The Measure Word Usage of Chinese Heritage Speakers

WANG Yu-xiang

University of California Davis, Davis, The United States

Measure words (MWs) are special in Chinese compared with English, and Chinese heritage speakers in the United States are also special. This study combines these two special groups to find the specialties of MW usage of Chinese heritage speakers and the reasons. Linguistic data were collected from four Chinese heritage speakers through topic-based interviews and the story-telling (narrative). Data are coded with both linguistic and social constraints. Mixed-model and crosstab analyses were done through Rbrul. MW usage type is the dependent variable while MW type, noun frequency, speech styles, family language, friend language, and gender are considered as independent variables. The mixed-model analysis shows both linguistic variables (MW type and noun frequency) and one social variable (family language) are significant, while other variables might not be. The crosstab analysis is consistent with the mixed-model analysis except for the noun frequency factor; this may be caused by the fact that high-frequency nouns in Chinese are usually modified by *ge*. However, further large-scale research is needed to prove this hypothesis. These two different analyses indicate that language environment has greatly affected both the MW usage and the MW choice of Chinese heritage speakers.

Keywords: measure word, *ge*, Chinese heritage speaker, Rbrul, variable

Introduction

Measure Words in Chinese

Measure words (MWs) are very special in the Chinese language compared with English. Unlike English, in which MWs are seldom used when counting things, MWs are always used in Chinese. For example, we can say (1) *three dogs* in English. But in Chinese, we must use a measure word between "*three*" and "*dog*", which is " *(zhī)*"; the phrase (1) in Chinese is (2) "*sān zhī gǒu (   )*". Chinese and English also share some MWs, such as (3) *a cup of water* (*yì bēi shuǐ* (, *bēi* is the measure word, whose meaning is *cup*)) and (4) *a group of people* (*yì qún rén* (, *qún* is the measure word, whose meaning is *group*)). In spite of some similarities, the differences are still very great, especially when we express the number plus a countable noun.

There are three types of MWs in Chinese: MWs for nouns (such as   *(zhī)*); MWs for action (such as *xià ( )*); and MWs for time (such as *tiān ( )*). The noun MWs are the most complicated type among these three groups and most of them do not have counterparts in English. Therefore, this study will focus on noun MWs. In the follow sections, MW will only refer to the noun MW.

The distribution of Chinese MWs is not even. "*Ge* can be used to modify almost every individual noun, no matter whether it has its specific measure word or not" (ZHU, 1982, p. 49). FAN (2007) counted the nouns that

WANG Yu-xiang, Chinese instructor, Ph.D. candidate, Department of Linguistics, University of California Davis.

can be used with *ge* in the noun corpus of *Dictionary of Modern Chinese Grammar Information* and found that of 2,240 nouns, 1,516 (67.7%) can be combined with *ge*, and of these, 666 (44%) can be combined only with *ge*.

Chinese Heritage Speakers

Chinese heritage speakers in the United States are a very special linguistic group. They can sometimes speak Chinese well. However, their primary language is English, rather than Chinese. Most of them acquire Chinese through their parents and only learn Chinese in Chinese school on weekends during elementary school, and seldom receive formal Chinese language education. Therefore, their Chinese language usage may be very different from Chinese native speakers.

The Research of the Language Use of Chinese Heritage Speakers

There is some research focusing on Chinese heritage learners. For example, the University of California Los Angeles (UCLA) Center for World Languages has conducted many studies on heritage learners, including Chinese heritage learners. Some other research achievements can be found in the book edited by HE and XIAO (2008). However, very few studies focus on the specific linguistic features of Chinese heritage speakers.

On the other hand, there have been many studies on the errors of MW usage made by Chinese second language (L2) learners. Based on a research done by Z. M. LI (2011), there are two categories of error types of Chinese L2 learners: incorrect usage and incorrect order. He also finds that the errors made by English-speaking learners are similar to those errors made by all L2 Chinese learners of various origins.

However, I have not found any research focusing on the MW usage of Chinese heritage speakers. Chinese heritage speakers can often understand and speak Chinese, some and can even perform at a high level. However, they have their own Chinese language systems, which may be very different from those of native Chinese speakers.

Because of these unique characteristics of Chinese MWs and Chinese heritage speakers, a study of the MW usage of Chinese heritage speakers will offer us a glimpse of the difference in language use between Chinese heritage speakers and native Chinese speakers.

Methodology and Informants

Informants

The participants of this study are four Chinese heritage students (two male and two female) from the Chinese 1 Bilingual class in the department of East Aisan Languages and Cultures (EALC), University of California Davis (UC Davis) in fall 2012. They had studied Chinese at UC Davis for about six weeks before my interview. Before this formal language education, all of them only studied Chinese when they were in elementary school and did not learn Chinese in their secondary education. As a result, their Chinese language learning at UC Davis has not yet strongly affected their language use.

Interviews

Four informants were interviewed individually. All four Interviews were conducted in my office, Sproul Hall 324, due to quiet environment. In order to make them feel comfortable, we sat around a table which is close to the window and the interviewees sat opposite the window. Before each interview, we chatted for a while to become familiar with each other. I told them my research aim is to compare the language usage between Chinese

heritage speakers and Mainland Chinese people, but did not tell them that my research only focuses on measure-word usage, so they would not pay much attention to the measure words during the process of the interview. The interviews were recorded by a SONY voice recorder. In order to make them not pay attention to the voice recorder, I plugged a mic into the voice recorder, and the mic was pinned on their collars. Each interview lasted for about 50-60 minutes. After the interviews, I transcribed the data based on what I would use—the sentences or phrases which contain MWs.

Each of the interviews in fact includes two parts: the interview part and the narrative part. The interview part is topic-based. The interviewees and I talked based on some topics which I designed in advance (see the Appendix). All of the topics are relevant to the interviewees' daily life. In the narrative part, interviewees first saw the video "Pear Story", and then recounted that story. Before they saw this video, I told them that I had not seen it yet and wanted them to tell me the story in as much detail as possible.

Coding and Variables

The aim of this study is to find the MW usage of Chinese heritage speakers, and explain the reasons. As a result, the MW usage of the informants is considered as a dependent variable. It includes two values—correct and incorrect. The data are coded for linguistics and social constraints. Linguistic variables include MW types and the frequencies of the nouns used with MWs. Social variables include speech styles, family language, friend language, and gender.

MWs are divided into three types, which are the general one (*ge*), the specific ones (MWs except *ge*), and the empty one (no MW was used) (see Table 2). There are only three empty measure word instances, all of which were produced by the two female interviewees:

(5) *   (*sān qiānbǐ. Three pencils*) (UC1)

(6) *   (*nà sān ge háizi zǒu dào nà shù de shíhou. When those three children got close to that tree…*) (UC1)

(7) *   (*dàgài shíèr chuānghu. About 12 windows.*) (UC2)

When being closely observed, (6) in fact can be excluded, because *nà shù* can be said in many situations in Chinese, and this sentence is also somewhat acceptable. However, (5) and (7) are not acceptable in Chinese. After instance (6) is removed, there are only two empty instances; each of the female interviewee produced one. These two instances were produced when they were trying very hard to count things, and they got the number while forgot to say the MW. Compared to the 460 MW tokens collected for this research, and also compared to these two interviewees' MW production (109 of UC1 and 90 of UC2; see Table 2), these two instances can be counted as slips in speaking. Therefore, in the following parts, this study will only focus on the non-empty MW usages. Then we have to factors in the MW type: general and specific.

Nouns' frequencies are decided by the Frequency List of Modern Chinese Corpus (The State Language and Writing Commission of China). Based on studies done by K. ZHANG (1997), Q. H. LI (1999), Nation (2001), and Schmitt (2008), the frequency of nouns is divided into three segments: high frequency words (1-2000[1]), middle frequency words (2001-4000) and low frequency words (4001+). The reason I consider the noun frequency as an independent variable is that people do not learn an MW alone; an MW has to be used with a noun

---

[1] The number here means the frequency sequence of a word while 1 is the highest frequency.

in the Chinese language. Of course, in some contexts, we can omit the noun, but this does not mean that MW can be used without a noun. Therefore, the noun frequency should have an influence on the MW usage.

As have been mentioned above, the data I collected are from both topic-based interviews and narrative recounting. These are two very different speech types. In the interview, interviewees focused more on language form. When they were asked, they had time to think, so they would try to use the right language form to answer. However, in the narrative part, they tried to tell the whole story to me in a detailed manner. Therefore, they did not have time to think much about language form; they focused on content.

Family languages are very important for a language speaker of that language. A child acquires his first language by listening to and speaking with his family. My interviewees' family languages are a little different from each other. The two male interviewees' families speak Chinese only. However, in one female interviewee's family, English is the dominant language, while in the other female interviewee's family, Chinese is the dominant. In fact, we cannot expect that a family in an English-speaking country does not speak any English at home. If they have an English-speaking visitor who cannot speak Chinese, they probably need to speak English. However, when only the family members are at home, they probably do not speak English at all.

What languages one's friends use is also very important to his language acquisition, because he can practice that language with his friend. If one practices more, one can acquire that language better. Among these four students, only the two male students have Chinese-speaking friends. One of them has some friends in Taiwan, and they speak Chinese through Skype every day. Another male student is Buddhist and he has some Buddhist friends who can speak Chinese and he meets with frequently.

Gender was also be coded as social variable, because males and females sometimes use language in very different ways. Therefore, in this study, each gender group includes two informants.

Speaker is coded as a random variable. "Including a speaker random effect takes into account that some individuals might favor a linguistic outcome while others might disfavor it, over and above (or 'under and below') what their gender, age, social class, etc. would predict" (Johnson, 2009, p. 365).

This study does not consider the length of Chinese learning as a social variable. First, these four informants learned Chinese at least six years ago. It can be supposed that they would forget the nuanced part of MW usage if they have not used it for six years, though their learning experiences might still have some influence on their language output. Second, their families and friends are the resources that are most important to them for learning Chinese and keeping their Chinese language competency. The variables I will use can be presented as Table 1.

Table 1

Variables

Dependent variables	MW usage	Correct
		Incorrect
Independent variables	Linguistic variables	MW types
		Noun frequency
	Social variables	Speech styles
		Family language
		Friend language
		Gender
Random variable	Speaker	UC1, UC2, UC3, UC4

Date Analysis and Discussion

Overview of the Data

Four hundred and sixty tokens were collected from my interviews. In these 460 tokens, there are 35 different measure words including the empty one: 33 specific MWs, one general MW (*ge*), and one empty MW. The token distribution is very uneven (see Table 2). First, 85.75% tokens are the general one. Second, different speaker's MW productions are different, from 90 to 134. Third, the correction percentage of specific MW usage is higher than the general one. Fourth, many specific MWs were produced only by one informant. For example, there are nine specific MWs that only were produced by UC4.

Table 2

Token Distribution

MW type	MW	UC1	UC2	UC3	UC4	Total tokens	Correct	Incorrect	Correct percentage (%)
	ba	4	0	2	2	8	8	0	100
	ban	0	0	5	0	5	5	0	100
	bao	1	1	0	1	3	3	0	100
	ben	1	1	1	2	5	5	0	100
	ceng	1	1	1	0	3	3	0	100
	chuan	0	0	0	1	1	1	0	100
	dai	2	0	0	1	3	3	0	100
	ding	0	0	0	1	1	1	0	100
	duan	0	0	0	1	1	1	0	100
	dui	0	0	0	1	1	1	0	100
	feng	1	0	0	1	2	2	0	100
	guan	0	0	2	0	2	1	1	50
	jian	0	0	0	2	2	2	0	100
	ke	0	0	2	1	3	3	0	100
	ke2	0	0	1	0	1	1	0	100
	kou	0	1	1	0	2	2	0	100
	lanzi	1	0	0	1	2	2	0	100
Specific	mian	0	0	0	1	1	0	1	0
	pai	1	0	4	2	7	7	0	100
	pian	1	0	0	0	1	1	0	100
	ping	2	0	0	1	3	3	0	100
	qun	1	0	0	0	1	1	0	100
	shuang	1	2	0	0	3	2	1	66.66667
	tai	0	0	1	2	3	3	0	100
	tang	2	0	0	2	4	4	0	100
	tiao	0	0	0	1	1	1	0	100
	tong	0	1	0	1	2	0	2	0
	tou	0	0	0	1	1	1	0	100
	zhan	0	0	0	1	1	1	0	100
	zhang	0	0	2	5	7	7	0	100
	zhi	2	1	6	10	19	19	0	100
	zhi2	0	2	1	1	4	4	0	100
	zhong	7	0	1	3	11	11	0	100
	total	28	10	30	46	114	109	5	0.95614

(Table 2 continued)

General	ge	79	79	104	81	343	159	184	46.35569
Empty	Empty	2	1	0	0	3	0	3	0
Total Tokens	Total	109	90	134	127	460	268	192	58.26087

R and Rbrul

> R is actually a programming language but you can think of it as a program that you will download onto your computer. …Rbrul is a program written in R. Whenever you want to use it, you will first open Rand then open Rbrul within R. (Tamminga, 2011, p. 1)

Rbrul is very useful in sociolinguistics data analysis. "Rbrul's superior performance is demonstrated on both simulated and real data sets" (Johnson, 2009, p. 359). Rbrul is also very convenient to use; its user interface is very friendly. In this study, I will use both mixed models and crosstabs to explore my data.

Mixed Model Analysis

After having deleted the three empty-MW instances, I first used the Rbrul mixed model to analyze my data with speaker as the random factor to get the best model. Because I only have four informants and 460 tokens, when I analyzed the data, many problems appeared, especially mismatches between the step-up and step-down analyses. After several attempts, I decided to analyze linguistic variables and social variables separately.

The analysis of linguistic variables. When running R with the linguistic fixed factors (MW type and noun frequency) and the random factor (speaker), the best step-up model and best step-down model match each other, which means the best model has been attained:

BEST STEP-UP MODEL IS WITH Speaker [random] and MW.type (7.87e-24) + noun.frequency (0.0456). [*p*-values building from null model].

BEST STEP-UP MODEL IS WITH Speaker [random] and MW.type (2.07e-24) + noun.frequency (0.0456). [*p*-values dropping from full model].

The best step-up model will be used here to analyze the roles of linguistic variables (see Table 3).

Table 3

Best Step-Up Model for Linguistic Variables

	Factors	Logodds	Tokens	Correct/Correct + Incorrect	Centered Factor Weight
BEST STEP-UP MODEL IS WITH Speaker [random] and MW.type (2.07e-24) + noun.frequency (0.0456). [*p*-values dropping from full model].					
MW Type	Specific	1.642	114	0.956	0.838
	general	-1.642	343	0.464	0.162
Noun Frequency	< 2001	0.369	161	0.634	0.591
	2001-4000	-0.290	108	0.537	0.428
	4001 +	-0.079	188	0.574	0.48
Speaker std dev: 0.214					

Deviance *df* intercept grand mean centered input prob
507.612 5   1.439   0.586   0.808

When the speaker is considered as a random factor, the *p*-value of MW type is very small (2.07×10^{-24}), which means MW type is a very significant independent variable of the MW usage. The *p*-value of noun

frequency is 0.0456, which means noun frequency is also a significant independent variable. Therefore, both of the two fixed linguistic variables affect the MW usage of Chinese heritage speakers.

Within the MW type variable, the specific factor has a weight of 0.838, which is much higher than 0.5 (a weight of 0.5 equals no effect). This means that correct usage favors the specific factor considerably. On the other hand, the general factor only has a weight of 0.162, which means that the correct usage disfavors the specific factor considerably.

This result can be explained in this way: When a person does not know the specific MW for a noun, he would probably turn to the general one. That is why so many errors (53.6%) happened to the general one. On the other hand, a person can only use a specific MW after acquiring it. This explains why the proportion of correct usage of specific MWs is so high (95.6%).

Within the noun frequency variable, the high frequency (1-2000) factor has a weight of 0.591, which is slightly higher than 0.5, while the middle frequency (2001-4000) and low frequency (4000+) factors have the weights of 0.428 and 0.48, respectively. This result means that high-frequency nouns are somewhat more likely to be used with the correct measures, while middle- and low-frequency nouns are somewhat less likely to be used with the correct measures.

As mentioned above, an MW needs to be used with a noun, so when a person learns a Chinese noun, he probably learns the MW that is used with that noun. If he has not learned that noun, he usually cannot have learned that MW. High-frequency words are acquired earlier and better, and so too are the MW for these nouns. However, noun frequency in the language environment of Chinese heritage speakers can be very different from China. This may can explain why the noun frequency is a not a very significant variable. To understand this, we will need to investigate the noun frequency of the Chinese language environments of heritage speakers.

The analysis of social variables. After running R with the social fixed factors (gender, family language, friend language, and speech type) and the random factor (speaker), the best step-up model and best step-up model match each other, which also means the best model has been attained:

BEST STEP-UP MODEL IS WITH Speaker [random] and family.language (0.00933).

[*p*-values building from null model]

BEST STEP-DOWN MODEL IS WITH Speaker [random] and family.language (0.00933).

[*p*-values dropping from full model]

The best step-up model will be used here to analyze the role of linguistic variables (see Table 4).

Table 4

Best Step-Up Model for Social Variables

BEST STEP-UP MODEL IS WITH Speaker [random] and family.language (0.00933). [*p*-values building from null model]					
	Factor	Logodds	Tokens	Correct/Correct + Incorrect	Centered Factor Weight
Family Language	Chinese	0.388	261	0.644	0.596
	Chinese = English	0.156	107	0.589	0.539
	English > Chinese	-0.544	89	0.416	0.367
Speaker std dev: 0.317					
Deviance *df* intercept grand mean centered input prob					
605.746 4 0.203 0.586 0.551					

The *p*-value of the family language factor is small enough (0.00933) to indicate that the family language is a significant variable in the MW usage of Chinese heritage speakers. However, the weights of Chinese and Chinese = English factors are only slightly over 0.5 (0.596 and 0.539), which means MW usage somewhat favors these two factors. On the other hand, the English > Chinese factor has a weight of 0.367, which is much lower than 0.5. This means that MW usage disfavors this factor considerably.

Family plays a very important role in the heritage language learning and maintenance (Rohani et al., 2012, p. 2). The current study confirms this result. With families that speak Chinese no less than English, Chinese heritage speakers will use correct measure words more frequently. However, the family language environment alone is not enough to help the speaker have a very high correction level. The difference between the first two factors' weights is only 0.057, which tells us that after above the equal level, the role of family language environment in MW usage will not raise much. However, a family language environment of English more than Chinese has an obvious negative effect on the heritage speakers' MW usage.

This best model does not provide any information as to the roles of the other three social variables (friend language, speech type, and gender). In order to find their roles, we pick out the steps that include these variables (see Table 5).

Table 5

Other Social Variables in the Step-Up Analysis

	Factor	Logodds	Tokens	Correct/Correct + Incorrect	Centered Factor Weight
Friend Language ($p = 0.0548$)	Chinese & English	0.28	261	0.644	0.57
	English	-0.28	196	0.510	0.43
Speech Type ($p = 0.612$)	Interview	0.039	343	0.598	0.51
	Narrative	-0.039	114	0.553	0.49
Gender ($p = 0.0548$)	M	0.28	261	0.644	0.57
	F	-0.28	196	0.510	0.43

The *p*-values of these three variables are larger than 0.05, which means they are not significant. Besides, the weights of the factors included in these three variables are very close to 0.5. Nonetheless, some useful information can still be seen in Table 5.

Examining the friend language variable first, we see that the weights of Chinese & English and English factors are 0.57 and 0.43, which means that having some friends who speak Chinese is better for MW usage than just having English-speaking friends.

Informants used MW correctly in the interview part more often than in the narrative part, because their attention had been paid to different things (form or content).

The male informants' MWs usage is a little better than that of the female informants. As mentioned above, these two gender groups have very different backgrounds, from Chinese learning time to family language and friend language. The different usage between them is much more likely caused by their different backgrounds. However, because of the limited data, we cannot make a conclusion here.

Rbrul Crosstabs Analysis

As having been mentioned above, informants made more errors when they used the general MW (*ge*) than when they used the specific MWs. They used specific MWs only when they had acquired them, so they made

very few errors on using them. In order to find the relation between the specific MW usage and other variables, a crosstab analysis was done through Rbrul (see Table 6).

Table 6

Relation Between the Specific MW Usage and Other Variables

MW Type	MW amount and percentage			Total
	1. Family Language			
	Chinese	Chinese = English	English > Chinese	
Specific	76	28	10	114
General	185	79	79	343
Total	261	107	89	457
	2. Noun Frequency			
	< 2001	2001-4000	4001 +	
Specific	34	29	51	114
General	127	79	137	343
Total	161	108	188	457
	3. Friend Language			
	Chinese & English	English		
Specific	76	38		114
General	185	158		343
Total	261	196		457
.	4. Speech Type			
	Interview	Narrative		
Specific	106	8		114
General	237	106		343
Total	343	114		457
	5. Gender			
	F	M		
Specific	38	76		114
General	158	185		343
Total	196	261		457
	6. MW Usage Type			
	Correct	Incorrect		
Specific	109	5		114
General	159	184		343
Total	268	189		457

Based on the above table, we can make such conclusions:

First, the more Chinese spoken in family, the more specific MWs were produced. The family language environment has a positive effect on the specific MW usage.

Second, having Chinese-speaking friends is helpful in specific MW use. Again, this shows that the language environment is very important.

Third, more specific MWs were used in the topic-based interview than the narrative. Attention can have a positive effect on the specific MW usage.

Fourth, the female informants used less specific MWs than the male informants. This may be also because of the different backgrounds of these two gender groups. The Chinese linguistic backgrounds of the female group

are considerably inferior to those of the male group.

Fifth, the correct usage includes much more specific MWs than the incorrect usage. Informants only tried to use the specific MWs that they had acquired previously. Before they had learned the specific ones, they just used the general one.

Based on the above discussion, we can see that these five variables play the same roles as in the MW usage type. However, the role of noun frequency on specific MW usage is different. Informants used less specific MWs for the high-frequency nouns than for the middle- and low-frequency nouns. Compared with Table 3, in which the percentage of correct MW usage is higher for the high-frequency nouns, we can hypothesize that the high-frequency nouns in Chinese are more likely to be used with *ge*, the general MW. However, because of my data limitation and time limitation, I will not do a further research here.

Conclusion

The MW usage of Chinese heritage speakers is very different from the MW usage in China. Several linguistic and social variables affect the MW usage of Chinese heritage speakers. With Rbrul mixed-model and crosstab analysis, I examined the different roles of variables while considering speaker as a random variable.

Both of the two linguistic variables (MW type and noun frequency) affect MW usage. The correct usage favors specific MWs and high-frequency nouns while disfavors the general one (*ge*) and middle- and low-frequency nouns. This means that the informants in this study were going to use *ge* when they did not know the specific ones, and they only used specific ones that they had acquired. MW learning is connected with noun learning. The high-frequency nouns were learned before the middle- and low-frequency nouns, so my informants' MW usage is better when the MWs were used for the high-frequency nouns.

Not all of the four social variables have significant effects on the MW usage. Only the family language variable is significant ($p < 0.05$). The correct usage favors the Chinese and Chinese = English factors while disfavoring the English > Chinese factor. This means that family language does have an effect on the MW usage of Chinese heritage speakers. Because the weights of the two favored factors are very close, it can be hypothesized that a family language environment with Chinese spoken no less than 50% is very important to the heritage language maintenance.

Different variables also have different effects on MW choice. Chinese language environment (family language and friend language) is very important to specific MW usage; the topic-based interview can elicit more specific MWs; gender may also play a role on specific MW usage. Specific MWs were used correctly because they are usually only used when the speaker has acquired them. Less specific MWs were used with the high-frequency nouns, which may be the result of characteristics of the Chinese language—high-frequency nouns are more likely to be used with *ge*.

Based on the above analysis, we can also find that for Chinese heritage speakers, language environment plays an important role in both MW usage and MW choice.

Sociolinguistics is very useful in studying the language usage of heritage speakers. With Rbrul, we can analyze the variables that affect the language usage of heritage speakers. The data for this study is very limited, so the results of this study may vary when more data are collected.

References

DU, Z. H. (2011). Analysis and countermeasures of foreign language teaching in quantifier error from the perspective of Chinese and English quantifier comparison. *Social Sciences Education, 2,* 13-14.

FAN, Z. Y. (2007). A brief talk on general measure word "ge". *Language teaching and research, 16,* 60-61.

HE, A. W., & XIAO, Y. (Eds.). (2008). *Chinese as a heritage language: Fostering rooted world citizenry*. Honolulu: University of Hawaii Press.

Johnson, D. E. (2009). Getting off the GoldVarb Standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass, 3*(1), 359-383.

Labov, W. (1984). Field methods of the project on language change and variation. In J. Baugh & J. Sherzer (Eds.), *Language in use: Readings in sociolinguistics* (pp. 28-53). Englewood Cliffs, NJ: Prentice-Hall.

LI, Q. H. (1999). The problems of vocabulary size in "the Outline of HSK Vocabulary". *Language Teaching and Linguistic Studies, 1,* 50-59.

LI, Z. M. (2011). *Errors analysis on the noun measure words of English-speaking Chinese learners* (Master's thesis). Fudan University, Shanghai, China.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Rohani, S., Choi, C., Amjad, R. N., Burnett, C., & Colahan, C. (2012). *Language maintenance and the role of the family amongst immigrant groups in the United States: Persian-speaking Bahá'ís, Cantonese, Urdu, Spanish, and Japanese (An Exploratory Study)*. Retrieved from http://www.tc.columbia.edu/centers/cmll/publish/PDFs/LangMaintFamily.pdf

Schilling-Estes, N. (2007). Sociolinguistic fieldwork. In R. Bayley & C. Lucas (Eds.), *Sociolinguistic variation: Theories, methods, and applications* (pp. 165-189). Cambridge: Cambridge University Press.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329-363.

Tamminga, M. (2011). Getting started with Rbrul for the completely clueless: A basic illustrated guide to the quantitative analysis of categorical linguistic variables. Retrieved from http://www.danielezrajohnson.com/Getting_started_with_Rbrul.pdf

The State Language and Writing Commission of China. (2012). Frequency list of modern Chinese corpus. Retrieved from http://www.cncorpus.org

UCLA Center. (2012). Articles for world languages. Retrieved from http://www.international.ucla.edu/languages/heritage languages/journal/

UCLA. (2012). Survey of heritage language teaching. Retrieved from http://www.international.ucla.edu/languages/heritage languages/survey/survey-hl-c.asp

ZHANG, K. (1997). A statistical analysis on basic Chinese word-building characters. *Language Teaching and Linguistic Studies, 1,* 42-51.

ZHANG, Q. (2005). A Chinese yuppie in Beijing: Phonological variation and the construction of a new professional identity. *Language in Society, 32,* 431-466.

ZHU, D. X. (1982). *Lecture notes on grammar* (p. 49). Beijing: Business Press.

Appendix: Topics for Interview

1. Warming-up topics. (How do you go to campus every day? Do you have a car/bike? Do you prefer the quarter system or semester system? Does it rain in Davis? How many days does it rain?)

2. About Halloween: What do people buy for it? How many things? (Pumpkin, Candies, costumes, lanterns) What do people do during this holiday? (Trick-or-treat: how many candies?) What did you do? Describe one Halloween party you attend.

3. Describe our interview environment and count the things I put on the table. (I will put several things on a table, such as one bottle of water, several pens and pencils, some books, some pieces of paper, a piece of clothes, some boxes of gum, a computer, a pair of gloves, some mail, and so on.) (*If interviewees forget some, then I will ask: Do I have X? How many X?*)

4. Describe your personal life. (Such as: family members and their jobs, house size, how many rooms/doors/windows/beds? What do you eat for breakfast/lunch/dinner? How about your room (things and number, such as do you have lamps in your room))?

5. Discuss your classes. (How many courses do you have? Where? Which one do you like the best? Describe it: How many students (From how many countries) take that course? How many chairs does the room have? How much homework/how many

papers do you have for this course and for all of you courses? How many hours do you need to finish your homework every day?)

6. Describe the place you like the most. What kinds of things does that place have? Why do you like it? Does it have many trees? Fruits? Mountains? Rivers? Lakes? Birds?

7. Have you ever been blamed for something you did not do? Did that kind of thing happen several times? Describe what happened. Did you feel sad for a long time? Then what did you do?

8. Describe a zoo you have visited before. How many kinds of animals are there? Tell me the numbers of each kind of animal you saw there. (Or describe the cartoon movie Madagascar (prepare a picture for it).)

9. Which sports do you like? Tell me more about the details. (How many teams/people does it have? How long? How many championships are held in one year?)