.43** |
| Pinyin input assisted meaning-based character production (part 2) | <u>1.69</u> | 1.58 | **1.92** |
| Spoken form recognition (part 3) | 0.83 | <u>0.92</u> | **1.17** |
| Character-based Pinyin form production (part 4a) | **0.67** | 0.53 | <u>0.65</u> |
| Character form recognition (part 4b) | <u>1.08</u> | 0.88 | **1.15** |
| Total Mean | <u>4.52</u> | 4.23 | **5.32** |

It is noticeable that the mean total score of the text + Pinyin + audio annotation group (5.32) is higher than for the text + Pinyin group (4.52) and the text + audio group (4.23). In addition, the means scores for the text + Pinyin + audio group are also higher in all of this experiment's vocabulary posttests with the exception of the character-based Pinyin form production test. This result may suggest that in this study

the text + Pinyin + audio annotation has a better effect than the other two annotation types in acquiring more types of word knowledge. The exception noted could be due to redundant information, as it is not necessary to provide much sound-related information for a character form recognition test.

Moreover, the scores for the character form-related test increased with the help of the text + Pinyin annotations, while the higher scores for the sound-related word knowledge tests were connected to the text + audio annotations. This result indicates that audio annotation is more helpful than Pinyin annotation in terms of acquiring both the spoken and Pinyin forms of the target words in this study.

Therefore, unlike the Friedman's test results, the in-depth analysis indicates that annotation type may influence incidental vocabulary learning. When the results of the statistical tests are being interpreted, it should thus be borne in mind that the text + Pinyin + audio annotation may be the most helpful type of annotation as it assists many – although not all – types of vocabulary knowledge (details are presented in Section 6.2.2).

### 6.2.2 Amounts of incidental word knowledge gain

The first experiment indicated that some types of word knowledge relating to the sound of a word can be acquired incidentally, to varying degrees, through reading in L2 Mandarin Chinese; the results of the second experiment support this finding. The amounts of related word knowledge gain in the context of online reading are presented in this section, as the percentages of correct answers for each type of vocabulary posttest. Taking the possible word effect into consideration (i.e. the suggestion from the first experiment's results that some target words are easier than others), this section also provides detailed knowledge gain on a word-by-word basis to identify whether the ease of certain words still exists. Finally, to provide more detailed information on incidental vocabulary acquisition, the amount of partial knowledge gain is generalised (which is possible given that this study employs partial

knowledge-sensitive criteria).

First of all, the mean percentages of the correct answers of all 15 target words for each part of the posttest demonstrate the amount of word knowledge gain, as shown in Table 6.5 below. The mean percentages, which range from 10.87% to 19.40% for each part of the posttest, clearly demonstrate that participants were able to acquire relevant types of word knowledge through incidental learning in L2 Mandarin Chinese reading. This word knowledge includes character form; sound-related (i.e. Pinyin and spoken) form; and links between character and spoken form, character and word meaning, and spoken form and word meaning.

Table 6.5 Mean percentage for the correct answers in the vocabulary posttests

| Vocabulary posttest | Correct answers (%) |
|---|---|
| Meaning-based Pinyin production | 10.87 |
| Pinyin input assisted meaning-based character production | 12.20 |
| Spoken form recognition | 17.40 |
| Character recognition | 19.40 |
| Character-based Pinyin production | 10.67 |

The mean score percentages on the recognition tests for Pinyin and character forms are close, which suggests a similar amount of knowledge gain in receptive knowledge related to word sound and character form. This indicates that providing sound or related information in the reading material enables learners to obtain sound-related word knowledge through reading; more importantly, they may receive similar amounts of sound-related and character form-related knowledge. As such, instead of a just serving as resource for acquiring word meaning, reading could be more helpful in L2 vocabulary learning than often expected.

It is also noticeable that the participants obtained higher scores on both of the recognition tests than on the production test, which suggests that the former were less difficult – or, as widely acknowledged, that receptive knowledge is easier to obtain than productive knowledge. Given the idea that different types of vocabulary tests

measure different types of word knowledge (which was discussed in previous chapters), the reason for the ease of acquiring receptive knowledge could be further explored from the perspective of knowledge type and amounts thereof measured by different tests (see Section 6.1).

Now that the general amount of word knowledge gain demonstrated by the percentages of correct answers for each vocabulary posttest is known, this section demonstrates the knowledge gain on a word-by-word basis to identify whether the word effect exists. The main reason for using a word-by-word analysis was that the scores for words "1 丝绸(sīchóu) [silk]" and "3 耐克 (Nàikè) [Nike]" were much higher than for the other words involved in the first experiment. In the second experiment, it was ensured that these two words appeared only once throughout the reading and were not required for the reading comprehension questions. With possible influential factors (i.e. high frequency of encountering the target words and high task-involvement loads) being controlled in this experiment (see Section 5.3.2), it is interesting to see whether both words still have scores that are higher than those for the other words.

Each word had a total maximum score of five points for the whole vocabulary posttest (with one point for each of the posttest's parts). Table 6.6 presents the mean scores for each part of the vocabulary posttest on a word-by-word basis as well as the mean total score for the entire vocabulary posttest. The three highest mean scores for each test are marked in bold for each part of the posttest.

Table 6.6 Mean scores for each target word in the vocabulary posttests

| Target word | Meaning-based Pinyin form production test | Pinyin input assisted meaning-based character production test | Character form recognition test | Spoken form recognition test | Character-based Pinyin form | Mean total |
|---|---|---|---|---|---|---|
| 1 丝绸(sīchóu) [silk] | **0.24** | **0.31** | **0.33** | **0.44** | **0.25** | **1.58** |
| 2 价值(jiàzhí) [value] | **0.13** | 0.06 | 0.19 | 0.19 | 0.13 | 0.7 |
| 3 耐克 (Nàikè) [Nike] | **0.40** | **0.44** | **0.75** | **0.58** | **0.27** | **2.44** |
| 4 判断(pànduàn) [judge] | **0.13** | 0.08 | 0.04 | 0.00 | 0.03 | 0.28 |
| 5 折扣(zhékòu) [discount] | 0.09 | **0.13** | 0.17 | **0.38** | 0.14 | 0.89 |
| 6 减缓 (jiǎnhuǎn) [slowdown] | 0.00 | 0.02 | 0.02 | 0.02 | 0.01 | 0.07 |
| 7 眩晕(xuànyūn) [dizziness] | 0.02 | 0.06 | 0.08 | 0.15 | 0.02 | 0.33 |
| 8 尽量(jìnliàng) [to the best of one's ability] | 0.03 | 0.04 | 0.10 | 0.04 | 0.03 | 0.25 |
| 9 症状(zhèngzhuàng) [symptom] | 0.08 | 0.08 | 0.06 | 0.08 | 0.07 | 0.38 |
| 10 稍微(shāo wēi) [slightly] | 0.07 | 0.08 | 0.00 | 0.02 | 0.10 | 0.27 |
| 11 庄稼(zhuāngjia) [crops] | 0.07 | 0.06 | 0.10 | 0.04 | **0.16** | 0.44 |
| 12 辛苦(xīnkǔ) [to work hard; laborious] | 0.10 | 0.10 | 0.15 | 0.23 | 0.15 | **0.73** |
| 13 培养(péiyǎng) [cultivate] | 0.09 | **0.13** | 0.17 | 0.15 | 0.05 | 0.58 |
| 14 保持(bǎochí) [to keep; to maintain] | 0.07 | 0.08 | 0.00 | 0.06 | 0.06 | 0.28 |
| 15 懒惰(lǎnduò) [lazy] | 0.11 | 0.08 | **0.21** | 0.13 | 0.06 | 0.59 |

This table clearly demonstrates that despite increasing the number of target words from five in the first experiment to 15 in this one, the target words “3 耐克 (Nàikè) [Nike]” and “1 丝绸 (sīchóu) [silk]” still gained the two highest total scores (2.44 and 1.58, respectively); and the next highest score (0.89), which was achieved for the word “折扣 (zhé kòu) [discount]”, was much lower. In addition, these top two words also achieved the highest scores in each part of the vocabulary test. Such a result obviously suggests the ease of both words and therefore calls for further discussion (which is presented in the next chapter).

The percentages of correct answers for each part of the target words’ Pinyin syllables are presented next. Table 6.7 provides information on partial knowledge gain in the meaning-based Pinyin form production test, whereas Table 6.8 presents this information for the character-based Pinyin form production test. Both tables in the following pages use a word-by-word format to help to identify whether a word effect exists.

Table 6.7 Percentage of correct answers for each part of a word's Pinyin (posttest: meaning-based Pinyin form production test)

| Target word (with serial number) | Percentage of correct answers for each part of the Pinyin for the first character | | | Percentage of correct answers for each part of the Pinyin for the second character | | |
|---|---|---|---|---|---|---|
| | Initial | Final | Tone | Initial | Final | Tone |
| 1 丝绸(sīchóu) [silk] | 41.67 | 37.5 | 16.67 | 25 | 20.83 | 4.17 |
| 2 价值(jiàzhí) [value] | 20.83 | 16.67 | 4.17 | 16.67 | 16.67 | 4.17 |
| 3 耐克(Nàikè) [Nike] | 62.5 | 41.67 | 8.33 | 62.5 | 58.33 | 25 |
| 4 判断(pànduàn) [judge] | 20.83 | 16.67 | 4.17 | 16.67 | 12.5 | 8.33 |
| 5 折扣(zhékòu) [discount] | 8.33 | 8.33 | 0 | 12.5 | 12.5 | 8.33 |
| 6 减缓(jiǎnhuǎn) [slowdown] | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 眩晕(xuànyūn) [dizziness] | 0 | 0 | 0 | 4.17 | 4.17 | 4.17 |
| 8 尽量(jìnliàng) [to the best of one's ability] | 4.17 | 4.17 | 0 | 4.17 | 4.17 | 0 |
| 9 症状(zhèngzhuàng) [symptom] | 12.5 | 12.5 | 4.17 | 8.33 | 8.33 | 0 |
| 10 稍微(shāo wēi) [slightly] | 8.33 | 8.33 | 4.17 | 8.33 | 8.33 | 4.17 |
| 11 庄稼(zhuāngjia) [crops] | 8.33 | 8.33 | 4.17 | 8.33 | 8.33 | 4.17 |
| 12 辛苦(xīnkǔ) [to work hard; laborious] | 12.5 | 12.5 | 4.17 | 12.5 | 12.5 | 4.17 |
| 13 培养(péiyǎng) [cultivate] | 12.5 | 12.5 | 0 | 12.5 | 12.5 | 4.17 |
| 14 保持(bǎochí) [to keep; to maintain] | 8.33 | 8.33 | 4.17 | 8.33 | 8.33 | 4.17 |
| 15 懒惰(lǎnduò) [lazy] | 20.83 | 20.83 | 4.17 | 8.33 | 8.33 | 4.17 |
| Mean for all 15 words | 16.11 | 13.89 | 3.89 | 13.89 | 13.05 | 5.28 |

Table 6.8 Mean percentage of correct answers for each part of a word's Pinyin (posttest: character-based Pinyin form production test)

| Target word (with serial number) | Percentage of correct answers for each part of the Pinyin for the first character | | | Percentage of correct answers for each part of the Pinyin for the second character | | |
|---|---|---|---|---|---|---|
| | Initial | Final | Tone | Initial | Final | Tone |
| 1 丝绸(sīchóu) [silk] | 41.67 | 41.67 | 25 | 29.17 | 29.17 | 8.33 |
| 2 价值(jiàzhí) [value] | 33.33 | 20.83 | 8.33 | 33.33 | 25 | 0 |
| 3 耐克(Nàikè) [Nike] | 41.67 | 41.67 | 4.17 | 50 | 50 | 12.5 |
| 4 判断(pànduàn) [judge] | 8.33 | 8.33 | 4.17 | 8.33 | 8.33 | 4.17 |
| 5 折扣(zhékòu) [discount] | 20.83 | 20.83 | 8.33 | 16.67 | 12.5 | 4.17 |
| 6 减缓(jiǎnhuǎn) [slowdown] | 4.17 | 4.17 | 0 | 4.17 | 4.17 | 0 |
| 7 眩晕(xuànyūn) [dizziness] | 0 | 0 | 0 | 4.17 | 4.17 | 4.17 |
| 8 尽量(jìnliàng) [to the best of one's ability] | 4.17 | 4.17 | 0 | 8.33 | 8.33 | 8.33 |
| 9 症状(zhèngzhuàng) [symptom] | 16.67 | 12.5 | 8.33 | 8.33 | 8.33 | 0 |
| 10 稍微(shāo wēi) [slightly] | 12.5 | 12.5 | 8.33 | 8.33 | 8.33 | 8.33 |
| 11 庄稼(zhuāngjia) [crops] | 16.67 | 12.5 | 4.17 | 25 | 25 | 12.5 |
| 12 辛苦(xīnkǔ) [to work hard; laborious] | 25 | 25 | 12.5 | 25 | 25 | 4.17 |
| 13 培养(péiyǎng) [cultivate] | 8.33 | 8.33 | 0 | 8.33 | 8.33 | 4.17 |
| 14 保持(bǎochí) [to keep; to maintain] | 16.67 | 16.67 | 4.17 | 12.5 | 12.5 | 4.17 |
| 15 懒惰(lǎnduò) [lazy] | 12.5 | 12.5 | 4.17 | 4.17 | 4.17 | 0 |
| Mean for all 15 words | 17.50 | 16.11 | 6.11 | 16.39 | 15.56 | 5.00 |

Generally speaking, the mean percentages of the correct answers for each part of the Pinyin forms of both characters in the target words range from 3.89% to 16.11% in the meaning-based Pinyin from production test and from 5% to 17.50% in the character-based Pinyin form production test. However, the amount of knowledge gain for each part of the Pinyin ranges largely. Factors that may help to explain these differences are explored below.

Firstly, a strong tie seems to exist between the initial and final parts of a character's Pinyin, especially in the Pinyin forms for which participants provided relatively few correct answers. For example, in the meaning-based Pinyin form production test (Table 6.7), participants had no correct answers for the initial, final, and tone parts of "6 减缓 (jiǎnhuǎn) [slowdown]"; the same is true for the first character in "7 眩晕 (xuànyūn) [dizziness]", although the three parts of the Pinyin of the word's second character all had a production rate of 4.17%. In the character-based Pinyin form production test (Table 6.8), the production rates for all three parts of the Pinyin syllables of the word "7 眩晕 (xuànyūn) [dizziness]" are the same as in the first test. Moreover, the production rates of the three parts of the Pinyin of the second character in "8 尽量 (jìnliàng) [to the best of one's ability]" and "10 稍微 (shāowéi) [slightly]" are equal (8.33%).

More specifically, in the meaning-based Pinyin form production test (Table 6.7), the correct answer percentages for the initial part of the first and second characters are the same in 11 of the 15 target words; the same is true for 12 of the target words' second characters. In contrast, for the question that asked participants to write the Pinyin of a word according to its character form (Table 6.8), the correct answer percentages are equal for the initial and final parts of the first character in 13 target words; the same holds true for the second character in 13 target words.

However, these effects disappear when the participants demonstrated more knowledge gain on the target words' Pinyin form. The word "1 丝绸 (sīchóu) [silk]" can be used as an example. The correct answer percentages for the initial and final parts of the first character were respectively 41.67% and 37.50% (see Table 6.7). A parallel

situation exists in the results for the word's second character, although the percentages of correct answers for the initial part (25.00%) and final part (20.83%) were lower than for the first character. It is also noteworthy that among all of the characters that achieved different percentages of correct answers for their Initial and final parts, the initial parts had higher correct answer percentages than the final parts – which suggests it is particularly difficult to remember these characters' final parts.

Regardless of whether the final part is harder than the initial part, the tone is clearly the most difficult part of almost all of the words in this experiment. It was also noted that the correct answer percentages for some words are much higher than others, with "1 丝绸 (sīchóu) [silk]" and "3 耐克 (Nàikè) [Nike]" being the top two. At the same time, some words have percentages that are much lower than other words; for example, no participant was able to produce even a partially correct answer for "6 减缓 (jiǎnhuǎn) [slowdown]", as shown in Table 6.7.

One additional finding needs to be addressed here, namely that 11 of the 15 target words have correct answer percentages for Pinyin in the meaning-based Pinyin form production test that exceed, or are equal to, what they achieved in the character-based Pinyin form production test. This finding suggests that the link between the target words' Pinyin and character forms has been better established than the link between the Pinyin form and word meaning.

### 6.2.3 Participants' attitudes towards differing types of annotations

In this section the findings in relation to the last research question are presented. The aim was to gauge the participants' attitudes towards the information provided in annotations to assist L2 Mandarin Chinese vocabulary learning in the context of reading, to reiterate: word meaning, Pinyin and audio (spoken form of target words). The mean scores and standard deviation of each type of annotation are presented in Table 6.9 below, together with detailed information on the percentages of the participants' responses on a 5-point Likert scale, in which one represents strongly disagree and five represents strongly agree.

Table 6.9 Participants' views on different types of information provided in annotation

| Annotation | Percentage of participants' choices of each response | | | | | Mean | Std Deviation |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | | |
| Meaning | 0.00 | 0.00 | 15.40 | 11.46 | 73.10 | 4.58 | 0.75 |
| Pinyin | 0.00 | 11.54 | 19.23 | 38.46 | 26.92 | 3.69 | 1.21 |
| Audio | 3.80 | 26.92 | 19.23 | 23.08 | 26.92 | 3.42 | 1.26 |

n = 25. 1 = strongly disagree, 2 = disagree, 3 = normal, 4 = agree, 5 = strongly agree

A "related samples Friedman's two-way analysis of variance" test indicated significant differences between the three types of information provided in annotations ($\chi 2$ (2) = 26.205, p = 0.000). Post-hoc analysis with a Wilcoxon signed-rank test was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < 0.0167$ (See Section 5.1.4 for a more detailed explanation). A significant difference was seen between the following pairs of annotations: word meaning and Pinyin (Z = -3.917, p = 0.00), word meaning and audio (Z = -4.303, p = 0.00).

Table 6.9 shows that the meaning of a word, with a high mean score of up to 4.58 and a relatively low standard deviation of 0.75 is considered to be the most valuable information in annotations, compared with the sound-related information of a word provided in the form of Pinyin or audio. As for the sound information, both the Pinyin and the audio turned out to be much less useful than word meaning, according to the participants, with a mean score of 3.69 for Pinyin and 3.42 for audio annotation. This result suggests that most of the participants agreed on the usefulness of word meanings in annotations in terms of assisting incidental vocabulary learning through reading.

That concludes the presentation of the results of the second experiment. Before moving on to the next chapter, the important findings of this experiment will be recapped briefly. Firstly, clear knowledge gains result from many types of sound-related word knowledge, including that of the Pinyin form of a word, the links between the Pinyin form and word meaning, and the link between the Pinyin and character form of a word. Secondly, the text + Pinyin + audio annotation seems to be the most effective

annotation in terms of helping the participants to achieve the highest scores in the vocabulary posttest, although no statistically significant difference was found between this type of annotation and the Pinyin or audio annotation. Thirdly, the amount of knowledge gain demonstrated by the mean percentage of the correct answers for each part of the vocabulary posttest was 10.87% for the meaning-based Pinyin form production test, 10.67% for the character-based Pinyin form production testing, 12.2% for meaning-based character production test, 17.4% for the spoken form recognition test and 19.4% for the character form recognition test. Notably, the results of the last two tests were close, although various types of knowledge of word form were tested. Moreover, in this experiment the first one was repeated and it was found that some target words were much easier, while others could be more difficult. For example, "1 丝绸 (sīchóu) [silk]" and "3 耐克 (Nàikè) [Nike]" outperformed all the other target words in most parts of the vocabulary posttest, while "6 减缓 (jiǎnhuǎn) [slowdown]" seemed to be the word that got the lowest scores in many tests. Lastly, the results of the attitude question suggested that the participants felt that the meaning of the word was the most useful information in annotation and a significant difference in attitude was found between word meaning and Pinyin ($Z = -3.917$, $p = 0.00$), as well as between word meaning and audio ($Z = -4.303$, $p = 0.00$).

### 6.2.4 summary of findings

Before moving on to the next chapter (where the above results are further discussed), a brief summary is needed to recapitulate the most important findings of this experiment. First, knowledge gain is clearly achieved in relation to many types of sound-related word knowledge, including a word's Pinyin form, the links between the Pinyin form and word meaning, and the link between a word's Pinyin and character forms. Second, the text + Pinyin + audio annotation seems to be the most effective annotation in terms of helping participants to achieve the highest scores in the vocabulary posttest, although no statistically significant difference was found between this type of annotation and the Pinyin or audio annotations. Third, the amount of

knowledge gain (as measured by the mean percentage of correct answers for each part of the vocabulary posttest) is as follows: 10.87% for the meaning-based Pinyin form production test, 10.67% for the character-based Pinyin form production testing, 12.2% for the meaning-based character production test, 17.4% for the spoken form recognition test, and 19.4% for the character form recognition test. Notably, the last two tests have similar results, although they test different types of word form knowledge. Fourth, this experiment, similar to the first one, found that some target words are either much easier or more difficult than others, as "1 丝绸 (sīchóu) [silk]" and "3 耐克 (Nàikè) [Nike]" outperformed all other target words in most parts of the vocabulary posttest while "6 减缓 (jiǎnhuǎn) [slowdown]" had the lowest scores in many tests. Fifth, the results for the attitude question suggest that participants feel that word meaning is the most useful information to receive in annotations and that it was more helpful than both Pinyin and audio information.

# Chapter 7: General discussion

This chapter provides a detailed discussion of key findings from both the first and second experiments (as presented respectively in Chapters 5 and 6), with reference to the study's research questions. The results of the research are also discussed in relation to previous studies. The chapter starts by exploring the incidental acquisition of word knowledge, especially sound-related knowledge through L2 Mandarin Chinese reading. Thereafter the differences – although not statistically significant – observed among different types of annotations (namely text + Pinyin, text + audio, and text + Pinyin + audio) are explained. Findings pertaining to learners' difficulties with the final part in the Pinyin and reasons for certain target words being easier than others are examined next, followed by a comparison of the results of the vocabulary posttests. Finally, participants' attitudes towards different types of annotations are explored and some useful information revealed through the questionnaire and log file is highlighted.

## 7.1 Understanding the incremental nature of incidental vocabulary acquisition

This section discusses findings related to the first research question (which addresses the effects of sound-related information provided by the Pinyin and audio annotations in L2 Mandarin learning) and its two sub-questions (which explore whether knowledge related to the sound of word can be acquired incidentally and the amount of different types of word knowledge gained, as measured through the various vocabulary tests used in this study). Exploring these issues helps to clarify the incremental nature of incidental vocabulary acquisition.

As discussed in Section 2.2.1.1, Nagy et al.'s (1985) incidental vocabulary learning hypothesis asserts that the core of incidental vocabulary learning concerns incidentally acquiring a small amount of knowledge related to word meaning through reading (with such acquisition being possible even through a single encounter with a target word). The results of the current study's second experiment confirm the hypothesis by

demonstrating in the context of an under-researched L2 (i.e. Mandarin Chinese) that learners can acquire various types of word knowledge incidentally through reading when encountering words once or twice. Moreover, the fact that these results were achieved with a sample composed of L2 Mandarin Chinese learners at a beginner's level also gives rise to new topics for future incidental vocabulary acquisition research. The sound-related word knowledge that this study's participants obtained concerns the Pinyin form (both reception and production) and the links between Pinyin form and word meaning, Pinyin form and character form, and spoken form and word meaning. As this study has shown it is possible to acquire many types of word knowledge at the same time and that the word knowledge acquired in L2 Mandarin Chinese is not only receptive but also productive, it is reasonable to believe that previous studies (which focused primarily on testing word meaning) underestimated the effects of incidental vocabulary learning in several ways. As such, reading can be considered a more useful resource for L2 learning than normally expected, as long as it includes appropriate annotations.

Regarding the "small amount" of knowledge gain mentioned above, the second experiment's finding of a 10.87% – 19.40% knowledge gain (see Section 6.2) reveals the incremental nature of incidental vocabulary learning as proposed by Nagy et al. (1985). Previous researchers have already made a great effort to identify how much knowledge can be acquired through incidental learning. However, as discussed in Chapter 2, knowledge gain in incidental vocabulary learning has been reported under many different conditions and largely varies from study to study, which makes the related learning outcomes found incomparable. Nonetheless, it appears that the knowledge gain reported tends to be closer when studies use tests that measure the same type of knowledge. For example, in their first and second studies Chun and Plass (1996) found a 24.1% - 26.5% correct rate for a recognition translation test (L2 to L1). Likewise, in a study conducted by Waring and Takaki (2003), the results showed a mean score for the translation test of 4.6 points (18.4%). Based on these findings, a word knowledge gain of approximately 20% can be expected in a translation test (L2 to L1),

regardless of L2 type. In other words, receptive knowledge of both word form and the form-meaning link increases 20% via incidental vocabulary learning through reading.

However, the percentage of increase in receptive knowledge of word form and meaning stated in Chun and Plass's (1996) study (i.e. 24.1% – 26.5%) slightly exceeds the percentage found in the current research (i.e. 19.4%). The variation could be caused by the different annotation types utilised in the two studies (in particular, the picture annotation employed in Chun and Plass's study). Another possibility is the slightly different methods used to calculate knowledge gain: in the current study's second experiment, the amount of word knowledge gain equals the difference between the scores of the vocabulary pre- and posttests, whereas Chun and Plass (1996) calculated this rate based on the final score of only the posttest. Moreover, the different target languages may also account for the variation (in Chun and Plass's study the target language was German). Finally, the amount of word knowledge gain may also have been affected by the target words involved in the research; more details concerning this issue are provided in Section 6.3.

When interpreting results related to the amount of word knowledge gain in incidental vocabulary acquisition research, it should also be borne in mind that adopting different types of vocabulary tests illustrates varying types of knowledge gain. More importantly, one vocabulary test may measure several types of vocabulary knowledge. For example, completing a meaning-based Pinyin production test requires learners to have knowledge on the Pinyin form and to be able to link this form with its meaning. Instead of simply labelling the knowledge tested as productive knowledge, for clarification purposes the current study distinguished it as knowledge of both the Pinyin form and the link between Pinyin form and word meaning. As little interest in either this issue or the criteria for vocabulary tests has been demonstrated in the literature, the question of how to measure the results of incidental vocabulary acquisition seems to require further research. It is expected that if such research uses proper evaluation tools, its results will contribute to explaining how L2 Chinese Mandarin learners gradually master a word incidentally through reading.

As for improving the evaluation tools, the marking criteria employed in the current study serve as a good example as using criteria sensitive to partial knowledge enabled the word knowledge gained incidentally through reading to be further understood. Such criteria were also helpful for explaining certain findings. For example, they made it possible to report in the second experiment that the final part of Pinyin was more difficult than the initial part (see Section 6.2.2) and subsequently led to conducting further analysis to determine why (as discussed in Section 7.3).

Finally, connecting different types of vocabulary tests with various types of knowledge gain made it possible to further discuss issues concerning the amount of knowledge gain. Matters related to the similar amount of knowledge gain observed in both the spoken and character form recognition tests and the higher scores attained for the Pinyin input assisted meaning-based character form production test in comparison to the Pinyin form production test are discussed in Section 7.5.

## 7.2 The effects of different sound-related information provided in the annotations

This study's second research question was developed to investigate the effects of different types of annotations or, more precisely, the difference between various types of annotations with sound or sound-related information provided in Pinyin, audio, or both. As discussed in Section 2.2.3.6, annotation providing both Pinyin and audio would require information to be processed through both visual and aural channels, thus potentially leading to better learning results. This assumption, however, was not fully supported by the results of this research. Although the in-depth analysis suggests that the scores achieved for words with text with Pinyin and audio annotation are higher than those for the other two types of annotation in each part of the vocabulary post-test, the differences are not statistically significant. This section then attempts to explain this situation by revisiting the discussion of the CTML and features of Mandarin Chinese presented in Chapter 2.

The better effects of the text with Pinyin and audio annotation seem to be in

consistency with Mayer's CTML, as his theory involves the sensory modality channels (visual and aural) in addition to the original presentation mode channels (verbal and non-verbal). According to Mayer's CTML, the details regarding how each type of annotation presented in this study is processed are presented for comparative purposes. In the following pages, figures 7.1–7.3 illustrate how learners process the different forms provided by audio annotations, namely Pinyin, spoken, and a combination of the two forms. The solid arrows indicate that processing occurred, while the dotted arrows mean it did not.

Figure 7.1 The processing of words presented in Pinyin form (Pinyin annotation)

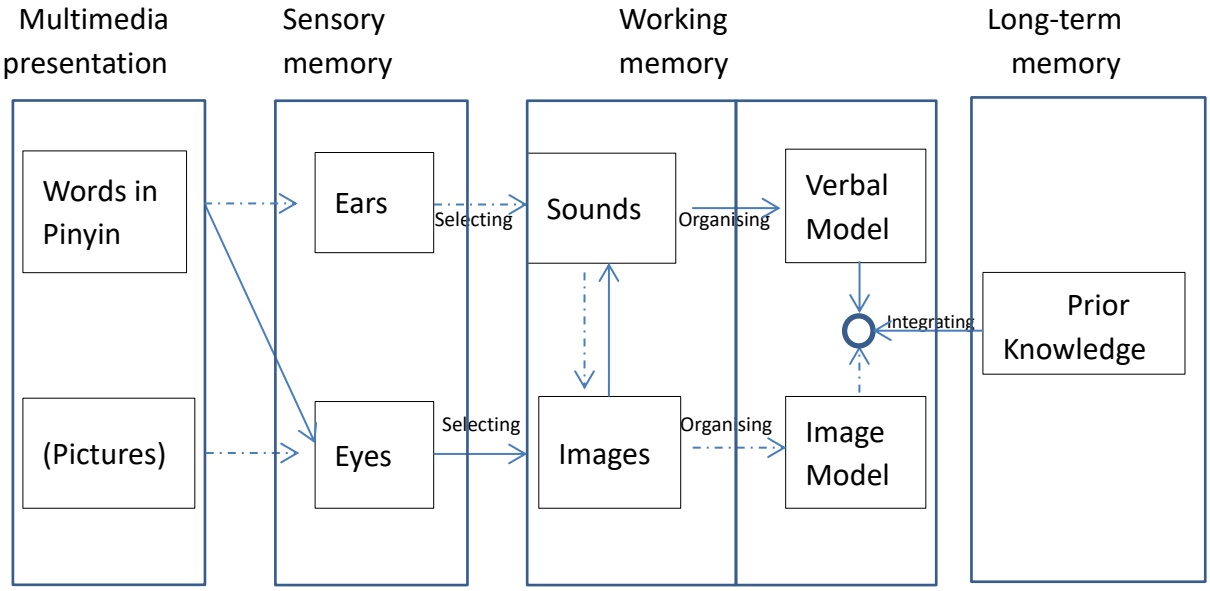



Figure 7.2 The processing of words presented in spoken form (audio annotation)

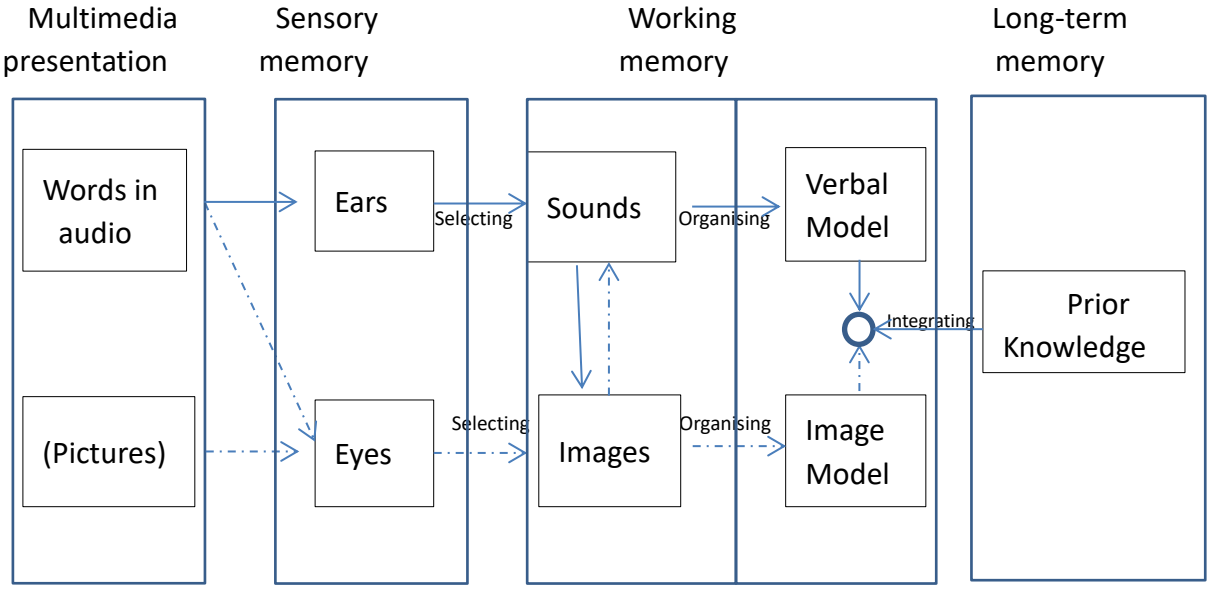
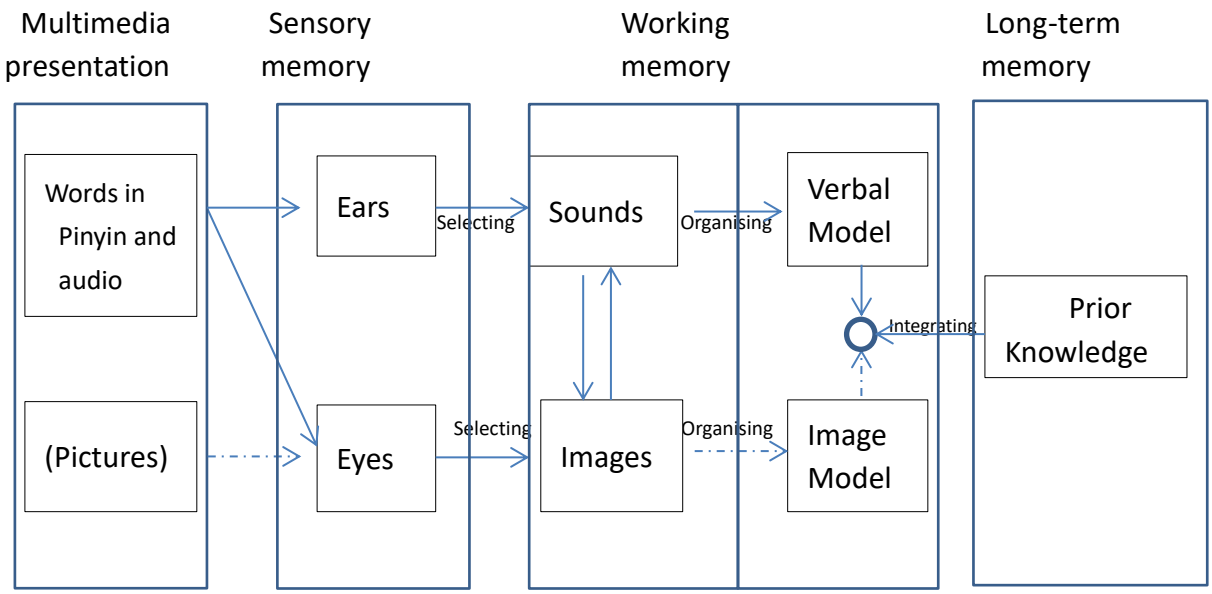


Figure 7.3 The processing of words presented in both Pinyin and spoken forms (Pinyin + audio annotation)




These figures reveal that Pinyin information entered through the visual channel and was transferred to the verbal channel in the working memory showed in the figures above, while the audio information entered through the visual line and kept moving through the verbal line in the working memory. As for the mixed information, both the visual and aural channels were used, although in both cases the information was transferred to the verbal channel, leaving the non-verbal channel empty.

To reiterate, the five critical cognitive processes are: selecting relevant words; selecting relevant images; organizing selected words; organizing selected images; and integrating word- and image-based representations. Apparently, none of the three types of annotation went through all the processes as described by Mayer in the CTML. However, it was observed that processing the Pinyin and audio annotation of a word together does involve four of the crucial processes described by the CTML, while processing the Pinyin or audio annotation is engaged in only two crucial processes. This could account for the better learning results of the text with Pinyin and audio annotation. Similarly, the analogous effects of the Pinyin and audio information could be caused by the fact that they underwent the same number of key processes. From this perspective, it is reasonable to believe that CTML also accounts for the vocabulary

learning assisted by multimedia annotation in the context of reading.

If this is the case, the insignificant differences might be caused by reasons including, for example, the non-parametric tests conducted (see Section 4.2.6 for discussion on the weakness of this statistical test). Beyond that, the automatic activation of phonological information may also come to play (see Section 3.3). If this is the case, the process in Figure 7.1 would be same as that in Figure 7.3, thus reducing the advantage of the multimodal sound information presented.

Alternatively, instead of employing CTML to analyse the better results of the text with Pinyin and audio annotation, it is possible that the better results can be attributed to the fact that this annotation meets the needs of different learners. Aural learners may prefer audio annotations more than visual learners. However, the text with Pinyin and audio annotation makes it easy for various types of learners to find the information they need. Although a question about participants' learning style was included in the questionnaire, the data collected are insufficient to draw any conclusion on this issue. Therefore, future studies should explore the relationship between the effects of different types of annotations and learners' learning style. Moreover, further research on the effects of annotation providing sound information in multimodality should be undertaken, with improvements made in order to control the factors mentioned above.

Finally, it should be noted that CTML was originally developed to explain the better comprehension and knowledge retention brought about by multimedia learning, but vocabulary learning specifically. The learning objectives of multimedia learning are often compromised by the use of complicated principles or phenomena (for instance the formation of lightning or migration of birds in Europe), rather than focusing on learning new words in an L2. As such, more research is required to explore the rationale underlying incidental vocabulary learning.

## 7.3 Difficulties related to learning the Pinyin form

This section addresses the findings from the in-depth analysis of the amount of word partial knowledge gain in the second experiment (Section 6.2.2). Firstly, it was

not surprising to see that tone is the most difficult part in the Pinyin form for the participants to remember. This difficulty was often observed by language teachers in L2 Mandarin Chinese teaching as well. The reason tone may be the most difficult part of a Pinyin syllable for L2 learners could be unfamiliarity with the form. In Chinese, each character is associated with a tone – which is not the case in English or other European languages. In a Pinyin syllable, the initial and the final parts are represented by Roman letters, which are not new for many learners. As such, the initial and final parts seem much easier for learners to cope with than the tones, which require greater effort on behalf of the learners and more help from language teachers and material developers.

Secondly, this study's second experiment also indicate that a Pinyin syllable's final part is more difficult than its initial part (see Section 6.2.2), although the reason is not very clear. The original answers of participants were examined to understand this phenomenon. Table 7.1 shows all of the mistakes found related to the 15 target words. As entering tone marks is difficult on the computer, the four tone marks were replaced by the numbers one to four (seeing as the tones are usually referred to as first, second, third, and fourth in teaching); related instructions were also provided within the questions themselves.

Table 7.1 Pinyin mistakes made by the participants in the second experiment

| Target word with serial number | Correct answer | Wrong answers in the meaning-based Pinyin form production test | Wrong answers in the character-based Pinyin form production test |
|---|---|---|---|
| 1 丝绸 (sīchóu) [silk] | si1 chou2 | su (chou)*; | |
| | | (si1) chao2 | |
| 2 价值 (jiàzhí) [value] | jia4 zhi2 | jin; | jie2 zhi1; |
| | | jie zhe; | (jia4) zhe1; |
| | | jie (zhi) | jie (zhi) |
| | | (jia) zhe | |
| 3 耐克 (Nàikè) [Nike] | nai4 ke4 | ni (ke); | dai(ke); |
| | | nai (ke); | ni(ke) |
| | | ni2 (ke3); | |
| | | ni1 (ke4) | |
| 4 判断 (pànduàn)[judge] | pan4 duan4 | (pan) ding; | |
| | | pai (duan) | |

| | | | |
|---|---|---|---|
| 5 折扣 (zhékòu)[discount] | zhe2 kou4 | | (Zhe2) ke4 |
| 11 庄稼 (zhuāngjia)[crops] | zhuang1 jia | zhuan2 (jia4) | |

* The initial and final parts of the Pinyin of the character within the bracket were correct in participants' answers.

The above mistakes can be categorised mainly into four types. The first category involves mistakes caused by a character's phonetic components. For example, while the character "价" in "2 价值 (jiàzhí) [value] has is a phonetic component "介 (jiè) [to introduce]", the component is not pronounced exactly the same as the whole character. The participants' answers reveal that mistakes of this nature are among the most major found in this study. The second category entails mistakes related to a word's English equivalent, which were particularly found in answers for "3 耐克 (Nàikè) [Nike]". As this is an English loan word, the English equivalent actually provides participants with a pronunciation tip; however, the similarity between the English word and its Pinyin form may have also caused many participants to ignore the small spelling difference between them. The third category involves mistakes related to dropping letter(s) of an entire Pinyin syllable (e.g. missing "g" in "11 庄稼 (zhuāngjia) [crops]") and the final "ou" (in "5 折扣 (zhékòu) [discount]" and "1 丝绸 (sīchóu) [silk]"). The fourth category entails typing mistakes, for instance, participants who gave the wrong answer of "su chou" for "1 丝绸 (sīchóu) [silk]". It is very likely that this mistake was just due to mistyping as the letters u and i are placed next to each other in a computer keyboard.

The typing error aside, the prevalence of mistakes made in the three categories described above merit attention from both language teachers and material developers. For instance, when teaching materials are being developed, it might be helpful to highlight the differences between the Pinyin for the character and its phonetic component, as well as that between the Pinyin form and the word's English equivalent.

If a CALL environment is involved, such differences could be easily highlighted using various colour, size, or even cartoons and audio annotations that involve related explanations.

## 7.4 The word effect

This section deals with the finding of both experiments concerning certain target words being easier than others for the participants in this study. Firstly, as the first experiment revealed a possibility that the higher scores for "1 丝绸 (sīchóu) [silk]" and "3 耐克 (Nàikè) [Nike]" were connected to these words' higher encounter frequencies, the second experiment limited their frequency to only one time. However, the scores for both words were still much higher than the scores for others, as shown in Table 6.6 in Chapter 6 – which indicates that the frequency of encounter does not seem to be responsible for the high acquisition rates. This section thus attempts to identify reasons that account for the better acquisition of these words, taking factors that affect studies on incidental vocabulary acquisition as discussed in Section 2.2.3.5 into consideration as well as the properties related to Mandarin Chinese words that were mentioned in Section 3.3.

Considering the large range in the amount of word knowledge gain reported in previous studies (see Sections 5.2 and 6.2), two reasons may account for the above phenomenon. Apart from the effects of vocabulary tests that were intensively discussed in earlier chapters, the other factor needs to be addressed is word difference. The current study observed a large gap on incidental vocabulary learning results for different target words. Both of the current study's experiments suggest a possible word effect for "1 丝绸 (sīchóu) [silk]" and "3 耐克 (Nàikè) [Nike]", and it is clear that the amount of knowledge gain would change significantly if these two words were not included. The ease of some words was also observed by Waring and Takaki (2003) with English as the L2. In their study, these researchers reported that "jurg/s" and "molden" were the highest rated words in the vocabulary tests, they together accounted for 65% of all of the correct answers given on the tests. Similarly, using another research set (i.e. with different frequency of encountering target words) they found an ease for "yoot", as it accounted for approximately 58% of the total score for the vocabulary test. The researchers therefore determined that they may have over-estimated the results for vocabulary learning and subsequently re-calculated the results for the remaining

words.

The current study's results are consistent with Waring and Takaki's above findings and reconfirm the ease of some words in the process of incidental learning through reading. It is reasonable to believe easy words might lead to high amount of knowledge gain. However, their solution of simply removing these words from final analysis is questionable. Following Waring and Takaki's logic, do the most difficult words then need to be removed from the final analysis in the current study as well? If yes, what standard should be followed to identify those words? More importantly, it is difficult to tell whether the existence of such words in a language represents an extreme situation, that is, only a few of them can be found in the vocabulary. If this is true, it would be correct to exclude these words from the final analysis. However, if it is not, they should be involved given that they may represent an important category of words that should be considered when choosing target words. In this case, it is important to determine what makes these words easy before deciding whether to remove them from the research pool. To this end the following paragraphs attempt to identify the special features of the two easy words found in this study.

When the frequency of encountering words and high task-induced involvement are excluded as possible reasons, the factors that may be connected to the ease or difficulty of words that remain relate to word features; examples include part of speech (e.g. Lin, 2010; Paribakht & Wesche, 1997; Qian, 2003) and conceptual difficulty (Nagy et al., 1987), although neither factor has been fully investigated within incidental vocabulary acquisition research. In this case, both target words are nouns that represent a conceptual idea that is very easy to understand. In addition to the factors mentioned above, the factors relating to word features mentioned in Chapter 3 may also need to be taken into consideration (although they have not appeared in previous research).

The word "3 耐克 (Nàikè) [Nike]" was noticed due to the high scores participants in both the treatment and control groups attained in the matching test in the first experiment. This word's Pinyin form and English equivalent are very close to each

other as they share both “n” and “ke”, which makes correct matching very easy for learners. It may be argued that the high score might be caused by random guessing. However, the word’s higher scores in the spoken form recognition test in the second experiment also support the view above. The audio annotation provided the spoken form of the target word, which could also be easily connected to the right meaning as its pronunciation is also very similar to that of its English equivalent. This trait was also found in “1 丝绸 (sīchóu) [silk]”, where “si” is common to both the Pinyin and English forms of the word. Other evidence comes from the error data in the Pinyin production tests presented in Chapter 6: it was found that some participants answered “nike” for “3 耐克 (Nàikè) [Nike]”.

From the perspective of the meaning-based character production test adopted in the first experiment, partial knowledge gain results for “3 耐克 (Nàikè) [Nike]” indicate that the high score for this word was caused mainly by participants’ ability to get the second character (i.e. “克 (kè) [gram]”) fully or partially correct. The participants’ answers to the pre-knowledge question in the questionnaire demonstrated a very low amount of pre-knowledge of this word. If the answers to the pre-knowledge question are reliable, a reason for the high score could be participants’ familiarity with the components in this word. According to the *GB.13000.1 Character Set for Information Processing-Specification of Chinese Character Components* (The National Language Committee, 1997), “克 (kè) [gram]” consists of three commonly used components: “十 (shí) [ten]”, “口 (kǒu) [mouth/measure word]”, and “儿 (ér) [son]” or used in “一点儿 (yìdiǎnr) [a little]”. Moreover, three is also a relatively low number of components, as Chinese characters can be made of up to 13 components (Zhou & Chen, 1998; see Section 3.3.2).

The word “1 丝绸 (sīchóu) [silk]” may have obtained a higher score for three reasons. The first is the repetition of the word’s components (the component “纟 [silk], also known as the silk radical, suggesting that the character might be related to silk or other textual” appeared three times in this word). The second is that the learners had already been taught all of the components involved in this word, which should have

made it easier for them. Finally, the word was repeated seven times in the article that was used for the experiment.

The above-mentioned similarities between each words' Pinyin and English forms could have helped learners with the vocabulary tests on sound-related knowledge. However, another feature – namely the repetition of components in the target word – may have assisted learners on tests that involved word character form; for example, the silk radical "纟" appears three times in "1 丝绸 (sīchóu) [silk]". It is possible that the second feature of Chinese words also helped the participants to acquire "5 折扣 (zhékòu) [discount]", as the hand radical "扌" appears twice in this word.

This explanation is further supported by the other two words containing repeated components used in this study: "2 价值 (jiàzhí) [value]", with a repeated "亻" (people radical), and "15 懒惰 (lǎnduò) [lazy]", with a repeated "忄" (heart radical). With mean scores of respectively 0.7 and 0.59, these words attained the fifth and sixth positions in the vocabulary posttests (see Table 6.6 in Section 6.2.2). It is therefore possible that the repetition of word components enhanced incidental vocabulary learning.

In addition to the above phenomena, it was also noted that scores decreased as character complexity increased. Moreover, familiarity with components or a component's frequency may also contribute to better learning results; for instance, the commonly used and learnt component "口 (kǒu) [mouth]" could account for the higher score for "5 折扣 (zhé kòu) [discount]".

On the basis of the above observations, repeated components and familiarity with components appear to be features that could also account for the better learning results achieved through incidental vocabulary learning in this study. However, the finding is just an inference at this stage and requires solid support from further empirical studies.

The results also suggest that some of the words used in this study were more difficult than others. For example, in the second experiment, "6 减缓 (jiǎnhuǎn) [slowdown]" and "8 尽量 (jǐnliàng) [to the best of one's ability]", which respectively

attained mean total scores of 0.07 and 0.25 on the vocabulary test, had the two lowest scores among all 15 target words (see Table 6.6 in Section 6.2.2). Their difficulty could be caused by the abstract meanings that they carry. It is also noted that "6 减缓 (jiǎnhuǎn) [slowdown]", which is a verb in Mandarin Chinese, obtained a much lower score than other verbs in this study. Its score was also lower than the score for the adverb "8 尽量 (jǐnliàng) [to the best of one's ability]", which conflicts with Laufer's (1990) assertion that adverbs are the most difficult type of word for learners. This might be caused by the verb-complement structure of "6 减缓 (jiǎnhuǎn) [slowdown]". The other three verbs used in this test, namely "14 保持 (bǎochí) [to keep; to maintain]", "4 判断 (pànduàn) [judge]", and "13 培养 (péiyǎng) [cultivate]" happen to have the same parallel structure (i.e. each word's two characters have the same or similar meanings). The internal structure of "6 减缓 (jiǎnhuǎn) [slowdown]" is more complicated, as "减 (jiǎ) [reduce; decrease]" carries the verb's main meaning whereas "缓 (huǎn) [slow]" implies the results of the action. This complication may also be reflected in the word's English translation, which is a combination of the meaning of "slow" and "down". Moreover, none of these verbs contain repeated components or have Pinyin and English forms that are similar (as mentioned earlier).

To summarise, several word properties may account for the word effect found in this study, namely: similarity between a word's Pinyin form and English equivalent, number of components, inclusion of special components (e.g. phonetic component and repeated component), and presence of a difficult word structure (e.g. a verb-complement structure) while others may increase the difficulty (e.g. the involvement of a verb-complement structure). However, without sufficient evidence of these findings, they are merely conjecture and require further research. As such, although not an original objective of this research, an attempt to identify influential factors is made in Chapter 8 using scores attained for all 15 target words in the second experiment. It should also be noted that other word properties that could affect incidental vocabulary acquisition may also exist beyond those demonstrated by this study's target words.

## 7.5 Issues identified through a cross-posttest comparison

This section compares results obtained across different types of vocabulary tests used in the second experiment. The similar results found for the spoken form recognition test and the meaning-base Pinyin form production test are examined first; thereafter the better results achieved for the Pinyin input assisted meaning-based character form production test in relation to both Pinyin form production tests are explored.

Regarding the first finding, it is interesting to discover that participants obtained similar scores for the two types of form recognition tests, which suggests that recognising a word's character form is not more difficult than recognising its Pinyin form. In L2 Mandarin Chinese learning, it is commonly accepted that characters are very difficult for Western learners. In the past 20 years, a number of teachers and researchers have proposed separating character recognition from handwriting in the curriculum to reduce the difficulty of character writing (Song, 2000; Jiang, 2007). The findings of the current study support the recommendation to emphasise the ease of recognising words' character forms based on another perspective.

Turning to the second issue mentioned above, the current study's detailed results concerning knowledge type and the amount of knowledge gain reveal many interesting findings. In particular, the amount of knowledge gain in the meaning-based character production test (12.20%) is higher – although not by much – than that in both the meaning-based Pinyin production test (10.87%) and the character-based Pinyin production test (10.67%). This result seems to be problematic, as the character production test not only covers all of the word knowledge types assessed by the other two tests but also requires learners to provide an extra type of word knowledge (see Table 6.1 in Section 6.1.2). Put in another way, when completing the character production test online by using Pinyin input, learners cannot choose character forms without correctly producing the Pinyin forms; as such, this test's score should theoretically not exceed that of the Pinyin production test. Given the 12.20% knowledge gain in the character production test and gain of only 10.87% in the Pinyin

production test, this is apparently not the case. This conflict could be explained by the adoption of the Pinyin Input method as following.

Typing a Mandarin Chinese word using Pinyin input does not require the Pinyin's tone. This means that when learners were entering the Pinyin of a two-character word, they did not need to have mastered one-third of the knowledge (according to the partial knowledge-sensitive criteria). Moreover, the in-depth analysis of the Pinyin production tests in both the first and second experiments revealed that tone happened to be the most difficult part of Pinyin for learners to acquire (see Chapters 5 and 6 for details). This phenomenon may to some extent have increased the chances that participants obtained correct answers in this test. In addition, learners frequently do not need to master the complete Pinyin of a word (i.e. its other two-thirds) before they can type it into a computer. This is because the Google Input method's "word recommend" function provides several options for users to choose from according to the Pinyin they are inputting (as mentioned in Section 4.2.7). Table 7.2 below shows the pieces of Pinyin for each word typed into the computer when a word's correct character form is shown with corresponding its option number appears in the recommendation bar.

Table 7.2 Target words presented when Pinyin is typed into Google Pinyin Input

| Target word with serial number | Pinyin typed in | Option number of the word |
|---|---|---|
| 1 丝绸(sīchóu) [silk] | si c, si ch<br>si cho<br>si chou | 3<br>4<br>1 (other options are all single character) |
| 2 价值(jiàzhí) [value] | jia z, jia zh<br>jia zhi | 2<br>1 |
| 3 耐克 (Nàikè) [Nike] | Nai k, nai ke | 1 |
| 4 判断(pànduàn) [judge] | pan d,<br>pan du, pan dua<br>pan duan | 1<br>Not in the list<br>1 |

| | | |
|---|---|---|
| 5 折扣(zhékòu) [discount] | zhe k<br>zhe ko<br>zhe kou | 4<br>5<br>1 |
| 6 减缓(jiǎnhuǎn)[slowdown] | jiang huan | 1 |
| 7 眩晕(xuànyūn) [dizziness] | xuan yun | 1 |
| 8 尽量(jìnliàng) [to the best of one's ability] | jin l<br>jin li, jin lia, jin lian<br>jin liang | 1<br>Not in the list<br>1 |
| 9 症状(zhèngzhuàng) [symptom] | zheng z<br>zheng zh<br>zheng zhu, zheng zhua, zheng zhuan<br>zheng zhuang | 4<br>3<br>Not in the list<br>1 |
| 10 稍微(shāowéi) [slightly] | shao w<br>shao we<br>shao wei | 1<br>2<br>1 |
| 11 庄稼(zhuāngjia) [crops] | zhuang j<br>zhuang ji<br>zhuang jia | 5<br>Not in the list<br>3 |
| Target word with serial number | Pinyin typed in | Option number of the word |
| 12 辛苦(xīnkǔ) [to work hard; laborious] | xin k, xin ku | 1 |
| 13 培养(péiyǎng) [cultivate] | pei y<br>pei ya, pei yan<br>pei yang | 1<br>Not in the list<br>1 |
| 14 保持(bǎochí) [to keep; to maintain] | bao c, bao ch, bao chi | 1 |
| 15 懒惰(lǎnduò) [lazy] | lan d<br>lan du<br>lan duo | 3<br>Not in the list<br>1 |

This table reveals that correctly entering the Pinyin of a word's first character and the initial part of its second character leads to 13 of the 15 target words (87%) in the second experiment being presented among the suggested options (the exceptions are

“6 减缓(jiǎnhuǎn)[slowdown]” and “7 眩晕 (xuànyūn) [dizziness]”). In addition, if the initial part of the second character is a double-consonant initial (i.e. “zh”, “ch”, or “sh”), only the first letter is required to call the right word up. As such, for most of the target words used in this experiment, participants only had to produce three out of the total six parts (50%) of a word’s entire Pinyin form to be able to select the word from options presented by Google Input and subsequently produce the whole word in character form. In other words, a 0.5-point gain in the Pinyin production test may contribute maximally to a 1-point gain in the test that requires the character form to be provided. The Pinyin Input application could thus serve as a good practice tool for L2 Mandarin Chinese learners, as it would reduce the difficulty involved in handwriting and increase the successfulness of producing the character form in another way. Practically, in Mandarin Chinese teaching, exercises could involve typing with Pinyin input; for example, presenting a word or a sentence in both character and Pinyin, which the learners could then type using Pinyin input. This process could enhance multiple types of word knowledge, including Pinyin form, character form, link between Pinyin and character forms, and link between Pinyin form and meaning.

## 7.6 Participants’ attitudes towards various types of annotations

In this section the participants’ attitudes towards different types of annotation are dealt with and the discussion is based mainly on the results of the second experiment, because the sample size for this test was relatively larger than the first one and all the participants had access to all types of annotations. However, the results of the first experiment are also consulted where necessary.

Significant differences in attitude were found between word meaning and Pinyin ($Z = -3.917$, $p = 0.00$), as well as between word meaning and audio ($Z = -4.303$, $p = 0.00$). The much higher mean score, 4.58, for word meaning indicates that this type of information was the most helpful. The reason for this is apparently connected to the nature of the reading-comprehension tasks. In order to understand the articles, priority was given to the meaning of unknown words rather than to their sound.

With respect to the attitudes towards the Pinyin and the audio annotation, although the mean scores were very close (3.42 for the audio annotation and 3.69 for the Pinyin annotation), the participants' choices were dispersed. With the majority of the participants (44 out of 48) giving it between three and five points, Pinyin was considered to be generally helpful to incidental vocabulary learning through reading. By contrast, the scores for the audio annotation show an even tendency from two to five points, suggesting that fewer respondents regarded it as being as helpful as the Pinyin annotation. A possible reason for this could be the information modality. Using Pinyin annotation, which provides the phonetic information of words in a written form, seems to be more in line with written reading material. Alternatively, it could be related to the participants' perceptual learning styles.

It is also quite interesting to see that their opinions about the audio information varied largely. Compared with the results of the first experiment, the participants' attitudes towards the meaning and Pinyin annotations were quite similar, while in the first experiment the marks related to the audio information provided in annotations were relatively focused around three and four; however, the participants chose scores of two and five in equal ratio to the scores of three and four in the second experiment. The results from the second experiment might be more accurate than those of the first one, because the audio annotation was not used in that test and the participants made their choices based on their assumptions about the effect of audio annotations at that time. If this is the case, it is interesting to discover that audio information is received so differently by different people, also suggesting a possible impact brought about by the learners' individual learning styles.

In response to the question about the participants' learning styles (whether they were visual, auditory or kinaesthetic learners?), 59.26% of the participants regarded themselves as visual learners, only 11.11% as auditory learners and 22.22% as kinaesthetic learners, while the rest had a combined learning style. For example, one participant described his learning style as 30% auditory and 70% visual. All the five auditory or partially auditory learners chose four or five points for the audio

annotation when asked about their attitudes towards different types of annotation, suggesting that the audio annotation was very helpful to them. The answers to the vocabulary tests seem to corroborate their choices. Four participants out of five achieved the highest scores for the words for which the audio annotation had been provided (three participants with the text + Pinyin + audio annotation and one participant with the text + audio annotation). This seems to conflict with the findings of Yeh and Wang (2003), who suggested that perceptual learning styles did not interact with annotation type. However, with only a small number of auditory learners participating in this research, insufficient information was gathered for drawing any conclusions.

## 7.7 Useful information gathered from the log file and questionnaire

In this section data elicited by the feedback question in the questionnaire and the log file of the tracking tool are presented as evidence that the research design meets two requirements: 1) investigating incidental vocabulary learning and 2) the well-controlled frequency of encountering the target words.

Based on the methodology definition of incidental vocabulary mentioned in Section 2.2.1.2, in order to keep vocabulary learning through reading incidental, the most crucial condition was to ensure the learners did not expect the vocabulary posttests. It is clear from the participants' answers to the feedback question that they did not realise that the real intention of this exercise was a vocabulary learning test rather than a reading exercise, even after completing the entire online program, which contained a vocabulary pretest. Several comments of the participants are presented as evidence:

“A little bit repeative [sic; this word should be ‘repetitive’] at points, but fine”;

“…what I've said above [sic; ‘about’] the test was interesting. I didn't really know what the point of the test was so didn't realise it was investigating my ability to learn Chinese”.

Regarding the first comment, it is clear that the participant did not have a clue that the real intention of this test was incidental vocabulary learning rather than reading comprehension, while the second comment suggests that the participant finally figured out the real intention of the test after completing it. There are many statements similar to the first one, but no others like the second one. In both cases, the participant did not expect the posttest and it is thus evident that this research design qualifies as an incidental research design, and that this was not affected by the pretest employed in the second experiment.

Turning to the log file data, when the students used the online program, a log file of their actions was recorded by an application that determined how many times they looked at the annotation of each word. It was found that almost every participant clicked each link for the annotated words, including that for the target words and distracters. In addition, none of the links were clicked twice, even when the students encountered the same word later in the article, suggesting that, firstly, none of the target words were ignored in the reading process and, secondly, that the participants did not repeatedly look at the annotations. This result confirms their answers to the question asking them how many times they looked at each target word.

With respect to the frequency with which they encountered the target words, only one thing needs to be addressed here, namely the difference between the occurrence and noticing or processing of a word. The high frequency of appearance does not necessarily lead to equal instances of noticing or processing a word, which is required for acquiring it, and vice versa – language learners can repeatedly look at the same word in reading if necessary. Therefore, the frequency of encountering the target words might be more reliable if it represents how many times language learners actually look at the word or annotation (see Section 2.2.3.6). Therefore, the participants' answers to the question "how many times did you look at the marginal annotation for each word on average?" was consulted for supplementary information. It was quite surprising to find that 56% of the participants in the first experiment indicated that they had looked at the annotations of each word three to six times on

average, and the rest reported a lower number of less than three times. This is quite different from the answers to this question collected from the second experiment. In the second experiment, 44 responses to this question were received. Only four of them, that is, 9% suggested that they had looked at each annotation three to six times, while the rest, 91%, indicated a lower number of times. The reason for this difference could be that both "丝绸 (sīchóu) [silk]" and "耐克 (Nàikè) [Nike]" appeared more times in the first experiment or, alternatively, it could have been due to the nature of the annotations in paper-based reading materials, because annotations do not disappear during the reading process.

However, in the pen-and-paper environment there is no way in which the reliability of the participants' answers could be further confirmed. On the other hand, with the tracking tool embedded in the online program for the second experiment, the records of the participants' online clicking provided an accurate source of evidence, which supported the participants' answers to that question. The tracking record clearly demonstrated that the participants tended to click the link of annotation only once while reading, and there were very few occasions when they missed one or two annotations. It could also be because of the underscored format of the target words, which usually suggested that extra information had been provided, that the participants clicked on the link. However, every word was annotated only the first time it appeared, so the participants might not have bothered to find and click the word when it appeared again.

Apart from these two issues, the need for sound information in Mandarin Chinese reading is also supported by the feedback of a participant who stated that "I think the point of this survey is great, to be able to see a word used in context is to know the word. The prerequisite for understanding its usage in a sentence is being able to pronounce the word in your head, and this mandates Pinyin usage". His statement indicates that the sound of the word is indispensable in the process of decoding the whole sentence, although this might not have been the case for every learner.

# Chapter 8: Further Analyses

As mentioned in the previous chapters, certain target words, i.e. “丝绸 (sīchóu)[silk]” and “耐克 (Nàikè) [Nike]”, are much easier to learn than the others in both experiments in this study. This chapter comprises an attempt to answer one question: what factors relating to word features make a word easier and thus should be considered when selecting target words for research into incidental vocabulary acquisition? Although this is not directly connected to the research questions of this study, it is believed that identifying these factors will benefit future studies in terms of selecting target words in the field of research into incidental vocabulary acquisition in L2 Mandarin Chinese, to help researchers to reduce the possible influences of such factors and thereby improve the quality of their research in the field.

Similar word effect was also reported by Waring and Takaki (2003), as discussed in Section 7.4. In their study, such word effect was explained by the ease of guessing the meaning of such words in L2 English incidental vocabulary acquisition through reading. Those words were simply excluded from the final analysis. However, by excluding such words in their study, they might have missed the chance to identify other possible factors that could contribute to the ease of learning the target words; for example, part of speech and conceptual difficulty, as suggested by other researchers who conducted related studies (see Section 2.2.3.5). As both “丝绸 (sīchóu)[silk]” and “耐克 (Nàikè) [Nike]” are nouns and easy to understand, these two factors do not seem to account for the ease of learning the two words found in this study. Other factors, though not used in previous studies on incidental vocabulary learning, thus, need to be considered and the possible ones are those related to Chinese character recognition.

In this case, preliminary efforts were made to identify such factors in the field of incidental L2 Mandarin Chinese vocabulary learning through reading. The first step was to start with those most often discussed factors relating to Chinese character recognition that were identified in Section 3.3.2, including number of characters, number of components and character structure. However, it is noteworthy that all of the target words in this research contained two characters and the previous studies

were focused mainly on recognition of single characters. The factors investigated in this chapter were thus modified as: 1) the number of strokes (including those of both the first and second characters and in the whole word); 2) the number of components (including the number of components in both the first and second characters and in the whole word, as well as repeated components in the word); and 3) the character's structure (including the structure of the first and second character). Relevant information on each character, such as the number of strokes, the structure and the phonetic component of the character were first analysed by the author and then double checked with another teacher, as well as in dictionaries[13].

All these factors are analysed in the next section, using data collected from the second experiment mainly due to its larger sample size, although it was still not large enough to conduct a regression test. Alternatively, these factors were examined one by one, with the results presented in Excel tables. In each table, the mean scores of each target word obtained in each part of the vocabulary posttests, as well as the mean total score of the whole posttests, are presented according to the factor that needed to be examined, which is shown in the second column. In each column of these mean scores, the bold figures show the highest five scores achieved by the target words and the lowest five scores are underscored. The column in which the factor is presented is arranged in ascending order. Using this treatment, the distribution of the cells of the same format might amass or congregate together, thereby suggesting the influences of each factor. Although it is reasonable to believe that these character-related factors might have influenced only the answers to the questions testing the recognition or production of the word character form, there was no evidence that such factors would not affect learning other aspects of word knowledge. Therefore, scores for all types of word knowledge tests were included in these tables.

[13] The dictionary mainly used was an online dictionary called "汉典 (hàndiǎn) [Chinese dictionary]". This online dictionary can be accessed via the link: http://www.zdic.net/

## 8.1 Analysis of the number of strokes in a target word

The first set of factors investigated in this section is that of the stroke number factors, including the stroke number of the first character, that of the second character and that of the complete word. Three tables are presented at the end of this section, showing the mean scores for each target word in each part of the posttests, as well as the mean total score for all the posttests in an ascending order in the second column in relation to the number of strokes in the first character in Table 8.1, the second character in Table 8.2 and the whole word in Table 8.3. If these factors do indeed affect the acquisition of the target words, one would expect to see higher scores associated with a lower number of strokes, because characters with fewer strokes are easier to recognise, as mentioned in Section 3.3.2.

In Table 8.1, the number of strokes in the first character in the target words ranged from five to 16. The distribution of the high scores (in bold) indicates that all the five highest mean total scores were attained for words in which the first character comprised less than nine strokes. However, both the highest scores and lowest scores (underscored) were found to be evenly distributed throughout the table, suggesting that the number of strokes in the first character did not affect the results of word learning in this study. Similar even distribution patterns were found in Table 8.2, suggesting that the number of strokes in the second character (ranging from six to 15) might also not affect word learning in this research. It is noted that the distribution pattern was different in Table 8.3, in which the number of strokes in the whole word, ranging from 13 to 28, was used as the controlling factor. The table does reflect the expectation of data distribution mentioned earlier, with 84% of the highest scores for each part of the posttest obtained by the five words with no more than 16 strokes and all five highest total mean scores obtained by these words. More interestingly, this tendency is clearly demonstrated by the results of the spoken form recognition test, rather than the character form recognition test, suggesting a possible relationship between the learning of the spoken form and the total number of strokes in a word.

Table 8.1 The mean scores of the post-test arranged according to the stroke number of the first character in the target words

| Target words | Number of strokes in the first character | Meaning-based Pinyin form production test | Pinyin input assisted meaning-based character form production test | Character form recognition test | Spoken form recognition test | Character-based Pinyin form production test | Total |
|---|---|---|---|---|---|---|---|
| 1 丝绸(sī chóu) [silk] | 5 | **0.24** | **0.31** | **0.33** | **0.44** | **0.25** | **1.58** |
| 2 价值(jià zhí) [value] | 6 | **0.13** | 0.06 | **0.19** | **0.19** | 0.13 | **0.70** |
| 8 尽量(jìn liàng) [to the best of one's ability] | 6 | 0.03 | 0.04 | 0.10 | 0.04 | 0.03 | 0.25 |
| 11 庄稼(zhuāng jia) [crops] | 6 | 0.07 | 0.06 | 0.10 | 0.04 | **0.16** | 0.44 |
| 4 判断(pàn duàn) [judge] | 7 | **0.13** | 0.08 | 0.04 | 0.00 | 0.03 | 0.28 |
| 5 折扣(zhé kòu) [discount] | 7 | 0.09 | **0.13** | **0.17** | **0.38** | **0.14** | **0.89** |
| 12 辛苦(xīn kǔ) [to work hard; laborious] | 7 | 0.10 | **0.10** | 0.15 | **0.23** | **0.15** | **0.73** |
| 3 耐克(Nài kè) [Nike] | 9 | **0.40** | **0.44** | **0.75** | **0.58** | **0.27** | **2.44** |
| 14 保持(bǎo chí) [to keep; to maintain] | 9 | 0.07 | 0.08 | 0.00 | 0.06 | 0.06 | 0.28 |
| 7 眩晕(xuànyūn) [dizziness] | 10 | 0.02 | 0.06 | 0.08 | 0.15 | 0.02 | 0.33 |
| 9 症状(zhèng zhuàng) [symptom] | 10 | 0.08 | 0.08 | 0.06 | 0.08 | 0.07 | 0.38 |
| 6 减缓(jiǎn huǎn) [slowdown] | 11 | 0.00 | 0.02 | 0.02 | 0.02 | 0.01 | 0.07 |
| 13 培养(péi yǎng) [cultivate] | 11 | 0.09 | **0.13** | **0.17** | 0.15 | 0.05 | 0.58 |
| 10 稍微(shāo wēi) [slightly] | 12 | 0.07 | 0.08 | 0.00 | 0.02 | 0.10 | 0.27 |
| 15 懒惰(lǎn duò) [lazy] | 16 | **0.11** | 0.08 | **0.21** | 0.13 | 0.06 | 0.59 |

Table 8.2 The mean scores of each test arranged according to the number of strokes in the second character of each target word

| Target words | Number of strokes in the second character | Meaning-based Pinyin form production test | Pinyin input assisted meaning-based character form production test | Character form recognition test | Spoken form recognition test | Character-based Pinyin form production test | Total |
|---|---|---|---|---|---|---|---|
| 5 折扣(zhé kòu) [discount] | 6 | 0.09 | **0.13** | **0.17** | **0.38** | **0.14** | **0.89** |
| 3 耐克 (Nài kè) [Nike] | 7 | **0.40** | **0.44** | **0.75** | **0.58** | **0.27** | **2.44** |
| 9 症状(zhèng zhuàng) [symptom] | 7 | 0.08 | 0.08 | 0.06 | 0.08 | 0.07 | 0.38 |
| 12 辛苦(xīn kǔ) [to work hard; laborious] | 8 | 0.10 | **0.10** | 0.15 | **0.23** | **0.15** | **0.73** |
| 13 培养(péi yǎng) [cultivate] | 9 | 0.09 | **0.13** | **0.17** | 0.15 | 0.05 | 0.58 |
| 14 保持(bǎo chí) [to keep; to maintain] | 9 | 0.07 | 0.08 | 0.00 | 0.06 | 0.06 | 0.28 |
| 2 价值(jià zhí) [value] | 10 | **0.13** | 0.06 | **0.19** | **0.19** | 0.13 | **0.70** |
| 7 眩晕(xuàn yūn) [dizziness] | 10 | 0.02 | 0.06 | 0.08 | 0.15 | 0.02 | 0.33 |
| 1 丝绸(sī chóu) [silk] | 11 | **0.24** | **0.31** | **0.33** | **0.44** | **0.25** | **1.58** |
| 4 判断(pàn duàn) [judge] | 11 | **0.13** | 0.08 | 0.04 | 0.00 | 0.03 | 0.28 |
| 6 减缓 (jiǎn huǎn) [slowdown] | 12 | 0.00 | 0.02 | 0.02 | 0.02 | 0.01 | 0.07 |
| 8 尽量(jìn liàng) [to the best of one's ability] | 12 | 0.03 | 0.04 | 0.10 | 0.04 | 0.03 | 0.25 |
| 15 懒惰(lǎn duò) [lazy] | 12 | **0.11** | 0.08 | **0.21** | 0.13 | 0.06 | 0.59 |
| 10 稍微(shāo wēi) [slightly] | 13 | 0.07 | 0.08 | 0.00 | 0.02 | 0.10 | 0.27 |
| 11 庄稼(zhuāng jia) [crops] | 15 | 0.07 | 0.06 | 0.10 | 0.04 | **0.16** | 0.44 |

Table 8.3 The mean scores for each test arranged according to the number of total strokes of the whole word

| Target words | Total number of strokes in word | Meaning-based Pinyin form production test | Pinyin input assisted meaning-based character form production test | Character form recognition test | Spoken form recognition test | Character-based Pinyin form production test | Total |
|---|---|---|---|---|---|---|---|
| 5 折扣(zhé kòu) [discount] | 13 | 0.09 | **0.13** | **0.17** | **0.38** | **0.14** | **0.89** |
| 12 辛苦(xīn kǔ) [to work hard; laborious] | 15 | 0.10 | **0.10** | 0.15 | **0.23** | **0.15** | **0.73** |
| 1 丝绸(sī chóu) [silk] | 16 | **0.24** | **0.31** | **0.33** | **0.44** | **0.25** | **1.58** |
| 2 价值(jià zhí) [value] | 16 | **0.13** | 0.06 | **0.19** | **0.19** | 0.13 | **0.70** |
| 3 耐克(Nài kè) [Nike] | 16 | **0.40** | **0.44** | **0.75** | **0.58** | **0.27** | **2.44** |
| 9 症状(zhèng zhuàng) [symptom] | 17 | 0.08 | 0.08 | 0.06 | 0.08 | 0.07 | 0.38 |
| 4 判断(pàn duàn) [judge] | 18 | **0.13** | 0.08 | 0.04 | 0.00 | 0.03 | 0.28 |
| 8 尽量(jìn liàng) [to the best of one's ability] | 18 | 0.03 | 0.04 | 0.10 | 0.04 | 0.03 | 0.25 |
| 14 保持(bǎo chí) [to keep; to maintain] | 18 | 0.07 | 0.08 | 0.00 | 0.06 | 0.06 | 0.28 |
| 7 眩晕(xuànyūn) [dizziness] | 20 | 0.02 | 0.06 | 0.08 | 0.15 | 0.02 | 0.33 |
| 13 培养(péi yǎng) [cultivate] | 20 | 0.09 | **0.13** | **0.17** | 0.15 | 0.05 | 0.58 |
| 11 庄稼(zhuāng jia) [crops] | 21 | 0.07 | 0.06 | 0.10 | 0.04 | **0.16** | 0.44 |
| 6 减缓(jiǎn huǎn) [slowdown] | 23 | 0.00 | 0.02 | 0.02 | 0.02 | 0.01 | 0.07 |
| 10 稍微(shāo wēi) [slightly] | 25 | 0.07 | 0.08 | 0.00 | 0.02 | 0.10 | 0.27 |
| 15 懒惰(lǎn duò) [lazy] | 28 | **0.11** | 0.08 | **0.21** | 0.13 | 0.06 | 0.59 |

## 8.2 Analysis the number of components in each target word

In this section, the set of factors being examined are those relating to the number of components in Chinese characters. Four tables are presented at the end of this section, showing the mean scores for each target word in each part of the posttest, as well as the mean total score for the whole posttest arranged in an ascending order in the second column, where the number of components of the first character is presented in Table 8.4, that of the second character in Table 8.5, that of the whole word in Table 8.6, as well as the number of repeated components in Table 8.7. If this factor does indeed affect the acquisition of the target words, one would expect to see higher scores associated with words with a lower number of components in the first three tables and a reversed distribution pattern in the last table. The format of the components and the code was taken from *GB.13000.1 Character Set for Information Processing-specification on Chinese Character Components,* which was mentioned in Section 4.2.4. In this document, some complete essential components of Chinese characters are further divided into subsets. However, only the primary components were counted in this research.

Firstly, Table 8.4 shows the mean scores for each target word in the vocabulary tests, arranged according to the number of components contained in the first character of each word, which were two to four components in this study. As shown in the table, 52% of the highest scores (13 out of 25) are associated with the five words in which the first character is a two-component character, that is, for each of them there is a chance of obtaining 10.4% of the highest scores on average. The percentages for words with the first character being a three-component character and a four-component character are, respectively, 5% and 4%. In this case, it seems that incidental word learning in L2 Mandarin Chinese is connected with the number of components in the first character of a word, and a word might be easier to learn if it has fewer components in the first character. A similar tendency was also found in Table 8.6; the percentage decreases from 16% for words containing four components to 7.2% for words containing five components, to 4.8% for words with six components and finally to 2.7% for words with

eight components. The only exception is the score for the word with seven components, which is 20%. However, the word containing seven components is “丝绸(sīchóu)[silk]”, one of the easiest words in this research, and it is very likely that other factors contributed to such a high percentage, as discussed in Section 7.4. Unlike the distribution patterns shown in the first two tables, the results shown in Table 8.5, with two to five components in the second character of the target words, do not suggest any noticeable pattern in the distribution of the scores.

The last part of this section deals with the number of repeated components in whole words. Among the 15 target words, five of them contained one repeated component, and 80% of the highest mean total scores were associated with them, as well as 68% of the highest scores for each part of the vocabulary posttest, suggesting that this factor could have affected word learning in this research and therefore needs to be further researched.

To sum up, the results indicate that the number of components in the first character and in a whole word, as well as in the first character of the word, might affect word learning. In addition, the repetition of components in a word might also contribute to reducing the learning difficulty of this word.

Table 8.4 Mean scores of vocabulary tests arranged according to the number of components in the first character of the target words

| Target words | Number of components in the first character | Components in the first character | Meaning-based Pinyin form production test | Pinyin input assisted meaning-based character form production test | Character form recognition test | Spoken form recognition test | Character--based Pinyin form production test | Total |
|---|---|---|---|---|---|---|---|---|
| 3 耐克(Nài kè) [Nike] | 2 | 而 255 寸 095 | **0.40** | **0.44** | **0.75** | **0.58** | **0.27** | **2.44** |
| 5 折扣(zhé kòu) [discount] | 2 | 扌 055 斤 208 | 0.09 | **0.13** | **0.17** | **0.38** | **0.14** | **0.89** |
| 8 尽量(jìn liàng) [to the best of one's ability] | 2 | 尺 429? 147 | 0.03 | 0.04 | 0.10 | 0.04 | 0.03 | 0.25 |
| 11 庄稼(zhuāng jia) [crops] | 2 | 广 107 土 014 | 0.07 | 0.06 | 0.10 | 0.04 | **0.16** | 0.44 |
| 12 辛苦(xīn kǔ) [to work hard; laborious] | 2 | 立 080 十 027 | 0.10 | **0.10** | 0.15 | **0.23** | **0.15** | **0.73** |
| 1 丝绸(sī chóu) [silk] | 3 | ?032?032 一 002 | **0.24** | **0.31** | **0.33** | **0.44** | **0.25** | **1.58** |
| 2 价值(jià zhí) [value] | 3 | 亻 009 人 012 丿丨 277 | **0.13** | 0.06 | **0.19** | **0.19** | 0.13 | **0.70** |
| 4 判断(pàn duàn) [judge] | 3 | 丷 004? 153 刂 099 | **0.13** | 0.08 | 0.04 | 0.00 | 0.03 | 0.28 |
| 7 眩晕(xuàn yūn) [dizziness] | 3 | 目 091 亠 048 幺 033 | 0.02 | 0.06 | 0.08 | 0.15 | 0.02 | 0.33 |
| 9 症状(zhèng zhuàng) [symptom] | 3 | 疒 139 一 002 止 050 | 0.08 | 0.08 | 0.06 | 0.08 | 0.07 | 0.38 |
| 10 稍微(shāo wēi) [slightly] | 3 | 禾 081?078 月 020 | 0.07 | 0.08 | 0.00 | 0.02 | 0.10 | 0.27 |
| 13 培养(péi yǎng) [cultivate] | 3 | 土 014 立 080 口 001 | 0.09 | **0.13** | **0.17** | 0.15 | 0.05 | 0.58 |
| 14 保持(bǎo chí) [to keep; to maintain] | 3 | 亻 009 口 001 木 005 | 0.07 | 0.08 | 0.00 | 0.06 | 0.06 | 0.28 |
| 6 减缓(jiǎn huǎn) [slowdown] | 4 | 氵 146 戊 221 一 002 口 001 | 0.00 | 0.02 | 0.02 | 0.02 | 0.01 | 0.07 |
| 15 懒惰(lǎn duò) [lazy] | 4 | 忄 087 束 280 ? 57 贝 038 | **0.11** | 0.08 | **0.21** | 0.13 | 0.06 | 0.59 |

Table 8.5 Mean scores of vocabulary tests arranged according to the number of components in the second character of the target words

| Target words | Number of components in the second character | Components in the second character | Meaning-based Pinyin form Production test | Pinyin input assisted meaning-based character form production test | Character form recognition test | Spoken form recognition test | Character-based Pinyin form production test | Total |
|---|---|---|---|---|---|---|---|---|
| 5 折扣(zhé kòu) [discount] | 2 | 扌05 口001 | 0.09 | **0.13** | **0.17** | **0.38** | **0.14** | **0.89** |
| 9 症状(zhèng zhuàng) [symptom] | 2 | 丬29 犬225 | 0.08 | 0.08 | 0.06 | 0.08 | 0.07 | 0.38 |
| 2 价值(jià zhí) [value] | 3 | 亻009 十027 ? 295 | **0.13** | 0.06 | **0.19** | **0.19** | 0.13 | **0.70** |
| 3 耐克(Nài kè) [Nike] | 3 | 十027 口 001 儿150 | **0.40** | **0.44** | **0.75** | **0.58** | **0.27** | **2.44** |
| 4 判断(pàn duàn) [judge] | 3 | 米116 凵183 斤208 | **0.13** | 0.08 | 0.04 | 0.00 | 0.03 | 0.28 |
| 7 眩晕(xuànyūn) [dizziness] | 3 | 日009 冖083 车097 | 0.02 | 0.06 | 0.08 | 0.15 | 0.02 | 0.33 |
| 8 尽量(jìn liàng) [to the best of one's ability] | 3 | 曰009 一002 里282 | 0.03 | 0.04 | 0.10 | 0.04 | 0.03 | 0.25 |
| 11 庄稼(zhuāng jia) [crops] | 3 | 禾081 宀058 豕184 | 0.07 | 0.06 | 0.10 | 0.04 | **0.16** | 0.44 |
| 12 辛苦(xīn kǔ) [to work hard; laborious] | 3 | 艹016 十027 口001 | 0.10 | **0.10** | 0.15 | **0.23** | **0.15** | **0.73** |
| 13 培养(péi yǎng) [cultivate] | 3 | 丷069 夫17 丿丨277 | 0.09 | **0.13** | **0.17** | 0.15 | 0.05 | 0.58 |
| 14 保持(bǎo chí) [to keep;to maintain] | 3 | 亻009 土014 寸095 | 0.07 | 0.08 | 0.00 | 0.06 | 0.06 | 0.28 |
| 1 丝绸(sī chóu) [silk] | 4 | 纟029 冂067 土014 口001 | **0.24** | **0.31** | **0.33** | **0.44** | **0.25** | **1.58** |
| 6 减缓 (jiǎn huǎn) [slowdown] | 4 | 纟029 爫 134 ? 194 又46 | 0.00 | 0.02 | 0.02 | 0.02 | 0.01 | 0.07 |
| 15 懒惰(lǎn duò) [lazy] | 4 | 忄087 力115 工109 月020 | **0.11** | 0.08 | **0.21** | 0.13 | 0.06 | 0.59 |
| 10 稍微(shāo wēi) [slightly] | 5 | 彳168 山070 一002 几 093 攵037 | 0.07 | 0.08 | 0.00 | 0.02 | 0.10 | 0.27 |

Table 8.6 Mean scores of vocabulary tests arranged according to the number of total components of the whole word

| Target words | Total number of components in this word | Meaning-based Pinyin form Production test | Pinyin input assisted meaning-based character form production test | Character form recognition test | Spoken form recognition test | Character-based Pinyin form production test | Total |
|---|---|---|---|---|---|---|---|
| 5 折扣(zhé kòu) [discount] | 4 | 0.09 | **0.13** | **0.17** | **0.38** | **0.14** | **0.89** |
| 3 耐克 (Nài kè) [Nike] | 5 | **0.40** | **0.44** | **0.75** | **0.58** | **0.27** | **2.44** |
| 8 尽量(jìn liàng) [to the best of one's ability] | 5 | 0.03 | 0.04 | 0.10 | 0.04 | 0.03 | 0.25 |
| 9 症状(zhèng zhuàng) [symptom] | 5 | 0.08 | 0.08 | 0.06 | 0.08 | 0.07 | 0.38 |
| 11 庄稼(zhuāng jia) [crops] | 5 | 0.07 | 0.06 | 0.10 | 0.04 | **0.16** | 0.44 |
| 12 辛苦(xīn kǔ) [to work hard; laborious] | 5 | 0.10 | **0.10** | 0.15 | **0.23** | **0.15** | **0.73** |
| 2 价值(jià zhí) [value] | 6 | **0.13** | 0.06 | **0.19** | **0.19** | 0.13 | **0.70** |
| 4 判断(pàn duàn) [judge] | 6 | **0.13** | 0.08 | 0.04 | 0.00 | 0.03 | 0.28 |
| 7 眩晕(xuànyūn) [dizziness] | 6 | 0.02 | 0.06 | 0.08 | 0.15 | 0.02 | 0.33 |
| 13 培养(péi yǎng) [cultivate] | 6 | 0.09 | **0.13** | **0.17** | 0.15 | 0.05 | 0.58 |
| 14 保持(bǎo chí) [to keep; to maintain] | 6 | 0.07 | 0.08 | 0.00 | 0.06 | 0.06 | 0.28 |
| 1 丝绸(sī chóu) [silk] | 7 | **0.24** | **0.31** | **0.33** | **0.44** | **0.25** | **1.58** |
| 6 减缓 (jiǎn huǎn) [slowdown] | 8 | 0.00 | 0.02 | 0.02 | 0.02 | 0.01 | 0.07 |
| 10 稍微(shāo wēi) [slightly] | 8 | 0.07 | 0.08 | 0.00 | 0.02 | 0.10 | 0.27 |
| 15 懒惰(lǎn duò) [lazy] | 8 | **0.11** | 0.08 | **0.21** | 0.13 | 0.06 | 0.59 |

Table 8.7 Mean scores of vocabulary tests arranged according to the number of repeated components

| Target words | Number of repeated components | Meaning-based Pinyin form Production test | Pinyin input assisted meaning-based character form production test | Character form recognition test | Spoken form recognition test | Character-based Pinyin form production test | Total |
|---|---|---|---|---|---|---|---|
| 3 耐克 (Nài kè) [Nike] | 0 | **0.40** | **0.44** | **0.75** | **0.58** | **0.27** | **2.44** |
| 4 判断(pàn duàn) [judge] | 0 | **0.13** | 0.08 | 0.04 | 0.00 | 0.03 | 0.28 |
| 6 减缓 (jiǎn huǎn) [slowdown] | 0 | 0.00 | 0.02 | 0.02 | 0.02 | 0.01 | 0.07 |
| 7 眩晕(xuàn yūn) [dizziness] | 0 | 0.02 | 0.06 | 0.08 | 0.15 | 0.02 | 0.33 |
| 8 尽量(jìn liàng) [to the best of one's ability] | 0 | 0.03 | 0.04 | 0.10 | 0.04 | 0.03 | 0.25 |
| 9 症状(zhèng zhuàng) [symptom] | 0 | 0.08 | 0.08 | 0.06 | 0.08 | 0.07 | 0.38 |
| 10 稍微(shāo wēi) [slightly] | 0 | 0.07 | 0.08 | 0.00 | 0.02 | 0.10 | 0.27 |
| 11 庄稼(zhuāng jia) [crops] | 0 | 0.07 | 0.06 | 0.10 | 0.04 | **0.16** | 0.44 |
| 13 培养(péi yǎng) [cultivate] | 0 | 0.09 | **0.13** | **0.17** | 0.15 | 0.05 | 0.58 |
| 14 保持(bǎo chí) [to keep; to maintain] | 0 | 0.07 | 0.08 | 0.00 | 0.06 | 0.06 | 0.28 |
| 1 丝绸(sī chóu) [silk] | 1 | **0.24** | **0.31** | **0.33** | **0.44** | **0.25** | **1.58** |
| 2 价值(jià zhí) [value] | 1 | **0.13** | 0.06 | **0.19** | **0.19** | 0.13 | **0.70** |
| 5 折扣(zhé kòu) [discount] | 1 | 0.09 | **0.13** | **0.17** | **0.38** | **0.14** | **0.89** |
| 12 辛苦(xīn kǔ) [to work hard; laborious] | 1 | 0.10 | **0.10** | 0.15 | **0.23** | **0.15** | **0.73** |
| 15 懒惰(lǎn duò) [lazy] | 1 | **0.11** | 0.08 | **0.21** | 0.13 | 0.06 | 0.59 |

## 8.3 Analysis of the structure of characters in the target words

Tables 8.8 and 8.9 were developed to analyse the relationship between the results of the vocabulary tests and the structure of different characters. To reiterate, Mandarin Chinese characters are generally structured in the following way: single structure, left and right structure, left, middle and right structure, top and bottom structure, top, middle and bottom structure and, finally, surrounded structure. These structures have been replaced by the numbers 1 to 6 in the next two tables, which are presented at the end of this section, in which 1 = single structure; 2 = left and right structure; 3 = top and bottom structure; 4 = left, middle and right structure; 5 = top, middle and bottom structure; 6 = other structure.

As Table 8.8 shows, more than half of the high scores marked in bold were attained in relation to words in which the first character was formed in the left and right structure. However, the data is not strong enough to support the idea that characters with left and right structures in the word will lead to a better performance in the vocabulary tests because most of the first characters in the target words were formed in a left and right structure. Therefore, no special distribution pattern of the data can be generalised. Likewise, the distribution of scores in the vocabulary tests, shown in Table 8.9, is similar. Although most of the high scores were attained in relation to target words in which the second character was formed in the left and right structure, no sufficient evidence was provided here as most of the second characters in these target words were left and right structured.

Therefore, no connections between the structure of the characters in the target words and the scores attained in the vocabulary posttests were suggested.

Table 8.8 The mean scores of the vocabulary tests arranged according to the structure of the first character

| Target words | Structure of the first character | Meaning-based Pinyin form Production test | Pinyin input assisted meaning-based character form production test | Character form recognition test | Spoken form recognition test | Character-based Pinyin form production test | Total |
|---|---|---|---|---|---|---|---|
| 12 辛苦(xīn kǔ) [to work hard; laborious] | 1 | 0.10 | **0.10** | 0.15 | **0.23** | **0.15** | **0.73** |
| 2 价值(jià zhí) [value] | 2 | **0.13** | 0.06 | **0.19** | **0.19** | 0.13 | **0.70** |
| 3 耐克 (Nài kè) [Nike] | 2 | **0.40** | **0.44** | **0.75** | **0.58** | **0.27** | **2.44** |
| 4 判断(pàn duàn) [judge] | 2 | **0.13** | 0.08 | 0.04 | 0.00 | 0.03 | 0.28 |
| 5 折扣(zhé kòu) [discount] | 2 | 0.09 | **0.13** | **0.17** | **0.38** | **0.14** | **0.89** |
| 6 减缓 (jiǎn huǎn) [slowdown] | 2 | 0.00 | 0.02 | 0.02 | 0.02 | 0.01 | 0.07 |
| 7 眩晕(xuànyūn) [dizziness] | 2 | 0.02 | 0.06 | 0.08 | 0.15 | 0.02 | 0.33 |
| 10 稍微(shāo wēi) [slightly] | 2 | 0.07 | 0.08 | 0.00 | 0.02 | 0.10 | 0.27 |
| 13 培养(péi yǎng) [cultivate] | 2 | 0.09 | **0.13** | **0.17** | 0.15 | 0.05 | 0.58 |
| 14 保持(bǎo chí) [to keep; to maintain] | 2 | 0.07 | 0.08 | 0.00 | 0.06 | 0.06 | 0.28 |
| 1 丝绸(sī chóu) [silk] | 3 | **0.24** | **0.31** | **0.33** | **0.44** | **0.25** | **1.58** |
| 8 尽量(jìn liàng) [to the best of one's ability] | 3 | 0.03 | 0.04 | 0.10 | 0.04 | 0.03 | 0.25 |
| 15 懒惰(lǎn duò) [lazy] | 4 | **0.11** | 0.08 | **0.21** | 0.13 | 0.06 | 0.59 |
| 9 症状(zhèng zhuàng) [symptom] | 6 | 0.08 | 0.08 | 0.06 | 0.08 | 0.07 | 0.38 |
| 11 庄稼(zhuāng jia) [crops] | 6 | 0.07 | 0.06 | 0.10 | 0.04 | **0.16** | 0.44 |

Table 8.9 The mean scores of the vocabulary tests arranged according to the structure of the second character

| Target word | Structure of the second character | Meaning-based Pinyin form Production test | Pinyin input assisted meaning-based character form production test | Character form recognition test | Spoken form recognition test | Character-based Pinyin form production test | Total |
|---|---|---|---|---|---|---|---|
| 1 丝绸(sī chóu) [silk] | 2 | **0.24** | **0.31** | **0.33** | **0.44** | **0.25** | **1.58** |
| 2 价值(jià zhí) [value] | 2 | **0.13** | 0.06 | **0.19** | **0.19** | 0.13 | **0.70** |
| 4 判断(pàn duàn) [judge] | 2 | **0.13** | 0.08 | 0.04 | 0.00 | 0.03 | 0.28 |
| 5 折扣(zhé kòu) [discount] | 2 | 0.09 | **0.13** | **0.17** | **0.38** | **0.14** | **0.89** |
| 6 减缓(jiǎn huǎn) [slowdown] | 2 | 0.00 | 0.02 | 0.02 | 0.02 | 0.01 | 0.07 |
| 9 症状(zhèng zhuàng) [symptom] | 2 | 0.08 | 0.08 | 0.06 | 0.08 | 0.07 | 0.38 |
| 11 庄稼(zhuāng jia) [crops] | 2 | 0.07 | 0.06 | 0.10 | 0.04 | **0.16** | 0.44 |
| 12 辛苦(xīn kǔ) [to work hard; laborious] | 2 | 0.10 | **0.10** | 0.15 | **0.23** | **0.15** | **0.73** |
| 14 保持(bǎo chí) [to keep; to maintain] | 2 | 0.07 | 0.08 | 0.00 | 0.06 | 0.06 | 0.28 |
| 15 懒惰(lǎn duò) [lazy] | 2 | 0.11 | 0.08 | 0.21 | 0.13 | 0.06 | 0.59 |
| 7 眩晕(xuànyūn) [dizziness] | 3 | 0.02 | 0.06 | 0.08 | 0.15 | 0.02 | 0.33 |
| 8 尽量(jìn liàng) [to the best of one's ability] | 3 | 0.03 | 0.04 | 0.10 | 0.04 | 0.03 | 0.25 |
| 13 培养(péi yǎng) [cultivate] | 3 | 0.09 | **0.13** | **0.17** | 0.15 | 0.05 | 0.58 |
| 10 稍微(shāo wēi) [slightly] | 4 | 0.07 | 0.08 | 0.00 | 0.02 | 0.10 | 0.27 |
| 3 耐克 (Nài kè) [Nike] | 5 | **0.40** | **0.44** | **0.75** | **0.58** | **0.27** | **2.44** |

To sum up, in this chapter some factors that might influence incidental vocabulary learning in L2 Mandarin Chinese were preliminarily identified. All the factors involved at this stage, namely number of strokes, number of components and character structure, originally came from research relating to Chinese character recognition. The distribution patterns of the data presented in the tables in the previous sections indicate that the factor that is most likely to affect the results of similar research would be the number of strokes in a whole word. Furthermore, it indicates that words with less than or equal to 16 strokes can be remembered more easily than those with more strokes. Other factors worth investigating are the number of components in the whole word and in the first character of this word, as well as whether it contains a repeated component.

However, it should also be borne in mind that no conclusion could be drawn regarding a small number of the target words used in this study, which means more research is required to verify the effects of these factors. Besides, it also needs to be remembered that the results might have been affected by other factors. For example, the special component of "口(kǒu) [mouth]" appears in four of the five words for which the highest scores were achieved, which could have affected incidental word learning in this research. Related research, therefore, is also required to identify other factors not addressed in this chapter.

In addition, it is quite interesting to discover that, apart from being formed with no more than 16 strokes, all of the five words with the highest mean total scores (see Table 8.3) have other features that might help to reduce learning difficulties, as discussed in this chapter. For example, four of them have repeated components, four of them have no more than six components in the whole word and four of them have only two or three components in the first character. It is, therefore, reasonable to believe that many factors work together to reduce the difficulty of a word. This result partially supports the idea of using "complexity" as an index for differentiating difficulty of learning, although the focus was on character when proposed by Hayes (1987), as mentioned in Section 3.3.2 above. This point could also be tested in future studies.

# Chapter 9: Conclusion

This chapter, which concludes the study, is divided into four subsections. Section 9.1 comprises a summary of this study, detailing the key findings related to the three research questions. After that, a consideration of the pedagogical implications for L2 Mandarin Chinese teaching is given in section 9.2, followed by a description of the limitations of this study in section 9.3. Lastly, in section 9.4, this chapter ends with remarks about the whole study.

## 9.1 Summary of findings

The literature review in Chapter 2 and 3 clearly indicated the dearth of research on L2 Mandarin Chinese in the field of incidental vocabulary learning through reading, as well on the effects of sound information in the annotation of previously unknown words. This study thus comprised an initial attempt to answer three interrelated questions regarding the effects of annotations, specifically sound or sound-related information, on words in the context of incidental vocabulary acquisition through reading in L2 Mandarin Chinese. The first question was whether such annotations helped L2 Mandarin Chinese learners to incidentally acquire sound-related word knowledge, as well as other types of knowledge relating to the character form of words through reading, and how much knowledge was gained in this process. The second question related to the effects of sound information provided in different modalities (Pinyin and audio), which was addressed by comparing three types of annotations in the second experiment, while the last research question related to the participants' attitudes towards different types of annotation.

Two experiments were conducted to seek answers to these research questions. The first one was carried out in a pen-and-paper environment, comparing the effects of text-only and text + Pinyin annotations on 25 beginner's level students, who were randomly assigned to the control and treatment groups. With the self-developed online computer reading program, the audio annotation, which provided the pronunciation of a word, could be added to the second experiment, and the effect of

the annotations, namely the text + Pinyin, text + audio and text + Pinyin + audio annotations, which provided different types of sound information, were consequently compared with 41 beginner's level L2 Mandarin Chinese learners.

Although the materials were presented in different ways at two points in time, the incidental research design was used for both experiments, that is, the participants in both groups were unaware of the impending vocabulary posttests. At beginning of both experiments, the participants were only asked to complete a reading comprehension exercise and the vocabulary posttests came unexpectedly after that. Information on the participants' language backgrounds and their feedback on the exercise etc. were collected by means of a questionnaire given to them after the test as an additional tool to screen the participants.

The results of both experiments answered the first research question, namely that various types of knowledge relating to the sound of a word, both receptive and productive, could be acquired by the participants incidentally through reading L2 Mandarin Chinese, although the help of sound-related annotations is required in the reading process.

In addition to often-investigated word meanings and the occasionally-researched association and collocation of words (see Chapter 2), this finding added evidence with respect to acquiring sound-related word knowledge in L2 incidental vocabulary learning, and thus expanded research topics in this field.

However, different types of sound information used in annotations were compared in the two experiments in this study. The only statistically significant difference found between the control and treatment groups was in the meaning-based Pinyin production test in the first experiment. There were no significant differences between the Pinyin and audio annotation in other types of vocabulary tests or even in the participants' attitudes towards them. However, it is notable that Pinyin annotation tends to help with acquiring not only sound-related knowledge, but also knowledge relating to the character form of words. The results of the second experiment indicate that the text + Pinyin + audio annotation is the most helpful annotation, compared with the text + Pinyin and text + audio annotation,

although the difference was not significant. Nevertheless, it should be also borne in mind that the non-parametric statistical tests used in this study might have affected the results, which therefore need to be interpreted with caution.

Another important finding of this research relates to the partial knowledge sensitive criteria. Based on discussions about word knowledge in Chapter 3, measurements of different types of word knowledge were associated with various vocabulary tests. More precisely, the increase of each type of knowledge was, for the first time, further measured by partial knowledge sensitive criteria for L2 Mandarin Chinese words, which was discussed in Section 4.2.4. The comparison between the vocabulary scores in the posttests under and not under the criteria used in the first experiment suggested a significant difference in the results ($F(6.12, 58.5) = 6.12$, $p = 0.03$), indicating the crucial role of such criteria in research into incidental vocabulary acquisition. In addition, with such criteria the amount of increased word knowledge can be more accurately revealed by the percentages of correct answers in the vocabulary posttests. As the type and amount of knowledge gain are closely connected with the vocabulary tests, the details of the relevant findings are summarised on the basis of the vocabulary posttests used in both experiments in Table 9.1 below, with a star indicating sound-related word knowledge.

Table 9.1 Type and amount of word knowledge acquired incidentally in this study

| Vocabulary posttests in the two experiments | Type of knowledge acquired | Word knowledge gain (%) |
|---|---|---|
| Meaning-based Pinyin form production test (Experiment 1 and 2) | Pinyin form (productive) *<br>Link between Pinyin form and word meaning * | 33.8 (Experiment 1)<br>10.87 (Experiment 2) |
| Meaning-based character form production test – handwritten (Experiment 1) | Character form (productive)<br>Link between character form and word meaning | 14.27 (Control group)<br>16.18 (Treatment group) |
| Pinyin input assisted meaning-based character form production test on computer (Experiment 2) | Pinyin form (productive)*<br>Character form (receptive)<br>Link between Pinyin form and word meaning*<br>Link between character form and Pinyin form* | 12.20 |

| | | |
|---|---|---|
| Spoken form recognition test (Experiment 2) | Spoken form (receptive)*<br>Link between spoken form and word meaning* | 17.4 |
| Character-based Pinyin form production test (Experiment 2) | Link between character form and Pinyin forms*<br>Character form (receptive)<br>Pinyin form (productive)* | 10.67 |
| Character form recognition test (Experiment 2) | Character form (receptive)<br>Link between character form and word meaning | 19.40 |

*sound-related word knowledge

It should be noted that the matching test included in the first test was not included in the table, because it was difficult to decide whether the word knowledge was obtained from incidental learning or from the process of completing the test (see Section 5.3.1 for a detailed discussion of this). In addition, as suggested by the results of the first experiment, the Pinyin annotation might help learners to acquire not only sound-related word knowledge, but also knowledge relating to the character form of words. The results of the related vocabulary posttests were therefore also included in the table.

Acknowledging that the frequency with which the target words were encountered was low in this study, especially in the second experiment (1 or 2 times), the amount of knowledge demonstrated in the table above contributes to understanding the initial stage of acquiring words through reading. In this case, by adopting similar designs, with words appearing at different frequencies, enough information might be gradually accumulated through further studies to illustrate the process of learning a word. However, it should be borne in mind that the set of partial knowledge sensitive criteria used in this research was just an initial draft, which requires further investigation and improvement. It should also be noted that the partial knowledge sensitive criteria will differ with different target languages.

Turing to the effects of sound information provided in annotations, the results of the comparison between the text-only and text + Pinyin annotation in the first experiment, outlined in 5.2, showed a significant difference between the control

group (without access to Pinyin annotation) and the treatment group (with access to Pinyin annotation) in the meaning-based Pinyin form production test ($U = 28.00$, $z = 2.897$, $p < 0.02$, $r = - 0.58$). It is easy to understand the finding of the first experiment about the Pinyin annotation being significantly more helpful for acquiring word knowledge of the Pinyin form on a production level. However, it should also be noted that the positive effect of the Pinyin form on assisting the treatment group to acquire more knowledge in the matching and meaning-based character form production tests was also observed in the first experiment by the higher scores attained by the treatment group than the control group, although the differences were not statistically significant. This finding indicates the possibility of sound-related information being helpful in acquiring knowledge related to the character form and the meaning of words. This phenomenon was also reported by Everson and Ke (1997) and Everson (1998), as noted in chapter 3. However, the difference between the control and treatment group was not significant in the current study and therefore requires further research.

In the comparison of the text + Pinyin, text + audio and text + Pinyin + audio annotation in the second experiment, no significant difference was found in the statistical tests. However, the in-depth analyses suggested that the text + Pinyin +audio annotation had a better effect on incidental vocabulary acquisition than the other two types. It was unexpected that no significant difference was found in this study. However, research on incidental vocabulary learning in L2 Mandarin Chinese is still in its infancy and many factors could have affected the results, especially those relating to Chinese character recognition, which has never been investigated in similar research. Although it was not the primary objective to identify such factors, the preliminary findings in Chapter 8 will help future researchers to choose target words. Moreover, according to the analysis on data collected in this study, three factors are very likely connected to word learning, namely: the number of strokes in a word, and the number of components in the first character and in a word, as well as whether a word contains repeated components. It seems that words with no more than 16 strokes and repeated components, with a low number of components in the

word, might be easier to acquire in the process of incidental learning. Lastly, regarding attitudes towards the different types of annotations, the participants regarded word meaning to be significantly more useful than both Pinyin ($Z = -3.917$, $p = 0.00$) and audio ($Z = -4.303$, $p = 0.00$). The attitudes to the Pinyin and audio annotation varied, as suggested by the distribution of the marks chosen by the participants, but they were not significantly different. This finding will help material developers to decide which types of annotations to present in their material and also indicates how audio information should be presented. The latter will be dealt with further in the next section.

## 9.2 Pedagogical implications

The findings of this study indicate some areas that are worth considering in classroom teaching. Firstly, they draw teachers' and material developers' attention to incidental vocabulary acquisition through reading. As the results of this study indicate, various types of word knowledge, over and above word meaning only, can be acquired incidentally by reading annotated articles. From this perspective, reading could be a better resource for vocabulary learning than expected and the learning results might be further improved if proper assistance is provided. For example, regarding the types of word knowledge attained, if teachers or material developers add corresponding exercises to increase task-induced involvement (Laufer & Hulstijn, 2001), the learning of words could be enhanced. Moreover, in the context of CALL, providing both Pinyin and audio annotation at the same time might be more helpful than providing Pinyin or audio annotation alone, because the results of the second experiment showed a tendency of the Pinyin + audio annotation to be more helpful in incidental vocabulary learning. It is notable that audio annotation has been widely adopted in online programs. However, no matter whether it is in an online reading program or an online dictionary, audio annotation is often available only when a "play" button being clicked. As audio information about a word is considered significantly less important by participants, as indicated in their answers to the attitude questions, it is very likely that when they understand the meaning of the word, readers might

ignore the audio information attached. Unlike these programs, the Pinyin and audio annotations for the target words showed spontaneously in this research without an extra click being required. It is therefore recommended that the auto played audio information be embedded in word annotation, together with Pinyin annotation.

Apart from using auto-played audio annotation, the discussions about the design and the creation of the online program might benefit L2 Mandarin Chinese material developers. Based on the structure of the website used in this research, more reading materials with different types of annotations can be easily added into the program.

Secondly, the discussions on the question of what knowing a word meant in Chapter 3 could be useful for teaching and learning Mandarin Chinese as a second/foreign language. Similarly, the discussion on word knowledge, vocabulary tests and criteria that are sensitive to partial knowledge could be useful for evaluating vocabulary learning more accurately, and hence help instructors understand how word knowledge is gradually obtained. When designing the current study, I struggled to find any references to evaluating the results of incidental vocabulary acquisition. It is hoped that the discussions of these issues in this study could be used as a starting point for future studies.

In addition, the finding that the matching test worked better as a word learning tool than as a testing tool would add one more useful vocabulary exercise to the Mandarin Chinese classroom, as discussed in Section 5.3. For learners who have received instruction on the phonetic component and mastered a number of basic phonetic components, the matching test could be used to expand their vocabulary by taking advantage of the large number of characters containing such components, as pointed out by previous researchers, including Li and Kang (1993), Zhang (2007) and Pan (2004) (see Section 3.3.2 for more details). In addition, when designing such exercises, due to the similarity between the Pinyin form and the English equivalent, the meaning components could also be used as clues for learners to answer such questions and consequently learn new words. However, it should be noted that there is a possibility of learners being affected by the phonetic components and

remembering an inaccurate pronunciation of a character, as discussed in Section 7.3. Therefore, the words used in the exercises need to be carefully selected.

Lastly, noticing the finding that Pinyin input might help learners to obtain higher scores in the meaning-based character form production test, as discussed in Section 7.3, typing with Pinyin input could be utilised as an exercise for new word learning. For example, learners could be asked to type a word in the character form, with both the character and Pinyin form provided if it is an unknown word. This type of exercise would enhance the learners' word knowledge, especially that relating to word character form and Pinyin form. In addition, this would also help to increase the production of words in character form but with the difficulty of handwriting reduced.

## 9.3 Limitations and suggestions for future research

This study has some limitations that readers are advised to bear in mind when interpreting the findings. First of all, due to the small number of learners who took part in this study, especially in the first experiment, the statistical results need to be interpreted with caution. Secondly, the statistical tests employed were, in most cases, non-parametric tests, and a contradiction was found in the statistical result when analysing the participants' attitudes to different types of annotations in the first experiment. This also means that the results should be read with caution. Thirdly, only immediate vocabulary posttests were used in this study and not the delayed vocabulary posttest that have been employed by many researchers in previous studies. There is one thing that needs to be addressed in relation to this point: it is well accepted that language learners will forget some words as time goes by if they do not have further exposure to or a chance to practise them. In this case, by simply presenting the percentage of knowledge gain derived from the immediate and delayed posttest without providing any in-depth analysis, researchers have simply repeatedly confirmed this well-known fact. A more valuable research question might be to identify what the knowledge loss is by comparing the answers to the immediate and delayed posttests. Such a comparison could lead to a deeper understanding of vocabulary knowledge retention and help to improve vocabulary teaching and

learning as a consequence. Lastly, regarding testing the participants' pre-knowledge of the target words, although the participants were asked to provide the details of what they knew about the words at a partial knowledge level, it might be more accurate if the components were also included in the pretest.

Reflecting on the design of this research, several issues are worth addressing in future studies. Firstly, in this study the types of word knowledge were connected to vocabulary tests to evaluate the results of incidental vocabulary acquisition. By doing so, vocabulary learnt via such a process could be understood better. For example, the results of the tests requiring the Pinyin form to be produced in both experiments, as well as the handwritten character form to be produced in the first experiment, demonstrated that incidental vocabulary acquisition goes beyond the receptive level; knowledge attained in this process can assist learners to write the target words in both the Pinyin form and the character form. This result is consistent with the findings of Bowles (2004) and Yanguas (2009) in L2 Spanish learning, who inadvertently demonstrated productive knowledge gain in incidental vocabulary acquisition. While there might be some advantages to evaluating the results of incidental vocabulary acquisition from this perspective, the reality is that there is no guidance on how a vocabulary test can be connected to types of word knowledge; it was just an attempt by this researcher to do so. It is hoped that this attempt will be treated as the starting point for establishing an evaluation system for incidental vocabulary acquisition, and future studies could contribute to this.

Secondly, partial knowledge of words, a concept referring to more elaborate knowledge related to different types of word knowledge, was employed to evaluate word knowledge gain. The crucial effect of criteria that are sensitive to partial knowledge was evident in this research. In addition, using such criteria produced the finding, for example, that learning the final part of the Pinyin was affected by the phonetic components and the English equivalent, which is similar to the Pinyin of the target words. However, such criteria, especially in relation to the forms of target words, were not included in previous studies, as discussed in Section 3.2. It might be a

good idea to conduct some research into this issue, so that word knowledge gain in the context of incidental vocabulary acquisition can be understood better.

Thirdly, in the second experiment three articles were used with the intention of accommodating more target words in the reading material and to keep the articles similar to the ones used in the first experiment. However, the participants thus needed to complete the comprehension questions for all the articles. This required more time to complete the comprehension section, which was unnecessary because reading comprehension was not the objective of this study. It might be more practical to use a longer article and present the vocabulary test directly after the reading.

In addition, the finding of "丝绸 (sīchóu) [silk]" and "耐克 (Nàikè) [Nike]" being easier in both the first and the second experiment makes the selection of target words worth further consideration, as some word features could affect learning difficulty in L2 Mandarin Chinese. While several word features mentioned before, e.g. conceptual difficulty and part of speech (Laufer, 1990; Lin, 2010; Nagy et al., 1987; Paribakht, & Wesche, 1997) do not completely account for the ease of learning these two words, further analyses on identifying possible influential factors were thus conducted in Chapter 8, starting with those relating to Mandarin Chinese character recognition. It is suggested that the number of strokes of a whole word, the number of components of a word, and whether a word contains repeated component could possibly affect incidental word learning in this research. More precisely, words that had no more than 16 strokes, had a small number of components in the first character and the whole word, and had repeated components could be easier. In addition, a combination of such factors was also noticed in the target words with high scores in this study, which is consistent with Hayes' (1987) view that the complexity of character affects learning, although his focus was on characters rather than words. Based on these preliminary findings, future studies are required to further verify the factors suggested above, as well as to identify factors that were not included in this research. It is hoped that research into such influential factors will benefit research into incidental vocabulary learning in L2 Mandrin Chinese in terms of selecting target words with a clear guideline.

Apart from the factors intensively discussed in this research, the differences between the traditional Chinese characters and the simplified ones could form another affecting factor. Although both sets of Chinese characters were adopted, they were used respectively in the two experiments, with the participants in the first experiment having started their learning with the traditional Chinese characters and the participants in the second experiment only having had experience of learning the simplified Chinese characters. Therefore, no comparison was made on this point in this study. However, it would also be interesting to know whether the results from learners who learn traditional characters first would be different from the results of those who learn simplified ones first.

In addition, in the second experiment, in which the audio annotation was included, no statistically significant differences were found between the three types of annotations. A possible reason for that could be that no comprehensive tests were done on the spoken forms of the words provided by the audio annotation, because they were tested only by means of a recognition posttest. It would be more comprehensive if the spoken form of a word was tested at the productive level as well. It might also be worth taking learners' learning styles into consideration, because annotations in differing modalities might interact with such learning styles.

Lastly, the design of the online program for this study included a tracking tool and log file. It was, however, used only to confirm the times of each annotation being clicked on by the participants. As a matter of fact, there was some other information included in the log file, such as the time they spent reading each article, completing the vocabulary pretest, posttest, questionnaire, etc. Such data could reveal more details, although not necessarily about incidental vocabulary acquisition but about L2 learning in CALL if carefully designed research were conducted.

## 9.4 Concluding remarks

Reading has been considered an important resource for expanding vocabulary in L2 learning. The aim of this study was to find ways to help L2 Mandarin Chinese learners at the beginner's level to gain more from vocabulary learning through

reading by investigating the effect of various types of annotation that provide sound-related word information on incidental vocabulary acquisition through reading. This approach has, to the best of my knowledge, not been adopted by previous researchers in the field of incidental vocabulary acquisition. While many researchers have compared the effects of a wide range of types of annotations, the effect of sound information on a word has rarely been paid any attention, even in the context of CALL, in which word pronunciation in the form of audio annotation can be, and has easily been, added to many language learning programs.

The results of this research provided preliminary evidence on some key points relating to incidental vocabulary learning in L2 incidental vocabulary learning from many perspectives, including: 1) selecting L2 Mandarin Chinese as the target language in this study contributes a further example to the literature, in which this language has rarely been investigated; 2) incidental learning of many types of knowledge relating to the sounds of words was identified; 3) with the self-developed online reading program, the effects of sound information provided in differing modalities were compared. There is tendency for the word + Pinyin + audio annotation to be more helpful in terms of helping participants acquire new words in reading, although no significant difference were found; 4) the criteria used for vocabulary tests influence the results of incidental vocabulary learning, because a significant difference was demonstrated in this study between the scores of the posttests marked under partial knowledge sensitive criteria and those marked under criteria that are not sensitive to partial knowledge; 5) by initially adopting the partial knowledge sensitive criteria and associating different types of word knowledge with vocabulary tests, the percentage of word knowledge gain was clearly reported in this research under the circumstance in which target words were encountered at a low word frequency; and 6) the further analysis conducted in Chapter 8 suggested that further investigation on factors that might affect learning results in similar research could start from three character-recognition related factors, namely the number of strokes in a whole word, the number of components in the first character and in a whole word, and repeated components in a word.

The first three points mentioned above expand the research boundary of L2 incidental vocabulary learning through reading in terms of types of word knowledge and target language. Points 4 and 5 will help with the development of standards for conducting research in this field so that results from different studies might be brought together and analysed to identify further features of incidental vocabulary learning in L2. As for the last point, further investigation of factors particularly relating to L2 Mandarin Chinese is required in terms of helping establish standards for selecting target words in future research.

Some findings of this research could also be used to improve teaching L2 Mandarin Chinese. Firstly, as well as learning word meanings, reading materials should also be considered a good resource for acquiring sound-related information about words, although relevant annotation is needed. Secondly, two types of exercises could be involved in vocabulary learning. One is the matching exercise, which is often treated as a tool to evaluate word learning, and the other is a typing exercise, which could be conducted in a CALL environment. In addition, with a deeper understanding of the connection between word knowledge and vocabulary tests, as well as an awareness of the partial knowledge of a word, learners' progress in vocabulary learning could thus be described more elaborately, and more appropriate exercises and learning targets could consequently be assigned.

Now that contributions made to the field of researching L2 incidental vocabulary learning through reading and pedagogical implications have been briefly summarised, the points that need to be further researched are as follows: apart from refining the framework of word knowledge in L2 Mandarin Chinese and improving the partial knowledge sensitive criteria, a couple of points about the methodology employed in this research are also worth considering in future studies. It could be useful to employ a delayed vocabulary posttest to identify the percentage of retention of word knowledge acquired in this process and, more importantly, to identify what the lost word knowledge is. In addition, affecting factors related to Mandarin Chinese words need to be identified and included in future studies. Furthermore, including a word pronunciation test in the online reading program seems to more in line with audio

annotation. Related to this point, learners' preferences with respect to different types of annotation could be considered as an affecting factor in relevant studies. Lastly, using tracking tools and conducting more analysis on tracking data could be considered by future researchers.

# Reference


Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis, *Computer Assisted Language Learning,* 21(3), 199-226.

AbuSeileek, A. F. (2008). Hypermedia annotation presentation: Learners' preferences and effect on EFL reading comprehension and vocabulary acquisition. *CALICO Journal*, 25(2), 260-275.

Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon*, 4th ed., Chichester: John Wiley & Sons.

Akbulut, Y. (2007). Effects of multimedia annotations on incidental vocabulary learning and reading comprehension of advanced learners of English as a foreign language. *Instructional Science*, 35(6), 499-517. doi: 10.1007/s11251-007-9016-7

Alessi, S., & Dwyer, A. (2008). Vocabulary assistance before and during reading. *Reading in a Foreign Language*, 20(2), 246-263.

Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: a comparative study. *Language Learning and Teaching*, 5(1), 202-232.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. Guthrie (Ed.), *Comprehension and Teaching: Research Reviews* (pp. 77-117). Newark, DE: International Reading Association.

Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. Hutson (Ed.), *Advances in reading/language research* (Vol. 2). Greenwich, CT: JAI Press, Inc.

Barcroft, J., & Rott, S. (2010). Partial Word Form Learning in the Written Mode in L2 German and Spanish. *Applied Linguistics*, 31(5), 623-650. doi: 10.1093/applin/amq017

Beatty, K. (2003). *Teaching and Researching Computer Assisted Language Learning*, New York: Longman.

Bourgerie, D. S. (2003). Computer Aided Language Learning for Chinese: A Survey and Annotated Bibliography. *Journal of the Chinese Language Teachers Association*, *38*(2), 17-47.

Bowles, M. A. (2004). L2 glossing: To CALL or not to CALL. Hispania, 541-552.

Bruton, A., López, M. G., & Mesa, R. E. (2011), Incidental L2 Vocabulary Learning: An Impracticable Term?. *TESOL Quarterly*, 45,759–768.

Chang, L., & Li, J. K. (2009). A correlational study of metacognitive strategies, listening comprehension and incidental vocabulary acquisition. *Foreign Language World*, 6, 50-59.

Chen, C., & Huang, X. (1999). Research on characteristics of visual recognition to symmetrical structural Chinese characters. *Acta Psychologica Sinica, 31(2), 154-161.*

Chen, Y., & Peng, R. (1994). Chinese character recognition: name and connection model. In H. Cheng, J. Huang, C. Wei, & J. L. Ovid, & T. Zeng (Eds.), *Advances in the study of Chinese language processing.*

Cho, K. S., & Krashen, S. (1994). Acquisition of vocabulary from the Sweet Valley Kids series. *Journal of Reading,* 37, 662-667.

Chun, D. M. (2006). CALL technologies for L2 reading In L. A. Ducate, N. (Ed.), *Calling on CALL: From theory and research to new directions in foreign language teaching* (pp. 69-98). San Marcos, TX: CALICO.

Chun, D. M., & Plass, J. L. (1996) Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal* 80 (2), 183-198.

Chung, K. H. (2002) Effective Use of Hanyu Pinyin and English Translations as Extra Stimulus Prompts on Learning of Chinese Characters, *Educational Psychology*, 22 (2), 149-164.

Coady, J. M. (1993). Research on ESL/EFL vocabulary acquisition: Putting it in context. In T. Huckin, Haynes, M., & Coady, J. (Ed.), *Second Language Reading and Vocabulary Learning*, (pp3-23). Norwood, NJ: Albex.

Cohen, L., Manion, L., & Morrison, K. (2007) *Research methods in education (6th ed.).* London, New York: Routledge.

Cook, V. J., & Newson, M. (1996). *Chomsky's universal grammar: An introduction (2nd ed.)*. Cambridge, MA: Blackwell Publisher Inc.

Craik, F. I. M., & Lockhart, R. S. (1972) Levels of processing: A frame work for memory research. *Journal of Verbal Learning and Verbal Behavior,* 11, 671-684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104, 268-294.

Crow, J. T., & Quigley, J. R. (1985). A semantic field approach to passive vocabulary acquisition for reading comprehension. *TESOL Quarterly*, 19(3), 497-513.

Day, R., Omura, C., & Hiramatsu, M. (1992). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language*, 7, 541-551.

De Ridden, I. (2002) Visible or invisible links: Does the highlighting of hyperlinks affect incidental vocabulary learning, text comprehension, and the reading process? *Language Learning & Technology*, 6(1), 123-146.

Daller, H., Milton, J., & Treffers-Daller, J. (Eds.). (2007). *Modelling and assessing vocabulary knowledge* (p. 150). Cambridge: Cambridge University Press.

Douglas, D. (2010). Book review: Understanding Language Testing.

Duan, S., & Yan, S. (2004) Effects of multiple choice annotation of English vocabulary incidental learning, *Foreign Language Teaching and Research,* 36(3), 213-218.

Dupuy, B., & Krashen, S. (1993). Incidental vocabulary acquisition in French as a foreign language. *Applied Language Learning, 4*(1), 55-63.

Eckerth, J., & Tavakoli, P. (2012) The effects of word exposure frequency and elaboration of word processing on incidental L2 vocabulary acquisition through reading. *Language Teaching Research*, 16 (2), 227-252.

Eller, G., Pappas, C. C., & Brown, E. (1988). The lexical development of kindergartners: Learning from written context. *Journal of Reading Behaviour*, 20, 5-24.

Elley, W. B. (1991). Acquiring literacy in a second language: the effect of book-based programs. *Language Learning*, 41(3), 375-411.

Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition, 21*(02), 285-301.

Everson, M. (1998). Word recognition among learners of Chinese as a foreign language: Investigating the relationship between naming and knowing. *Modern Language Journal*, *82*(2), 194-204.

Everson, M. E., & Ke, C. (1997). An inquiry into the reading strategies of intermediate and advanced learners of Chinese as a foreign language. *Journal-Chinese Language Teachers Association*, *32*, 1-20.

Eysenck, M. W. (1982). Incidental vocabulary learning and orienting tasks in C. R. Puff (Ed.). *Handbook of Research Methods in Human Memory and Cognition*. New York: Academic Press, 197-228.

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th edition). London: Sage publications.

Fisher, T., Sharples, M., Pemberton, R., Ogata, H., Uosaki, N., Edmonds, P., & Tschorn, P. (2012). Incidental second language vocabulary learning from reading novels: A comparison of three mobile modes. *International Journal of Mobile and Blended Learning (IJMBL)*, *4*(4), 47-61.

Folse, K. S. (2004). *Vocabulary myths*. Ann Arbor: University of Michigan Press. p26.

Fraser, C. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition,* 21, 225-241.

Gai, S. (2003). A Review of Incidental Vocabulary Acquisition. *Journal of PLA University of Foreign Languages,* 26(2), 73-76.

Gao, D., Zhong, Y., & Zeng, L. (1995). Research on how frequency influences the recognition speed of common Chinese characters. *Psychological Science*, 4, 225-229.

Gao, Z., & Liu, H. (2009). A review of incidental vocabulary learning with glosses outside China. *Journal of Southwest University for Nationalities (Humanities and Social Science)*, 5,272-277.

Gass, S. M., & Selinker, L. (2001). *Second language acquisition: An introduction.* Mahwah: Lawrence Erbium.

Granick, J. S. (1997). Learning word meanings of low frequency words from context during natural reading. *Dissertation Abstracts International, 58 (03A).* (University Microfilms No. AAI9723793)

Gu, Y., & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning,* 46(4), 643-679.

Hanban (2011a). Sample Test Paper for Level One, H11004. [online] Chinese Testing Net. Available: http://www.chinesetest.cn/godownload.do [Accessed 25 December 2014]

Hanban (2011b).Vocabulary list for the new HSK: level one to six (2011 version) [online] Chinese Testing Net. Available: http://www.chinesetest.cn/godownload.do [Accessed 20 December 2011]

Hayes, E. B. (1987). The relationship between Chinese character complexity and character recognition. *Journal of Chinese Language Teacher Association,* 22 (2), 45-57.

Haynes, M. (1998, March). Word form, attention and vocabulary development through reading. Paper presented at the Annual Conference of the American Association for Applied Linguistics, Seattle, WA.

He, H. (2009). An analysis of lexical errors in non-English majors' writing: A corpus-based study. *Foreign Language World*, 3, 2-9.

Henricksen, B. (1999). Three dimensions on vocabulary development. *Studies in Second Language Acquisition,* 21(2), 303-317.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a Clockwork Orange: acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11(2), 207-223.

Huang, H. S., & Hanley, J. R. (1995). Phonological awareness and visual skills in learning to read Chinese and English. *Cognition*, 54, 73-98.

Huang, S., Willson, V., & Eslami, Z. (2012). The Effects of Task Involvement Load on L2 Incidental Vocabulary Learning: A Meta-Analytic Study. *The Modern Language Journal*, *96*(4), 544-557.

Hubbard, P. (2009) (Ed.) *Computer Assisted language learning*, Volumes I-IV, London, UK and New York, NY: Routledge.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 21*(02), 181-193.

Hulstijn, J. H. (1992) Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. Arnaud & H. Bejoint, (eds). *Vocabulary and Applied Linguistics*, 113-125.

Hulstijn, J. H. (2001). Intentional and incidental second-language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In Robinson, Peter (Ed.). *Cognition and second language instruction*. Cambridge: C.U.P., 258-287.

Hulstijn, J. H., Hollander, M., & Greidanus, Y. (1996) Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal,* 80 (3), 327-339.

Hulstijn, J. H., & Laufer, B. (2001), Some Empirical Evidence for the Involvement Load Hypothesis in Vocabulary Acquisition. *Language Learning,* 51: 539–558. doi: 10.1111/0023-8333.00164

Jacobs, G., Dufon, P., & Hong, F. C. (1994) L1 and L2 glosses in reading passages: Their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading* 17 (1), 19-28.

James, B. (2009). Effects of Multimedia Glossary Annotations on Incidental Vocabulary Acquisition in L2 Learners of Japanese (Master of Arts), Brigham Young University.

Jenkins, J., Stein, M., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal*, 21, 767-787.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics, 21*(1), 47-77.

Jiang, X. (2000). A Survey on the Strategies for Learning Chinese among CSL Beginners. *Language Teaching and Linguistic Studies,* 1, 61-68.

Jiang, X. (2003). Relationship between knowing pronunciation and knowing meaning of Chinese characters among CSL learners. *Language Teaching and Linguistic* Studies, 6, 51-57.

Jiang, X. (2004a). The effects of orthography and phonology on reading of Chinese by American and European learners. *International Journal of Psychology*, 39(5/6) (Abstracts of the XXVIII International Congress of Psychology), 284.

Jiang, X. (2004b). The roles of frequency and stroke number of Chinese characters on learning by students from Alphabetic language background. *International Journal of Psychology*, 39(5/6) (Abstracts of the XXVIII International Congress of Psychology), 106.

Jiang, X. (2007). An experimental study on the effect of the method of 'Teaching the learner to recognize characters more than writing'. *Chinese Teaching in the World*, 2, 91-97.

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition. *Applied Linguistics*, 19(3), 357-377.

Kang, Y. D. (2006) *The Effects of Glossing on Incidental Vocabulary Learning by Foreign Language Learners of Different Proficiency Levels*, Fujian Normal University Master Dissertation in Education.

Khatib, M., & Zourzadeh, S. (2012). Incidental L2 vocabulary acquisition through reading comprehension: issues, agreements and controversies, *Journal of Academic and Applied Studies,* 2(2), 1- 11.

Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for learners of different verbal abilities. *Modern Language Journal*, 78, 285-299.

Koda, K. (1997). Orthographic knowledge in L2 lexical processing: a cross-linguistic perspective In J. H. Coady, T. (Ed.), *Second Language Vocabulary Acquisition.* Cambridge: Cambridge University Press.

Kost, C. R., Foss, P., & Lenzini, J. J. (1999). Textual and pictorial glosses: Effectiveness of incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals*, 32(1), 89-113.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the Input Hypothesis. *The Modern Language Journal* 73, 440-464.

Krashen, S. (1985). *The input hypothesis: issues and implications*. New York: Longman.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*(2), 149-174.

Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL*, *20*(03), 271-289.

Laufer, B., & Hulstijn, J. H. (2001) Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics,* 22, 1-26.

Laufer, B., & Nation, I. S. P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16, 307-322.

Laufer, B. (1990). Why are some words more difficult than others? Some intralexical factors that affect the learning of words. *International Review of Applied Linguistics in Language Teaching*, 28(4), 293-307.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian modern language review*, *59*(4), 567-587.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399-436.

Laufer, B., & Hill, M. (2000). What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention?. *Applied Linguistics*, 22(1), 1-26. p3.

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: do we need both to measure vocabulary knowledge? *Language Testing,* 21(2), 202-226.

Laufer, B. (2012). Word difficulty. In C.A. Chapelle (Ed.). *The Encyclopedia of Applied Linguistics.* Boston: Wiley-Blackwell

Lawson, M. J., & Hogben, D. (1996). The vocabulary-learning strategies of foreign-language students. *Language Learning*, 46(1): 101–35.

Lei, L. (2011). The effect of vocabulary enhancement conditions and language level on incidental vocabulary learning. *Foreign Language Research*, 1, 75-79.

Lei, L., Wei, Y., Ye, L., & Zhang, M. (2007). A Study of Incidental Vocabulary Acquisition Through Writing by Chinese EFL L earners. *Journal of PLA University of Foreign Languages*, 30(1), 53-61.

Lenders, O. (2008). Electronic glossing - is it worth the effort?. *Computer Assisted Language Learning*, 21, 457-481.

Levenston, E. A. (1979). Second language vocabulary acquisition: Issues and problems. *Interlanguage Studies Bulletin*, 4, 147-160.

Lewis, M. (1993). *The Lexical Approach*. Hove: Language teaching Publications.

Levy, M., & Hubbard, P. (2005). Why call CALL CALL?. *Computer Assisted Language Learning, 18*(3), 143–149.

Li, H. (2003). Error analysis on non-English major students' writing tasks. [M.A Dissertation]. Dalian: University of Dalian Maritime University.

Li, Y. (2008). The effect of involvement load on incidental learning of productive word knowledge. *Foreign Language Learning Theory and Practice*, 2, 6-9/22.

Liang, H. (2015). The development and prospects of computer-assisted teaching of Chinese as a foreign language. *Journal of International Chinese Teaching*, *4*, 48-56.

Lin, L. (2010). A video-based CALL program for proficient and less-proficient L2 learners' comprehension ability, incidental vocabulary acquisition. *Educational Media International*, *47*(3), 199-216.

Li, Y., & Kang, J. (1993). Research on phonetic components of characters in Modern Chinese. In Y. Chen (Ed.), *Analyses on Characters in Modern Chinese* (pp. 74-87). Shanghai: Shanghai Education Press.

Liu, H., & Zhao, M. (2012). Research on hot spots of computer assisted language learning between China and overseas based on keyword frequency analysis. *Oversea Chinese Education*, *4*, 352-360. doi:10.14095/j.cnki.oce.2012.04.001

Liu, L. X. (2007). *Elementary Chinese Reading and Writing Course (two).* Beijing: Peking University Press.

Llach, M. P. A. (2011). *Lexical errors and accuracy in foreign language writing*. Multilingual Matters.

Lomicka, L. L. (1998). "To gloss or not to gloss: an investigation of reading comprehension online", *Language Learning & Technology,* 1(2),41-50.

Lu, J. J. (1984). Interlanguage theory and phonetic errors made by L2 Chinese learners. *Language Teaching and Linguistic Studies,* 3, 44-56.

Luo, Y., Wang, P., Li, X., Shi, Y., Chen, M., Wang, P., Hu, S., & Luo, Y. (2010). The Effect of Character's Whole Recognition on the Processing of Components in the Processes of Chinese Characters. *Acta Psychologica Sinica, 42(6), 683-694.*

Luo, Y., Chen, M., & Peng, R. (2007). Influence of whole characters and radicals on stroke-form-character. *Journal of Clinical Rehabilitative Tissue Engineering Research*, 11(35), 7037-7040.

Ma, L. (2016). *Teaching Chinese as a foreign language in Chinese phonetic teaching problems and countermeasures_take Beijing Shuren Ribet private school for example*. [Master of Art]. Anyang Normal University.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions?. *Educational psychologist*, *32*(1), 1-19.

Mayer, R. E. (2002). Multimedia learning. *Psychology of learning and motivation*, *41*, 85-139.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, *38*(1), 43-52.

Mayer, R. E., & Johnson, C. (2008). Revising the redundancy principle in multimedia learning. *Journal of Educational Psychology*, 100, 380-386

Mayer, R. E. (2009). *Multimedia learning (2nd ed*). New York: Cambridge University Press.

McGraw, I., Yoshimoto, B., & Seneff, S. (2009). Speech-enabled card games for incidental vocabulary acquisition in a foreign language. *Speech Communication*, *51*(10), 1006-1023.

Meara, P. (1980). *Vocabulary acquisition: A neglected aspect of language learning.* Paper presented at the Language teaching and linguistics: Abstracts.

Meara, P. (1984). *The Study of Lexis in Interlanguage*. Edinburgh: Edinburgh University Press.

Meara, P. (1990). A note on passive vocabulary. *Second Language Research*, 6(2),150-154

Meara, P. (1996). The vocabulary knowledge framework. *Vocabulary Acquisition Research Group Virtual Library,* p4.

Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language testing*, *4*(2), 142-154.

Meara, P., & Milton, J. (2005). X_Lex: The Swansea Levels Test [CD-ROM]. *Newbury, UK: Express Publishing*.

Melka, F. (1997). in Schmitt, N. (2000) Vocabulary in Language Teaching. New York: Cambridge University Press.

Miyasako, N. (2002). Does text-glossing have any effects on incidental vocabulary learning through reading for Japanese senior high school students?. *Language Education & Technology*, 39, 1-20.

Moir, J., & Nation, I. S. P. (2002). Learners' use of strategies for effective vocabulary learning. *Prospect,* 17(1), 15-35.

Moreno, R., & Mayer, R. (1999). “Cognitive principles of multimedia learning: The role of modality and contiguity”. *Journal of Educational Psychology*, 91, 358–368.

Nagata, N. (I999). The effectiveness of computer-assisted interactive glosses. *Foreign Language Annuals*, 32(4), 469-479.

Nagata, N. (2000). *Japanese Courseware for Distance Learning.* Proceedings of the Twelfth World Congress of International Association of Applied Linguistics (AILA), 2000.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24(2), 237-270.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985) Learning words from context. *Reading Research Quarterly,* 20, 233-253.

Nan, J. (2000). "Lexical Representation and Development in a Second Language". *Applied Linguistics*, 21, 47–77.

Nassaji, H. (2003). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly*, 37 (4), 645-670.

Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. Boston: Heinle & Heinle.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language.* Cambridge: Cambridge University Press.

Nation, I. S. P. (2009). New roles for L2 vocabulary. In C. Vivian & W. Li (Eds.), *Contemporary Applied Linguistics: Language Teaching and Learning (Vol. 1).* York: Cotinuum International Publising Group.

Nation, I. S. P., & Gu, P. Y. (2007). *Focus on vocabulary*. Sydney: NCELTR, Macquarie University.

Newton, J. (2013). Incidental vocabulary learning in classroom communication tasks. *Language Teaching Research*, *17*(2), 164-187. doi:10.1177/1362168812460814

Paivio, A. (1986). *Mental representations: a dual coding approach.* Oxford. England: Oxford University Press.

Pan, J. (2004). *Research on issues relating to modern Chinses characters.* Yunnan: Yunnan University Press.

Paribakht, T. S., & Wesche, M. (1993). Reading comprehension and second language development in a comprehension based ESL program. *TESL Canada Journal,* 11(1), 9-29.

Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin

(Eds.), *Second Language Vocabulary Acquisition: A Rationale f or Pedagogy*. New York: Cambridge University Press.

Paribakht, T. S., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition. *Studies in Second Language Acquisition,* 21,195-224, p176.

Peng, R., & Wang, C. (1997). Basic processing unit of Chinese character recognition: evidence from stroke number effect and radical number effect. *Acta Psychological Sinica*, 29(1), 8-16.

Perfetti, C. A., & Tan, L. H. (1998). The constituency model of Chinese character identification. In: J. Wang, A. Inhoff, & H. C. Chen (eds.), *Reading Chinese script: A cognitive analysis. Hillsdale,* NJ: Erlbaum (in press).

Perfetti, C. A., & Zhang, S. (1991). Phonemic processes in reading Chinese words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17: 633–643.

Perfetti, C. A., & Zhang, S. (1995). Very early phonological activation in Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 24.

Perfetti, C. A., Zhang, S., & Berent, I. (1992). Reading in English and Chinese: Evidence for a “universal” phonological principle. *Advances in psychology*, *94*, 227-248.

Pitts, M., White, H., & Krashen, S. (1989) Acquiring second language through reading: A replication of the Clockwork Orange study using second language using second language acquirers. *Reading in a Foreign Language* (2), 271-275.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology,* 90(1), 25-36.

Qian, X. J. (2003). Research on IVA in Chinese reading. *Journal of Beijing University (Philosophy & Social Science Edition), 4*, 135-142.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10(3), 355-371. doi: 10.1177/026553229301000308

Read, J. (2000). *Assessing vocabulary*. Cambridge, England: Cambridge University Press.

Reynolds, B. L., & Bai, Y. L. (2013). Does the freedom of reader choice affect second language incidental vocabulary acquisition?. *British Journal f Educational Technology*, *44*(2), 42-44. doi:10.1111/j.1467-8535.2012.01322.x

Reynolds, B. L., & Wible, D. (2014). Frequency in incidental vocabulary acquisition research: An undefined concept and some consequences. *TESOL Quarterly, 48*(4), 843-861. doi:10.1002/tesq.197

Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly,* 10(1),77-89.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in second language acquisition,* 21, 589-619.

Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, *6*(2), 72-78.

Schmidt, R. (1990) The Role of Consciousness in Second Language Learning. *Applied Linguistics* 11,129-158.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness. In R. Schmidt (Ed.), *Attention and Awareness in Foreign Language Teaching and Learning* (Technical Report No. 9) (pp. 1-64). Honolulu: University of Hawai'i at Manoa.

Schmitt, N. (1999). The relationship between TOEFL vocabulary items and meaning, association, collocation and word-class knowledge. *Language testing*, *16*(2), 189-216.

Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.

Schmitt, N. (Ed.). (2010). *An introduction to applied linguistics*. London: Routledge Press.

Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework. *Studies in second language acquisition*, 19(1), 17-36.

Service, E., & Craik, F. (1993). Differences between young and older adults in learning a foreign vocabulary. *Journal of Memory and Language*, 32(5), 608-623.

Shahrokni, S. A. (2009). Second language incidental vocabulary learning: The effect of online textual, pictorial, and textual pictorial glosses. *TESL-EJ*, 13(3), 1–17.

Sheng, R. Z. (2011). Listening comprehension and incidental vocabulary under the interaction between metacognitive strategy and multimodality. *Heilongjiang Researches on Higher Education,* 9, 179-182.

Shu, H., Anderson, R. C., & Zhang, H. (1995). Incidental learning of word meanings while reading: A Chinese and American cross-cultural study. *Reading Research Quarterly*, 30, 76-95.

Singleton, D. (1997). Learning and processing L2 Vocabulary. *Language Teaching*, 30, 213-225.

Song, G. (2002). A general review of the second language vocabulary acquisition research abroad. *Language Teaching and Linguistic Studies*, 1, 54-60.

Song, L. (2000). 'An Investigation into the Feasibility of Emphasising Reading and De-emphasising Writing Characters in Teaching Chinese.' *Selected Papers for the Sixth International Symposium on Teaching Chinese as a Foreign Language*, 114-119. Beijing: Beijing University Press

Su, P. C. (1992). Brief introduction on modern Chinese character research. *Chinese Character Research and Regulation*, 7, 6-11.

Sun, X. M. (2007). A Review of Second Language Vocabulary Acquisition Research. *Language Teaching and Linguistic Studies*, 7, 54-62.

Swain, M., & Lapkin, S. (1995). "Problems in Output and the Cognitive Processes They Generate: A Step Towards Second Language Learning." *Applied Linguistics*, 16, 371-391.

Swanborn, M. S. L., & de Glopper, K. (1999). Incidental Word Learning while Reading: A Meta-Analysis. *Review of Educational Research,* 69(3),261-285

Tan, L., & Peng, R. (1991). Visual recognition processes of Chinese characters: a research to the effect of grapheme and phoneme. *Acta Psychological Sinica*, 3, 271-278.

Tan, L. H., & Perfetti, C. A. (1998). Phonological codes as early sources of constraint in Chinese word identification: A review of current discoveries

and theoretical accounts. *Reading and Writing: An Interdisciplinary Journal*, 10, 165–200.

The National Language Committee. (1997). *GB.13000.1 character set for information processing-- specification on Chinese character components*.

Todd, R., & Tepsuriwong, S. (2008). Mobile mazes: Investigating a mobile phone game for language learning. CALL-EJ Online, 10(1), 1–16.

van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensional approach. *System*, *41*(3), 609-624.

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61(1), 219-258.

Wan, Y. (2003). A synthetic study of Chinese character and its cognition. Language Teaching and Linguistic Studies, 25(2), 72-79.

Wan, Y. (2012). The role of pinyin transcription and Chinese characters and their relationship in teaching Chinese as a foreign language. *Chinese Teaching In The World*, *26*(3), 409-418. doi:10.13724/j.cnki.ctiw.2012.03.010

Wang, D. (2004). Sixty basic questions on Chinese character. Beijing: China Peace Publishing House.

Wang, H. (2005). The effects of subtitles on incidental vocabulary learning. *Computer-assisted Foreign Language Education,* 2, 92-94.

Wang, H. (2007). A comparison between text and video material on L2 acquisition. *Journal of Suzhou University (Philosophy & Social Science),*4,92-94.

Wang, X. (2012). Incidental vocabulary learning through English Movie and TV programme. *Teaching and Management*, 15, 117-118.

Wang, Y., Yao, Y., & Xu, Y. (2012). A comparative study of L2 incidental vocabulary acquisition: listening vs reading. *Foreign Languages and Their Teaching,* 6, 1-5.

Waring, R., & Nation, I. S. P. (2004) Second language reading and incidental vocabulary learning. *Angles on the English speaking World*, 4, 97-110.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader. Reading in a Foreign Language, 15(2), 130-163.

Watanabe, Y. (1997). Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition* 19 (3), 287-307.

Webb, S., Newton, J., & Chang, A. (2013). Incidental Learning of Collocation. *Language Learning,* 63(1), 91-120.

Wen, H. (2006). *Comparisons on errors from English major students' compositions*. [M.A Dissertation]. Chongqing: Xinan Normal University.

Wesche, M., & Paribakht, T. S. (Eds.) (1999). Incidental L2 Vocabulary acquisition: Theory, current research, and instructional implications [Special issue]. *Studies in Second Language Acquisition*, 21(2).

Wesche, M., & Paribakht, T. S. (2000). Reading-based exercises in second language vocabulary learning: An introspective study. *The Modern Language Journal*, 84, 196-213.

Wittrock, M. C. (1974). Learning as a generative process. *Educational Psychologist*, 11(2), 87-95.

Wu, J., Lang, J., & Duan, Q. (2007). Incidental vocabulary learning and the Involvement Load Hypothesis. *Foreign Language Teaching and Research, 5, 360-366+401.*

Wu, J., Yu, Q., & Wu, H. (2005). Statistic on Some Chinese Character Stroke Information. *Journal of Anhui University (Natural Sciences)*, 29 (3), 14-20.

Wu, W., & Xu, H. (2009). Effects of Different Types of Annotations on Incidental Vocabulary Learning in a Multimedia Environment. *Journal of PLA University of Foreign Languages,* 1, 44-49.

Xiao, X. (2002). An analysis of character errors of foreign learners of Chinese. *Chinese Teaching in the World*, 2, 79-85. doi: 10.13724/j.cnki.ctiw.2002.02.012

Xu, D. (2002). Incorporating multimedia in Chinese language teaching. Invited talk at Language Center, Union College, NY, May 15, 2002.

Xu, D. (2015). Issues in CALL studies. *Journal of Technology and Chinese Language Teaching*, *6*(2), 1-16.

Xu, P. (2006). Multimedia and Internet technology in teaching Chinese reading comprehension to foreigners. In P. Zhang, & T. Xie (Eds.), *Research on Digitalized*

*Chinese Teaching and its Application* (pp. 227-234). Beijing: Language and Literature Press.

Yang, H., & Hou, F. (2012). An Empirical Study of Output and IVA in Listening, *Shandong Foreign Language Teaching Journal,* 5, 5 3-58.

Yang, H., Gui, S., & Yang, D. (2005). *Analysis of English learner basing on CLEC*, Shanghai: Shanghai Foreign Language Education Press.

Yanguas, I. (2009). Multimedia glosses and their effects on L2 text comprehension and vocabulary learning. *Language Learning and Technology,* 13(2), 48-67.

Ye, Y. (2002). Problems in Developing Students' Writing Strategies: An Investigation of Syntactical and Lexical Errors from Students' Compositions. *Foreign Language Education*, 23(4), 77-81.

Yeh, Y., & Wang, C. (2003). Effects of Multimedia Vocabulary Annotations and Learning Styles on Vocabulary Learning. *CALICO Journal*, 21(1), 131-144.

Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology*, 10(3), 85-101.

Yoshii, M., & Flaitz, J. (2002). Second language incidental vocabulary retention: The effect of text and picture annotation types. *CALICO Journal*, 20(1), 33-58.

You, H. (2003). How is the acquisition of Chinese characters by learners from the non-Chinese character cultural sphere affected by stroke numbers, component numbers and topological structural patterns, *Chinese Teaching in the World, 2, 72-81.*

Yao, T. C. (2003). The current situation and future of Internet Chinese teaching in the United States. Paper presented at the 3rd International Conference on Internet Chinese Education, Taipei, Taiwan, Oct. 24-26, 2003.

Yan, X., & Wang, H. (2013). A review of computer assisted language learning and its application in Chinese classrooms. *Journal of International Chinese Studies*, 4(2), 59-68.

Zhang, J. (2007). Characteristics of the role of sound and grapheme of the Chinese characters in reading for foreign student with different mother languages. *Journal of South China Normal University,* 2, 119-123.

Zhang, J., & Zhang, H. (2001). The relationship between the wholes and their parts in recognition of Chinese characters. *Chinese Journal of Applied Psychology,* 7(3),57-62.

Zhang, W., & Feng, L. (1990). The visual recognition and capacity STM for Chinese disyllabic words. *Acta Psychologica Sinica,* 4, 383-390.

Zhang, W., & Feng, L. (1992). A study on the unit of processing in recognition of Chinese character. *Acta Psychologica Sinica*, 24(4), 379-358.

Zhou, H. (2008). A review on empirical studies on L2 acquisition in China. *Journal of Hunan Agricultural University (Social Sciences),* 9(4), 96-98.

Zhu, C., & Cui, H. (2005) Re-probing of incidental Vocabulary Learning in Chinese Reading, *Journal of Jinan University Chinese School*, 2, 16-22.

Zhu, Y. (2004) The effect of modified input methods of marginal glosses and dictionary-consulting on the incidental vocabulary learning by foreign students, *Chinese Teaching in the World*, 70(4),74-84.

Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition*, 27(4), 567-595.

Zheng, Y. (2014). On three basic issues about the realization of multimedia Chinese teaching. *TCSOL Studies*, *53*, 55-61.

Zhou, J. (2004). *Discussions on structures of Chinese vocabulary.* Shanghai: Shanghai Lexicographic Publishing House.

Zhou, X., & Chen, M. (1998). Census on regulations on Chinese characters. *Computer World*, *15*, 1-3.

# Appendix

## Appendix A: Reading materials

### Reading material for the first experiment

For the treatment group (with Pinyin annotation)

這雙耐克是用那塊絲綢換的

有個人到一家商店買絲綢， 想做窗簾。一塊絲綢，本來賣二百五十元，他說太貴了，後來，售貨員同意 兩百元賣給他。可是給錢的時候，他想：“我的窗簾還是新的，還能用很長時間”。就對售貨員說：“我開始的時候想買這塊絲綢做窗簾。現在我改變主意了。你給我一雙價值兩百元的耐克吧。”

售貨員拿出一雙耐克來，這個人拿著鞋就走。售貨員說：“你還沒給我錢呢！”

“付什麼錢？這雙耐克是用那塊絲綢換的。”

“可是你沒有給我絲綢錢啊！”

“奇怪！我沒拿你的絲綢，怎麼還要給你錢？”

sīchóu
絲綢：Silk

gǎibiàn
改變 ：change

jiàzhí
價值：value

nàikè
耐克：Nike

pànduàn
判 斷 ： judge

判斷對錯 （If it is wrong, correct it.）

1. 買絲綢的人付了兩百五十元錢。
2. 他開始的時候想買絲綢做窗簾，後來改變主意了。
3. 這個人用他的絲綢換了售貨員的一雙耐克。
4. 他應該給售貨員兩百五十元錢。

如果你是這個售貨員，你會怎麼回答？

Reading articles for the second experiment

## Article one

### 这双鞋是用那块布料换的（236 characters）

有一个人去商店里买东西，看到店里的布料很好，可以做衣服。有一块丝绸他很喜欢，但是太贵了，就问服务员能不能打个折扣，后来，服务员同意两百元一块卖给他。可是要给钱的时候，他又想，自己的衣服都是新的，还能穿很长时间，就对服务员说："我现在还不想做新衣服，不要这个了，我看你们的鞋也不错，给我换一双价值两百元的耐克鞋吧！"服务员拿出鞋来，这个人拿着鞋就要走。服务员说："你还没付钱呢！"。

"什么钱？这双鞋是我用那块布料换的。"

"可是你也没有给我那块布料钱啊！"

"奇怪！我没有拿你的布料，为什么要给你钱！"

朋友们，请你们判断一下，谁对谁错呢？

## Article Two

### 午睡（212 characters）

你午睡吗？你知道什么是健康的午睡吗？

调查说明：只有近 30%的人午睡，半数以上的人认为，不一定要午睡，实在太累了才有必要小睡一下。生理学上，睡眠分为深睡眠和浅睡眠。一般人在睡着 80-100 分钟后，从浅睡眠进入深睡眠，代谢减缓。如果在这时起来，就会出现眩晕、眼花、无力等症状。所以，健康的午睡应该在 30-60 分钟左右，尽量不要进入深睡眠。午睡时间最好在早上起床后的 7-8 个小时间。还有，吃饱午饭后不能马上午睡，应该稍微等一会儿，化消化一下，10 分钟后再睡。

## Article three

### 生活的故事 （208 characters）

有一个小孩对母亲说：“妈妈，您今天真漂亮。″妈妈问：“为什么？″小孩说：“因为妈妈今天没有生气。″

原来漂亮很简单，只要不生气就可以了。

有个农民,常常叫他的孩子在地里工作。朋友对他说：“你不要让孩子这么辛苦，庄稼一定也会长得很好的。″他回答说：“我培养的不是庄稼,是我的孩子。″

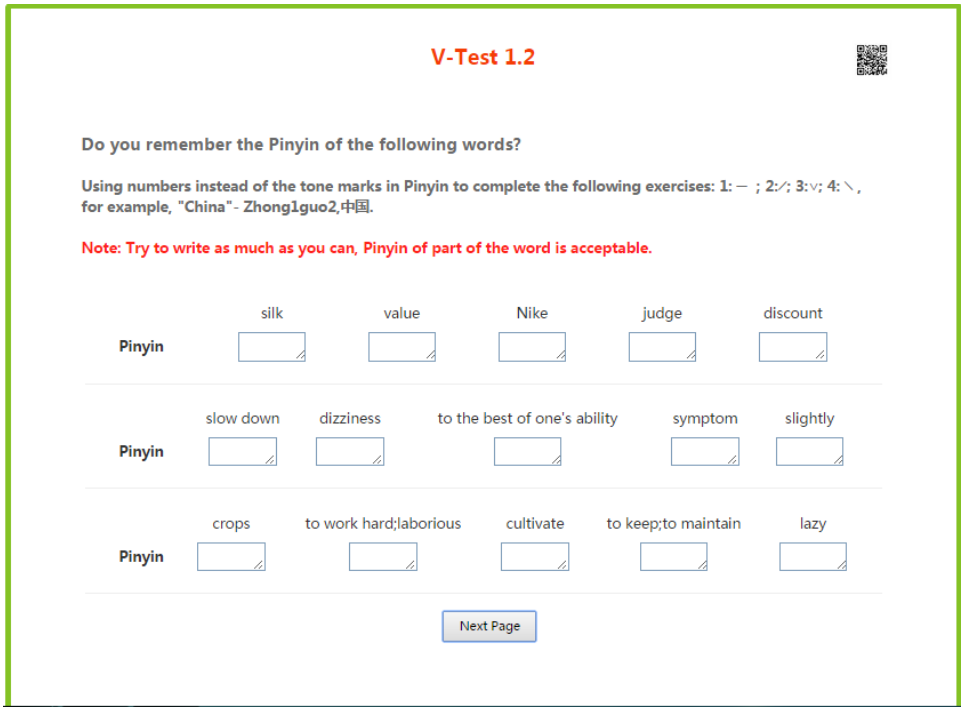
原来培养孩子很简单，让他辛苦一点儿就可以了。

有一个商店一直灯火通明，有人问：“你们店里用什么牌子的灯泡？用那么长时间？″店主回答说：“我们的灯泡也经常坏，只是坏了就马上换。″

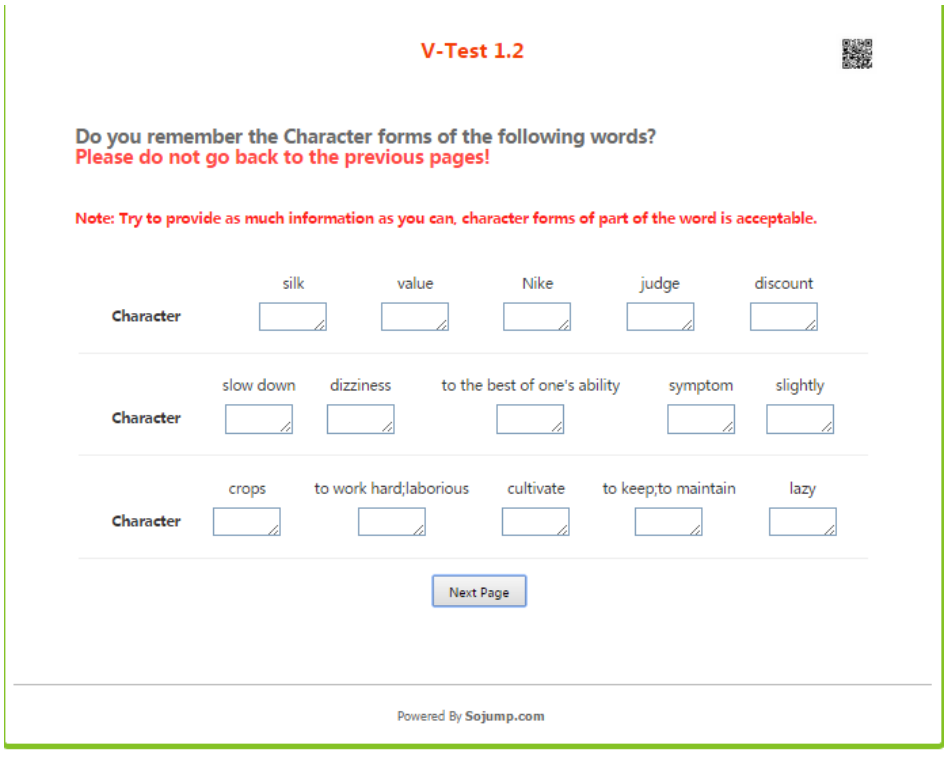
原来保持明亮很简单，只要不懒惰就可以了。

# Appendix B: Screenshots of the vocabulary posttest

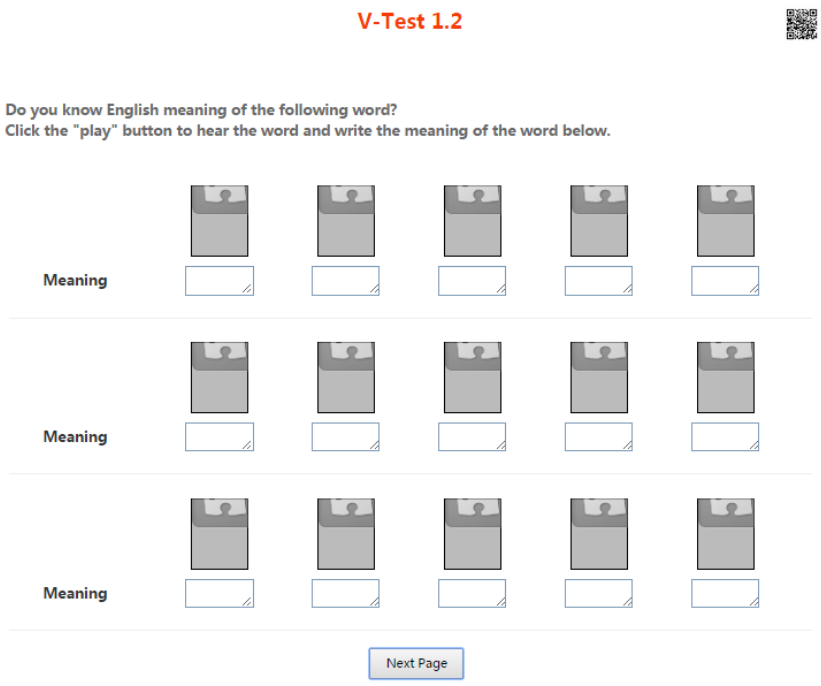
1) Meaning based Pinyin production test



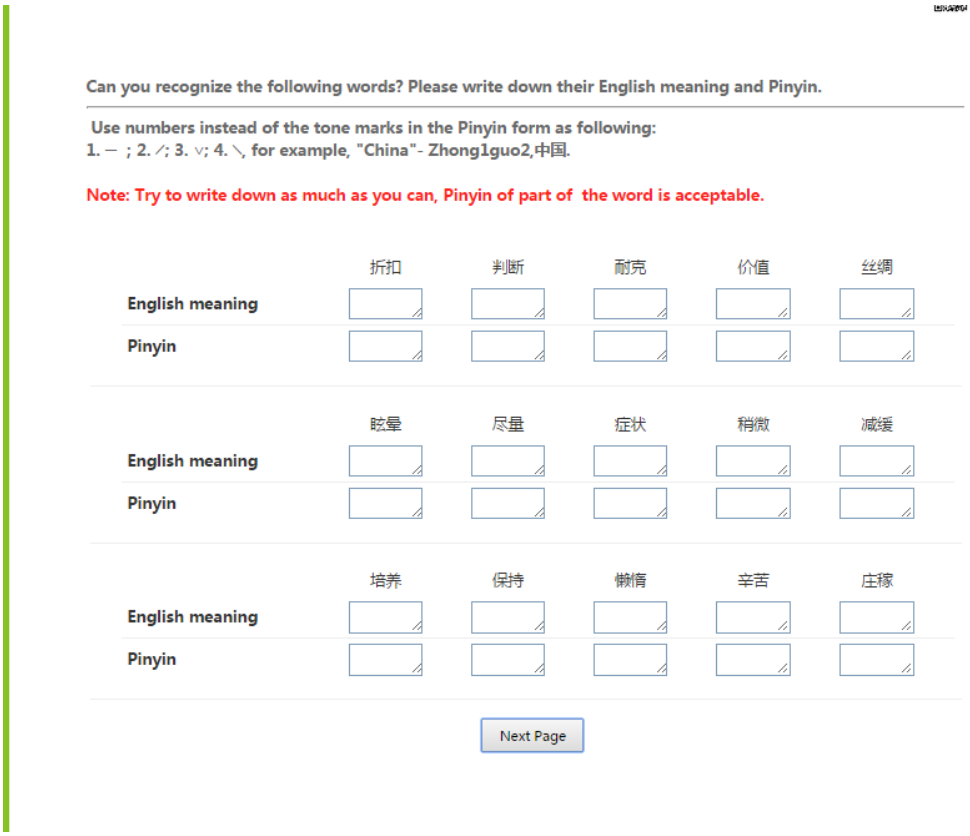

2) Meaning based character production test



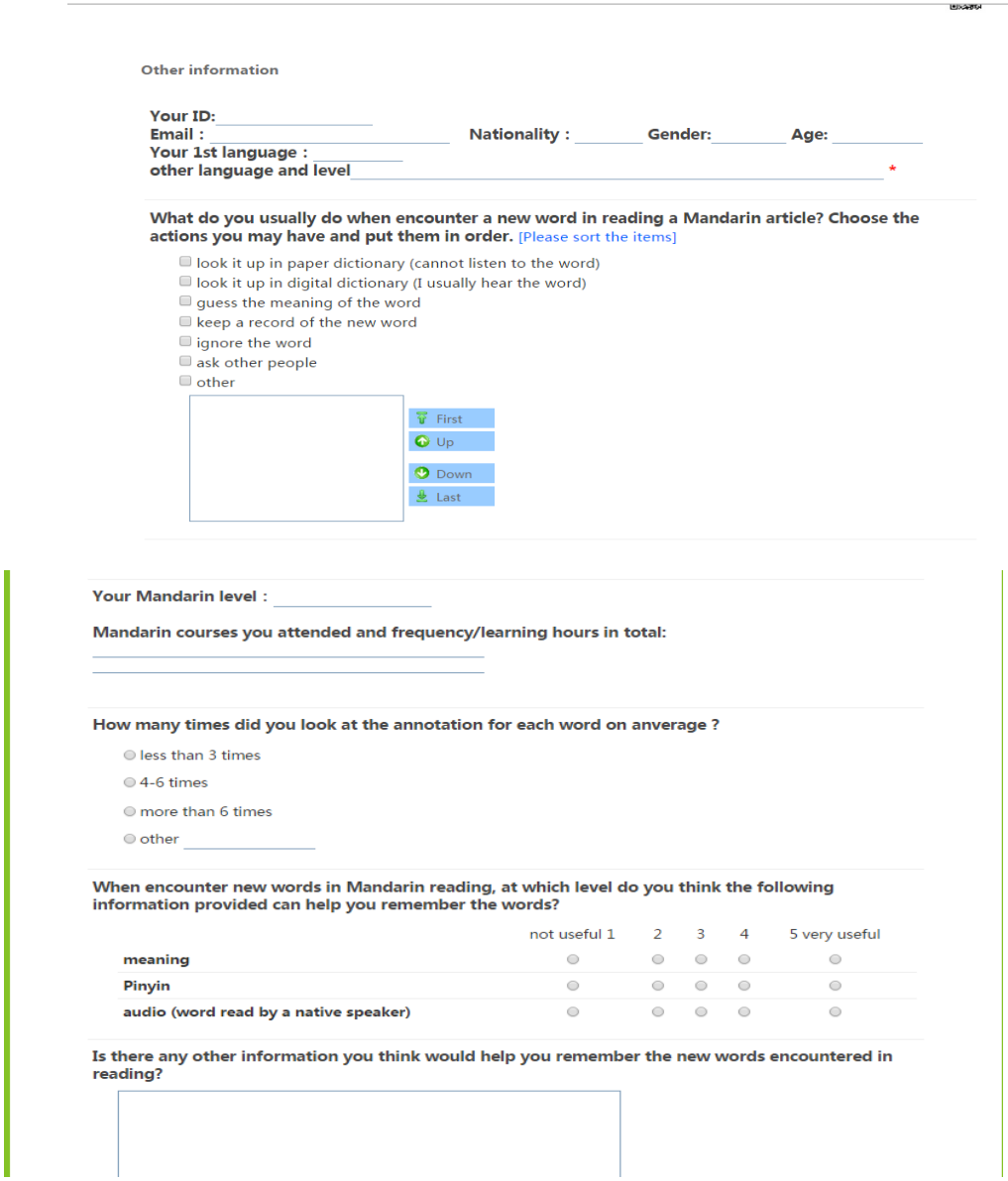

3) Audio recognition test




4) Character based Pinyin production and character recognition test (4a,4b)



# Appendix C: Screenshots of the questionnaire



Other information

**Your ID:** _____
**Email :** _____ **Nationality :** _____ **Gender:** _____ **Age:** _____
**Your 1st language :** _____
**other language and level** _____ *

**What do you usually do when encounter a new word in reading a Mandarin article? Choose the actions you may have and put them in order.** [Please sort the items]

- ☐ look it up in paper dictionary (cannot listen to the word)
- ☐ look it up in digital dictionary (I usually hear the word)
- ☐ guess the meaning of the word
- ☐ keep a record of the new word
- ☐ ignore the word
- ☐ ask other people
- ☐ other

First
Up
Down
Last

**Your Mandarin level :** _____

**Mandarin courses you attended and frequency/learning hours in total:**

_____
_____

**How many times did you look at the annotation for each word on anverage ?**

- ○ less than 3 times
- ○ 4-6 times
- ○ more than 6 times
- ○ other _____

**When encounter new words in Mandarin reading, at which level do you think the following information provided can help you remember the words?**

| | not useful 1 | 2 | 3 | 4 | 5 very useful |
|---|---|---|---|---|---|
| **meaning** | ○ | ○ | ○ | ○ | ○ |
| **Pinyin** | ○ | ○ | ○ | ○ | ○ |
| **audio (word read by a native speaker)** | ○ | ○ | ○ | ○ | ○ |

**Is there any other information you think would help you remember the new words encountered in reading?**

**Which kind of learner you think you are? Please write down the percentage of each item if you choose more than one.** [Multiple]

- ☐ Visual learner _____
- ☐ Auditory learner _____
- ☐ Kineathetic learner _____
- ☐ Other _____

**In which way you type Chinese characters in this exercise?**

- ○ I type the characters one by one. E.g. type "ni" ,choose the character"你",and then type "hao", choose character"好".
- ○ I type the whole word together. E.g. type "nihao" and choose "你好"
- ○ I mixed the two ways above.
- ○ Other _____

**Do you have any suggestions/comments on this exercise ?**

Appendix D: Written components in the target words (the first experiment)

| | 絲綢 silk | 判斷 judge | 耐克 Nike | 價值 Value | 改變 change |
|---|---|---|---|---|---|
| compoonent 1 | 糸 | 丷 | 而 | 亻 | 己 |
| compoonent 2 | 糸 | 丰 | 寸 | 西 | 攵 |
| compoonent 3 | 糸 | 刂 | 十 | 貝 | 糸 |
| compoonent 4 | 冂 | 幺 | 口 | 亻 | 言 |
| compoonent 5 | 士 | 幺 | 儿 | 十 | 糸 |
| compoonent 6 | 口 | 一 | — | 且 | 攵 |
| compoonent 7 | — | 幺 | — | — | — |
| compoonent 8 | — | 幺 | — | — | — |
| compoonent 9 | — | 乚 | — | — | — |
| compoonent 10 | — | 斤 | — | — | — |