Chapter Four

Results and Discussion

This chapter discusses the results of the present study. The answers to the five research questions stated in Chapter One will also be addressed with the findings interpreted. Section 4.1 reports the count-mass distinction observed in the subjects' performance in the two tasks. Section 4.2 examines the age effects found in the subjects' performance. Section 4.3 explores the methodological effects, and section 4.4 presents the hierarchy of difficulty in children's acquisition of count and mass classifiers. Section 4.5 illustrates the semantic feature governing the subjects' misuse of the classifier *ge*. Finally, section 4.6 summarizes the main points of this chapter.

4.1 The Count-Mass Distinction

This section mainly addresses the first research question—to find out whether Chinese children, at early stages of language acquisition, respond differently to the count-mass distinction. The results of all count and mass classifiers will be presented first, followed by the discussion about the subjects' performance difference in their use of count and mass classifiers.

To begin with, Table 4-1 presents our subjects' correct responses to overall count and mass classifiers:

Table 4-1: Subjects' Correct Responses to All Count/Mass Classifiers

Type	Mean	SD
Count Classifiers	0.48	0.16
Mass Classifiers	0.40	0.20

Table 4-1 illustrates that the mean score of count classifiers was 0.48 and that of mass classifies was 0.40. Again, evidently, our subjects generally showed better abilities in dealing with count classifiers. The Paired Samples t-test also confirmed that there was a significant correlation between their performance of count and mass classifiers (p=.000) and that they indeed performed better on count classifiers than on mass classifiers at a significant level (p=.000).

In sum, our subjects, at early stages of language acquisition, had significantly different abilities in their use of count and mass classifiers. That is, they were able to differentiate the grammatical count-mass distinction. Moreover, their way of dealing with count classifiers was better than that of dealing with mass classifiers.

The above findings are expected for the following reason. It has been suggested by various linguists (e.g., Chao 1968, Bach 1989, Chierchia 1998) that Chinese nouns do not encode a count-mass distinction and that all nouns in Chinese are like mass nouns. While indeed Chinese nouns are not marked for the count-mass distinction, this does not mean that the count-mass distinction is not relevant to Chinese nominal syntax. As mentioned in Chapter Two, the view put forward by Cheng and Sybesma (1998, 1999) says that the count-mass distinction is not encoded in nouns, but rather at the level of the classifier. In other words, the enumeration and quantification of Chinese nouns do require the presence of count and mass classifiers.

Accordingly, as Cheng and Sybesma's claim that the idea of cognitive count-mass distinction is indeed relevant in the Chinese grammar and is grammatically encoded in the classifier system, such distinction will emerge in children's language at very young age. That is, young Chinese children, at their stages of language acquisition, are expected to respond differently to the count-mass distinction. The present study, similar to Chien et al. (2003), has indeed provided experimental evidence showing that young Chinese children have knowledge of the

semantic difference between count classifiers and mass classifiers, and therefore supported the linguistic analysis proposed by Cheng and Sybesma.

Interestingly enough, in Chien et al.'s study, though their children performed slightly better on mass classifiers than on count classifiers, yet the difference was not statistically significant ($F(1, 75)=2.46$, $p>.05$). That is, their young children had roughly comparable abilities in dealing with count classifiers and mass classifiers. Nonetheless, the subjects of the present study performed better on count classifiers than on mass classifiers at a significant level ($p=.000$). The difference might lie in the fact that the four mass classifiers adopted in Chien et al.'s experiments were all container measure words, which were easier for children to differentiate. The present study, apart from container measures, also examined children's use of standard measures, partitive measures, and group measures, which posed greater difficulty for children. Furthermore, Chien et al. only conducted two comprehension tasks. The present researcher, on the other hand, conducted both comprehension and production tasks, hence different result[1].

In addition to the subjects' significant count-mass distinction, there is one phenomenon particularly worthy of mention. In Chien et al.'s study, it was found that their children rarely made a "cross-category" classifier-noun association. Scarcely did their subjects use count classifiers to quantify substances or use mass classifiers to enumerate individuated entities (Chien et al. 2003). The present study yielded similar results. That is, classifier-noun associations made by our subjects were mainly "within-category" associations. They were able to use count classifiers to count individuated objects and use mass classifiers to indicate the quantitative measurement of substances. They associated meanings with specific classifiers and applied the

[1] It is generally believed that comprehension is easier than production (e.g. McDaniel et al. 1996, Brown 2000).

specific classifiers generally with only the nouns whose referents belong to the same semantic category as identified by the classifiers. In other words, they were constrained in the choice of a count classifier to be used with a given referent. Seldom did they use a specific classifier non-restrictively with all nouns. This finding might be a reflection of their sensitivity to the nature of the nouns—the grammatical count-mass distinction.

Given that the subjects had better control over count classifiers than mass classifiers and their classifier-noun associations were mainly "within-category," when the subjects' misuses of "cross-category" classifier-noun associations were analyzed, we would naturally expect that they would use count classifiers to replace mass classifiers with a view to enumerating substances. Nonetheless, the findings were contrary to our expectations. It was found that even though our subjects were capable of differentiating the respective functions of count and mass classifiers, they, when making errors in "cross-category" classifier-noun associations, tended to use mass classifiers to quantify individuated objects more frequently than they used count classifiers to enumerate substances. For instance, **yi ping caomei* (*ke*) 'a bottle of strawberries', **yi gongjin/gongfen de qiqiu* (*ge*) 'one kilogram/centimeter of balloon', **yi gongjin de laoshi* (*wei*) 'one kilogram of teacher', **yi wan caomei* (*ke*) 'one bowl of strawberries', **yi wan tusi* (*pian*) 'one bowl of bread.' The classifiers in parentheses refer to the correct usages of count classifiers. The subjects' misuses of mass classifiers and their corresponding numbers of misuses are shown in Table 4-2:

Table 4-2: Subjects' Misuses of Mass Classifiers and Their Corresponding percentage

Type	Misuses	Misuses
M1 standard measures	**yi gongjin de qiqiu (ge)*	0.44
	**yi gongjin de laoshi (wei)*	0.20
	**yi gongfen de qiqiu (ge)*	0.47
	TOTAL	**0.56**
M2 container measures	**yi wan caomei (ke)*	0.22
	**yi wan tusi (pian)*	0.04
	**yi ping caomei (ke)*	0.22
	**yi ping xiaobaitu (zhi)*	0.18
	TOTAL	**0.33**
M3 partitive measures	**yi kuai shanyang (tou)*	0.07
	**yi kuai qianbi (zhi)*	0.20
	**yi pian runiu (tou)*	0.09
	TOTAL	**0.18**
M4 group measures	**yi qun shengzi (tiao)*	0.16
	**yi shuang xian (tiao)*	0.07
	**yi shuang xiaobaitu (zhi)*	0.09
	TOTAL	**0.16**

The subjects' misuses of mass classifiers actually cohere with the argument of Tai (1992, 1994). As mentioned in the literature review, the relationship between a count classifier and an entity denoted by a noun is much more fixed than the relationship between a mass classifier and an entity denoted by a noun. In many cases, different mass classifiers can be used to specify different quantities of the same entity, and the same mass classifier can be used to express the same quantity of many different entities, but count classifiers are not as flexible.

Furthermore, as shown in Table 4-2, when our subjects made "cross-category" errors, their preferred substitution of standard measures for count classifiers was easily detected. The nature of standard measures might contribute to their performance. Standard measures such as *gongjin* and *gonfen* are conventionalized units used to indicate only the weight and length of the following nouns. That is, they mainly serve to denote the unit of measurement of the associated nouns, revealing no intrinsic relation with the properties of these nouns. Therefore, standard measures can be applied to any entities that possess weight, length, area, etc.

To summarize, our subjects, at their early stages of language acquisition, responded significantly differently to the count-mass distinction. They gave better performance on count classifiers than on mass classifiers and, for the most part, they used count classifiers to enumerate individuated entities and used mass classifiers to quantify substances.

4.2 Age Effects

This section primarily concerns the investigation of age differences. We will probe into whether age, the amount of time the subjects were exposed to L1, is a determining factor for their success of comprehension and production of the Chinese count and mass classifiers. First the results of the three age groups will be presented, and then the relationship between the subjects' performance and age differences will be discussed.

The results of the subjects' correct response to each count classifier in the three age groups are shown in Table 4-3:

Table 4-3: Mean Scores of Correct Count Classifiers of Three Age Groups

Type		G1 (3yrs; n=15)		G2 (4yrs; n=15)		G3 (5yrs; n=15)	
		Mean	SD	Mean	SD	Mean	SD
C1-1-1	C1 *ge*	0.57	0.18	0.60	0.21	0.63	0.23
C1-1-2	C2 *wei*	0.17	0.24	0.37	0.23	0.37	0.23
C1-1 Human		0.37	0.29	0.48	0.25	0.50	0.26
C1-2-1	C3 *zhi*	0.30	0.32	0.63	0.35	0.77	0.26
C1-2-2	C4 *tou*	0.20	0.25	0.37	0.23	0.33	0.31
C1-2 Animal		0.25	0.29	0.50	0.32	0.55	0.36
C1 ANIMACY		0.31	0.29	0.49	0.28	0.53	0.31
C2-1-1	C5 *tiao*	0.40	0.21	0.43	0.26	0.57	0.18
C2-1-2	C6 *zhi*	0.30	0.32	0.63	0.30	0.67	0.24
C2-1 long saliently 1D		0.35	0.27	0.53	0.29	0.62	0.22
C2-2-1	C7 *zhang*	0.40	0.21	0.53	0.23	0.67	0.24
C2-2-2	C8 *mian*	0.20	0.25	0.40	0.21	0.33	0.24
C2-2 flat saliently 2D		0.30	0.25	0.47	0.22	0.50	0.29
C2-3-1	C9 *ge*	0.67	0.24	0.73	0.26	0.67	0.24
C2-3-2	C10 *ke*	0.33	0.24	0.60	0.34	0.57	0.26
C2-3 round saliently 3D		0.50	0.29	0.67	0.30	0.62	0.25
C2 INANIMACY		0.38	0.28	0.56	0.28	0.58	0.26
count classifiers overall		0.35	0.29	0.53	0.29	0.56	0.28

In Table 4-3, a developmental progress was found in terms of all count classifiers. We can see that the older subjects, as far as their overall performance is concerned, could better use count classifiers than the younger ones. The average score of correct responses for Group 3 (0.56) was higher than that for Group 2 (0.53) and Group 2 higher than Group1 (0.35). One-way ANOVA indicated that there was a

significant difference among the three age groups (F (2, 42) = 10.675, p= 0.000)[2]. The Scheffe post hoc further indicated that a significant difference indeed existed between Group 1 and Group 2 (p=0.003), and Group 1 and Group 3 (p=0.001). However, the difference in performance between Groups 2 and 3, did not reach a statistical difference (p=0.857).

Moreover, count-classifiers comprise two categories: animacy and inanimacy. With regard to animate/inanimate distinction, the age effects were easy to detect. There was an upward trend in mean scores as the subjects' age increased. One-way ANOVA showed that in both animacy and inanimacy classifiers there were significant differences among the three age groups (animacy: p=0.002; inanimacy: p=0.000). The Scheffe post hoc also showed that, in animacy and inanimacy classifiers alike, significant differences were found between Group 1 and Group 2 (animacy: p=0.020; inanimacy: p=0.005), and Group 1 and Group 3 (animacy: p=0.005; inanimacy: p=0.001). The difference in performance between Groups 2 and 3, again, did not reach a statistical difference (animacy: p=0.868; inanimacy: p=0.905).

As we take a closer look at each semantic feature, there was a gradual improvement on count classifiers as well. Within the animate subsection of the classifiers, there were human and animal classifiers. With regard to human classifiers *ge* and *wei*, the average scores of correct responses of Groups 3 and 2 were higher than that of Group 1. Nonetheless, one-way ANOVA showed that the three age groups did not perform significantly differently in the human classifiers (F (2, 42) = 2.850, p= 0.069). The major problem lies in the classifier *ge*, where no significance was found between the three age groups (p=0.675)[3]. As for animal classifiers, the situation was a bit different. Despite the fact that a general developmental progress was found

[2] See Appendix D for the one-way ANOVA table.
[3] The statistical significance was found in the classifier *wei* (p=0.034).

in terms of animal classifiers, Group 3 performed less well than Group 2 in responding to classifier *tou*[4]. Even so, One-way ANOVA showed that the three age groups did perform significantly differently in animal classifiers ($F(2, 42) = 6.781$, $p = 0.003$). To be more specific, the Scheffe post hoc further showed that there was a significant difference between the three-year-olds and the four-year-olds, and the three-year-olds and the five-year-olds. However, the difference between the four-year-olds and the five-year-olds was not significant ($p=0.849$).

As for shape classifiers, as mentioned earlier, group effects were found as well. That is, in terms of overall shape classifiers (INANIMACY classifiers), the tendency was obvious that the older subjects outperformed the younger subjects. To begin with, in response to 1D classifiers *tiao* and *zhi*, a developmental progress was easily found. One-way ANOVA showed that there was indeed a significant difference among the three groups ($F(2, 42) = 6.931$, $p = 0.003$). As for 2D classifiers *zhang* and *mian*, there was a general developmental progress and one-way ANOVA confirmed this ($F(2, 42) = 4.796$, $p = 0.013$). Nevertheless, when encountering the classifier *mian*, only ten subjects of Group 3 scored the full points in the case of classifier *mian* in the PI task, where the twelve subjects of Group 2 succeeded in identifying the correct picture. Although some subjects of Group 3 seemed to perform less well than most subjects of Group 2, we might want to attribute such a result to individual differences. One-way ANOVA showed that in 2D classifiers there was a significant difference between the three-year-olds and the five-year-olds ($p=0.022$). However, the difference between the four-year-olds and the five-year-olds ($p=0.891$) and the three-year-olds and the four-year-olds ($p=0.066$) was not statistically significant.

Despite the overall developmental progression in shape classifiers, such

[4] The statistical significance was found in the classifier *zhi* ($p=0.001$). But no statistical significance was found in the classifier *tou* ($p=0.204$).

tendency, nevertheless, did not seem true in response to 3D classifiers. Compared to the five-year-olds, the four-year-olds performed much better on both 3D classifiers *ge* and *ke*. Though one-way ANOVA indicated that, in terms of 3D classifiers, there was a significant difference among the three age groups (p=0.020), a significant difference was found in the classifier *ke* (p=0.027) rather than the classifier *ge* (p=0.701). The Scheffe post hoc further showed that between the four-year-olds and the five-year-olds there was no significant difference in their performance (*ge*: p= 0.765, *ke*: p= 0.949). However, a significant difference was found in the classifier *ke* between the three-year-olds and the four-year-olds (p=0.046)[5].

What is noteworthy is that our subjects predominantly used the general classifier *ge* for almost every noun in production (e.g., **liang ge qizi* 'two flags,' **si ge yang* 'four goats,' **si ge kapian* 'four cards,' and **wu ge xianglian* 'five necklaces'), as suggested by the previous studies of classifier acquisition. In response of the 3D classifier *ge*, while Group 2 scored the best, Group 1 shared the same mean score as Group 3. The lower scores of the older subjects might not mark their incapability to produce such classifier, but rather indicate that the four-year-olds and five year-olds had began to search for a particular specific classifier other than *ge* when seeing the corresponding referent. (For more discussion about *ge*, please see sections 4.3 and 4.5.)

Table 4-4 presents the mean scores of mass classifiers of our subjects:

[5] No statistical significance was found in the classifier *ge* between Group 1 and Group 2 (p=0.765), Group 2 and Group 3 (p=0.765), and Group 1 and Group 3 (p=1.000).

Table 4-4: Mean Scores of Correct Mass Classifiers of Three Age Groups

Type		G1 (3yrs; n=15)		G2 (4yrs; n=15)		G3 (5yrs; n=15)	
		Mean	SD	Mean	SD	Mean	SD
M1-1	M1 *gongjin*	0.10	0.21	0.23	0.26	0.27	0.32
M1-2	M2 *gongfen*	0.07	0.18	0.17	0.24	0.27	0.26
M1 standard measures		0.08	0.19	0.20	0.25	0.27	0.29
M2-1	M3 *wan*	0.30	0.32	0.57	0.42	0.77	0.26
M2-2	M4 *ping*	0.30	0.37	0.67	0.36	0.77	0.37
M2 container measures		0.30	0.34	0.62	0.39	0.77	0.31
M3-1	M5 *kuai*	0.20	0.25	0.40	0.21	0.57	0.26
M3-2	M6 *pian*	0.30	0.32	0.60	0.34	0.60	0.28
M3 partitive measures		0.25	0.29	0.50	0.29	0.58	0.27
M4-1	M7 *qun*	0.17	0.24	0.27	0.26	0.47	0.13
M4-2	M8 *shuang*	0.30	0.25	0.57	0.26	0.63	0.30
M4 group measures		0.23	0.25	0.42	0.30	0.55	0.24
mass classifiers overall		0.22	0.28	0.43	0.34	0.54	0.33

As shown in Table 4-4, developmental differences were also detected among the subjects' abilities to use mass classifiers—the older the children were, the better their performances were. As far as all mass classifiers are concerned, one-way ANOVA confirmed that a significant difference was found among the three age groups ($F(2, 42) = 17.526$, $p = 0.000$). Nevertheless, similar to the performance of overall count classifiers, the Scheffe post hoc test further indicated that a significant difference indeed existed between Group 1 and Group 2 ($p=0.002$), and Group 1 and Group 3 ($p=0.000$). However, Group 2 and Group 3 did not perform statistically differently ($p=0.166$).

As we take a closer look at each semantically-based subgroup, a gradual developmental progress was also found in mass classifiers. One-way ANOVA showed

that there was indeed a significant difference among the three groups in standard measures (p= 0.039), container measures (p= 0.000), partitive measures (p= 0.000) and group measures (p= 0.000). Again, the Scheffe post hoc further indicated that a significant difference existed between Group 1 and Group 2, and Group 1 and Group 3. Group 2 and Group 3, nonetheless, did not differ statistically.

It is intriguing to note that in container measures, partitive measures and group measures, the developmental progress seemed conspicuous while in standard measures, the progress seemed relatively slower, which might indicate that standard measures were far more difficult for our subjects to comprehend and produce.

To summarize, one-way ANOVA indicated that the three age groups responded significantly to overall count classifiers (p= 0.000) and mass classifiers (p= 0.000). Within the semantically-based subgroups, with the expectation of human classifiers, the tendency that the older outperformed the younger subjects was obvious and confirmed by SPSS. The Scheffe post hoc further indicated the statistical significances were found mainly between the three-year-olds and the four-year-olds, and between the three-year-olds and the five-year-olds, rather than between the four-year-olds and the five-year-olds.

The foregoing section has dealt with the results of the three age groups. Let's now turn to our research question: Do Chinese children of different age groups respond significantly differently to the count-mass distinction? The abovementioned data were analyzed by the **Two Way Mixed Design ANOVA** and confirmed that our subjects of different age groups did respond significantly differently to the count-mass distinction (p=.000)[6]. To put it another way, developmental differences were significant in response to the count and mass classifiers.

For decades, there has been substantial interest in the question of how age

[6] See Appendix E for the table of the Two Way ANOVA analysis.

affects language acquisition. Take children's semantic development for example, according to Piaget (1932), children have accomplished a lot by the age of one. They are able to repeat interesting invents and show preliminary indications of classification or meaning. Shortly after age one, children begin to utter words. To use words meaningfully, they begin to associate a sound with a meaning. By the age of 1;6, they have already acquired the object concept. It is also around this age that children begin to form a mental representation. Children in their second year have gradually become familiar with their surroundings, and they had formed some mental representations which were the meanings of the words they first acquired. At age three, children command some notions of time and aspectual distinctions and are beginning to develop a repertoire of modal meaning (Goodluck 1991).

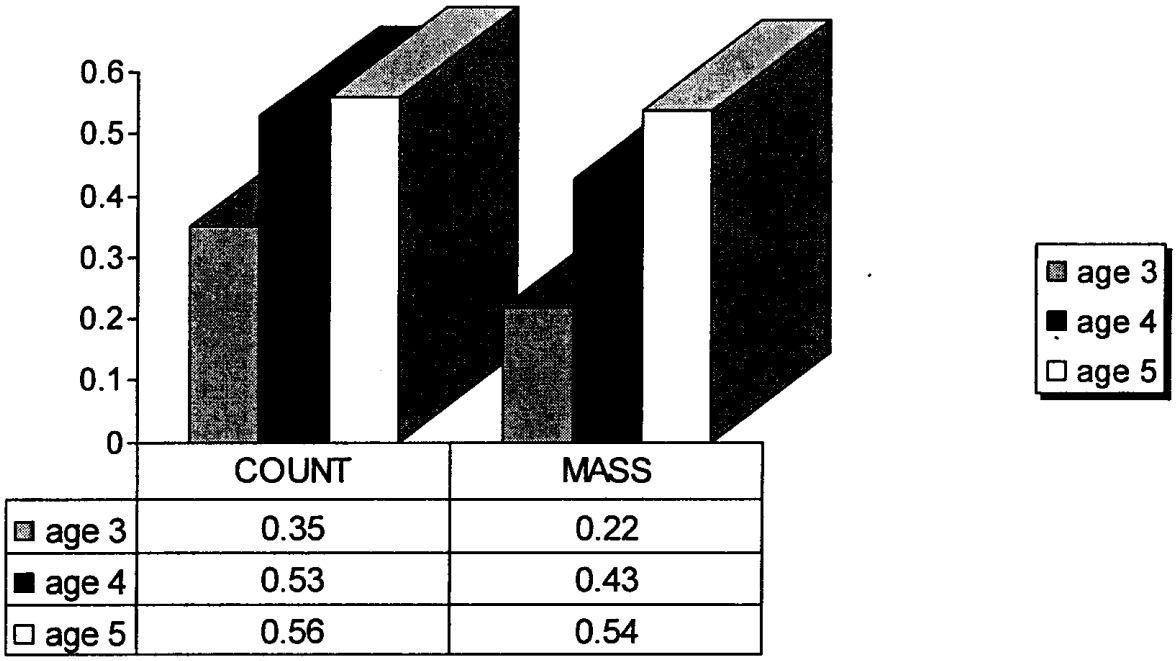
In the present study, age is also a determining factor for affecting our subjects' use of count and mass classifiers. The fact that our subjects' ability to detect the grammatical count and mass distinction increased with their age is understandable in that as our subjects grew older, they exhibited a higher level of linguistic competence and performance. Chinese language acquisition researchers (Chien et al. 2003, Fang 1985, Hu 1993, Ying 1983) also reported that children's acquisition of classifiers is strongly associated with their age.

In Chien et al.(2003), 80 children between the ages of three and eight and 16 adults, who served as the comparison group, were tested. In their experiment, where both count and mass classifies (exclusively container measures) were examined, developmental differences were also found among their subjects' abilities to comprehend count and mass classifiers. In other words, there were significant differences among five age groups in regard to their performance on count classifiers ($F(4, 75) = 37.66$, $p < 0.01$) as well as their performance on mass classifiers ($F(4, 75) = 16.08$, $p < 0.01$). To be more specific, significant differences in the subjects'

abilities to comprehend count-classifiers were found (1) between the 6- and the 7-year-olds between the 4- and the 5-year-olds, and (2) between the 4- and 5-year-olds and the 3-year-olds. No significant differences were found (1) between the 6- the 7-year-olds between adults, and (2) between the 4-year-olds and the 5-year-olds. In other words, in response to count classifiers, the 6- and the 7-year-olds performed as well as the adults, but they performed significantly better than the 4- and 5-year-olds. The 4-year-olds performed as well as the 5-year-olds, but both 4-year-olds and the 5-year-olds performed significantly better than the 3-year-olds. As for mass classifiers, significant differences were found only (1) between the adults and the 5-year-olds, (2) between the 6- the 7-year-olds and the 4-year-olds, and (3) between the 4- and the 5-year-olds and the 3-year-olds. In other words, in response to mass classifiers, the 6- and the 7-year-olds performed as well as the adults, but they performed significantly better than the 4-year-olds. The 4-year-olds performed as well as the 5-year-olds, but both 4-year-olds and the 5-year-olds performed significantly better than the 3-year-olds.

The present study, echoing the pervious studies, found that age indeed played a crucial role in the success of comprehension and production of Chinese count and mass classifiers.

It was also observed in the production data that those children aged three used only two to three classifiers. As Erbaugh (1982) found that children almost never used specific classifiers before the age of 2;6, specific classifiers seemed to emerge at around the age of three. In the present study, although the three-year-olds used the general classifier most of the time, the number of classifiers produced increased steadily with their age. By age four, they used four to five classifiers. By age five, they produced five to six classifiers. To put it another way, the frequency and the number of different classifiers our subjects used correlated positively with their age.



	COUNT	MASS
age 3	0.35	0.22
age 4	0.53	0.43
age 5	0.56	0.54

Figure 4-1: Distribution of the Subjects' Correct Response to the Two Types of Classifiers in PI and PD Tasks

Similarly, as can be seen in Figure 4-1, in terms of the count-mass distinction, as the subject's age increased, so did his/her performance of classifiers. Furthermore, in addition to the fact that the general age effects were found in their use of classifiers, the present study presents some new findings. The Scheffe post hoc, as mentioned earlier, indicated the statistical difference existed only between Group 1 and Group 2, and between Group 1 and Group 3. Group 2 and Group 3 did not perform statistically differently. The four-year-olds performed roughly similarly as the five-year-olds. To put it differently, in the present study it was found that, both in response to count and mass classifiers, the age between three and four was a critical stage in our children's classifier development. It is during this momentous period that our subjects began to rapidly learn to grasp the meaning of count and mass classifiers and use them in their daily life.

From age three to age four years, it is said that children start to learn how to

read and write and their verbal skills increase rapidly. For English children, for instance, vocabulary grows from about 900 words at age three to 1500 words by age four. During this period of speedy vocabulary growth, children enjoy learning new and unfamiliar words. They also enjoy being read to and telling stories. During this period most children can count from one to three, focus on a specific task, and recall things that happened in the recent past. This is also the age at which most children know what sex they are, and they are be able to name most of their body parts. All this may contribute to children's critical development in their acquisition of classifiers between age three and age four (Lee and Dasgupta 1995).

Children, by the age of four, have already acquired a good many classifiers so that after age four, the developmental progress is comparatively less obvious. This may cohere to Erbaugh's findings (1982). According to Erbaugh, the children were found to demonstrate a sound knowledge of the basic syntactic nature of classifiers at a very young age. Her two participants, aged two and three, were almost as reliable as the adults in using classifiers whenever one was required.

The results of the present study supported Chien et al. (2003). As mentioned above, with regard to both count and mass classifiers, their 4-year-olds performed as well as the 5-year-olds, and they performed significantly better than the 3-year-olds. To put it differently, the age between three and four is a critical stage in Chinese children's classifier acquisition.

Interestingly, in addition to the developmental progress, there are two phenomena worth mentioning. Firstly, if we compare the respective performances of the three age groups, we can see that our subjects' differences in count classifiers were smaller than those in mass classifiers, indicating that in the acquisition process of classifiers, children aged three to five may make much more progress with regard to mass classifiers than count classifiers. Secondly, despite the fact that our subjects

generally showed better abilities in dealing with count classifiers, their performance difference between count classifiers and mass classifiers became less marked as their age increased. To be more specific, for the three-year-olds, the result of count and mass classifiers was 0.35 to 0.22. The result of the four-year-olds was 0.53 to 0.43 and the result of the five-year-olds was 0.56 to 0.54. All this suggests that our children's ability to judge count-mass distinction should correlate positively with their age. By the age of five, they had roughly comparable performance on count and mass classifiers.

It was also found in the present study that the use of the general classifier *ge* decreased with the subjects' increasing age and acquisition of other classifiers. This result corresponded exactly to Hu's (1993) finding. In her study, half of the three-year-olds used the general classifier exclusively. Only two four-year-olds and one five-year-old did so. None of the six-year-old children used the general classifier exclusively. Furthermore, the present findings also supported the results of Loke (1991), who studied twenty-one young children's use of Mandarin shape classifiers, and found that the more shape classifiers children used, the less frequently *ge* was used. In consequence, the present findings provide another piece of evidence that children's predominant use of the general classifier decreases with their increase in classifier development.

As a whole, our findings support the previous linguistic studies of Chinese classifier acquisition. The age effects found in our subjects' comprehension and production serve as an empirical support for the previous linguistic analyses that age is a determining factor for the success of Chinese classifiers (Chien et al. 2003, Fang 1985, Hu 1993, Loke 1991, Ying 1983). More importantly, it was found that the age from three to four was a critical stage for our children's rapid development in their acquisition of classifiers. In addition, our subjects' predominant use of the general

classifier *ge* was found to decrease with their increase in classifier development, which supports the findings of Loke (1991) and Hu (1993).

4.3 Methodological Effects

Thus far, our main focus has been on whether the participants involved understand the grammatical count-mass distinction and the general performance of three respective age groups. Let's switch the focus to methodological effects.

A primary goal of language acquisition research has been to access the Chomskyan notion of grammatical competence. It is often more difficult to access young children's knowledge of language than adults'. Researchers have therefore devised various methods appropriate for assessing young children's early grammatical abilities. In the present study, with an eye to reducing the subjects' response biases to the minimum (White 1989) and evaluating how differently they might perform in the different tasks (Lightbown 1984, McDaniel et al. 1996), multiple tasks were designed to elicit children's linguistic knowledge within a short period of time – one comprehension task (the PI task) and one production task (the PD task). The PI task, on the one hand, aimed to examine whether or not the subjects comprehended the meaning of the tested count/mass classifiers by identifying one target object out of the three based on the classifier they heared from the verbal stimulus. The PD task, on the other hand, was to investigate how the subjects used the classifiers to describe the corresponding entity and to diagnose whether there was any misuse. As can be seen in the literature review in Chapter Two, some researchers in the field of classifier acquisition included only one type of task (Chien et al. 2003, Fang 1985), which turned out weak in making any substantial and satisfying claims. Accordingly, it is essential that both comprehension and production tasks be adopted. The former enables us to test children's sensitivity to aspects of language which they do not

produce in their spontaneous production while the latter makes it possible to uncover the extent of children's grammatical knowledge about classifiers and to evoke sentences corresponding to classifier structures which occur rarely in children's spontaneous speech (McDaniel et al. 1996).

Our subjects' correct responses to count/mass classifiers in the two tasks are shown in Figure 4-2:

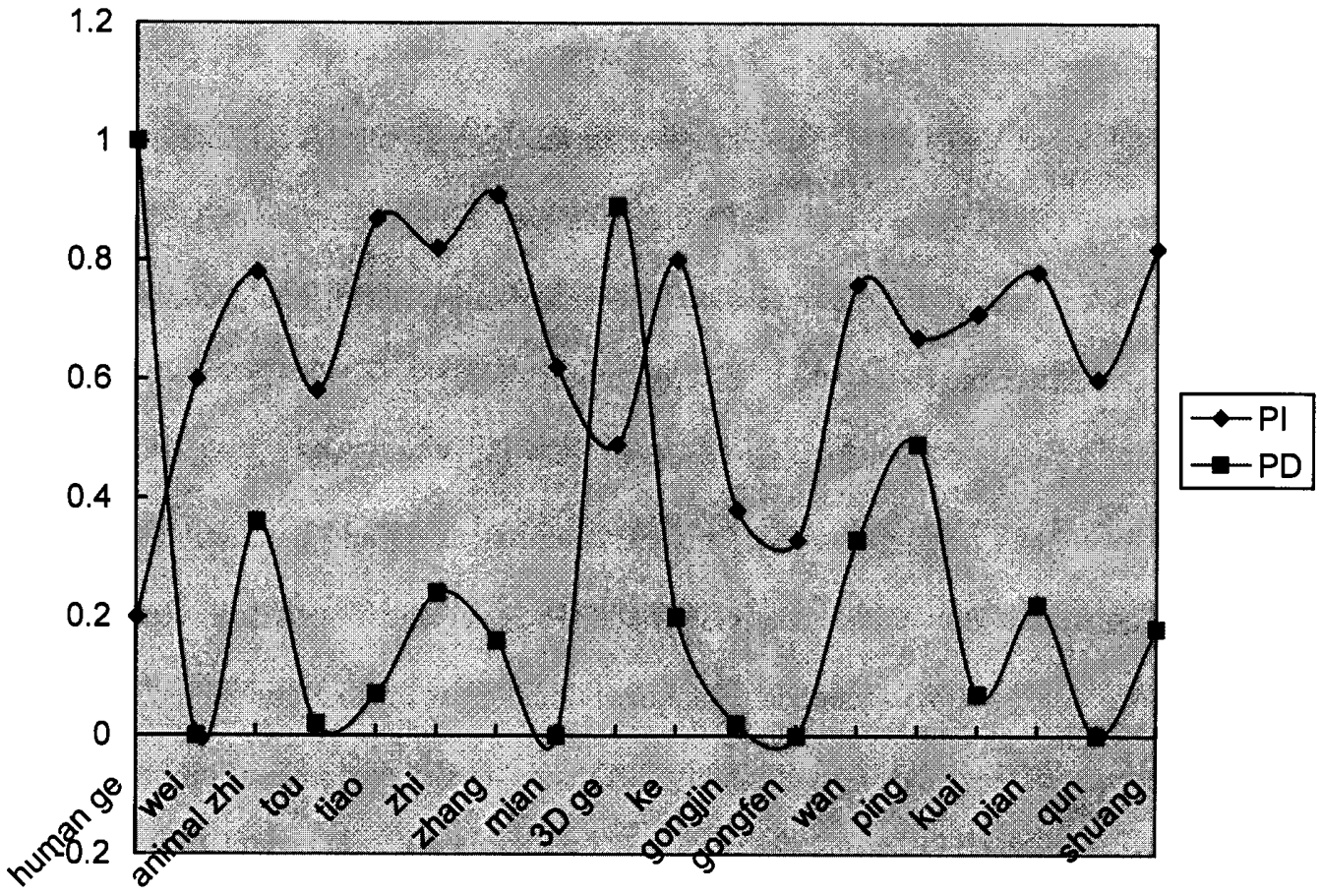


Figure 4-2: Distribution of the Subjects' Correct Responses in the PI and PD Tasks

Figure 4-2 indicates that the mean scores of the subjects' correct responses in the PI task were generally far higher than those of the PD task. They had more correct responses in the PI task than in the PD task. However, such tendency did not hold true in the case of the human classifier *ge* and the 3D classifier *ge*, which both exhibited a

reverse tendency as opposed to other classifiers. Such notable exception arised from the fact that when asked to generate the classifier phrases, our subjects predominately used the classifier *ge* with most noun referents regardless of their semantic properties. Therefore, the classifier *ge*, including human *ge* and round saliently 3D *ge*, received better scores in the production task. Meanwhile, the subjects' performance on the classifier *ge* seemed to go from one extreme to the other. In the PD task, its score was a lot higher than other classifiers. In the PI task, however, its score ranked lowest among count classifiers.

Also, there were classifiers that all the forty-five subjects failed to produce in the PD task. They were honorific classifier *wei*, 2D classifier *mian*, standard measure *gongfen*, and group measure *qun*. Yet, some subjects still could comprehend these classifiers in the PI task (to be discussed below).

Now, let us turn our focus to our third research question: Do Chinese preschoolers perform similarly on the comprehension and production tasks in their acquisition of count and mass classifiers? The abovementioned data were analyzed by Two-Way ANOVA (dependent samples) and confirmed that our subjects of the three groups did responded significantly differently to the comprehension and production tasks ($p=.024$)[7].

In our experiment, there was a significant difference in methodology. Our subjects, as shown in Figure 4-2, performed significantly better on the comprehension task (PI) rather than on the production task (PD). This result corresponds to the generally-accepted assumption that comprehension normally precedes production or competence precedes performance (McCarthy 1954, McDaniel et al. 1996). Moreover, different task formats indeed influenced our subjects' performance on count/mass classifiers, which supports the previous studies and claims (Lightbown 1984,

[7] See Appendix G for the table of the Two Way ANOVA analysis.

McDaniel et al. 1996, White 1989). Take the honorific classifier *wei*, the 2D classifier *mian*, standard measure *gongfen*, and group measure *qun* for instance. As mentioned earlier, no subjects used these classifiers to describe the picture stimuli. Children's failure to produce these four linguistic elements does not mean that they do not perceive or represent them in their linguistic capacity. As can be seen in the comprehension task, some children were sensitive to such classifiers and able to differentiate them from others. That is, the comprehension task successfully tested our children's sensitivity to aspects of language which they did not produce in their spontaneous speech. The present finding that the subjects were sensitive to these classifiers before they produced them also supports the assumption that comprehension is prior to production in language development.

In addition, our subjects' better performance on the PI task can be attributed to the fact that they overused *ge* in the PD task. They allowed *ge* to co-occur with almost any nouns which are considered to be "informal," "inappropriate," or "linguistically unsophisticated" in adult language, and require more specific count classifiers to be used. In other words, the strategy of our subjects was to use *ge* when they did not know which one to use. This finding echoes the general conclusion made in the previous studies (Chien et al. 2003, Erbaugh 1986, Fang 1985, Hu 1993).

As mentioned earlier, the subjects' performance on the classifier *ge* seemed to be at the opposite end of the scale—the highest score in the PD task, yet the lowest in the PI task. Our children's predominant use of *ge* in their production led to its highest score in the PD task. Nevertheless, in the PI task, which was to examine the subjects' comprehension, the subjects were asked to identify the target objects based on the classifier they heard from the verbal stimulus. If a child comprehended the grammatical count-mass distinction, then he or she would choose an entity corresponding to a count noun in the count-selective context and an entity

corresponding to a mass noun in the mass-selective context. Interestingly, here *ge* put obstacles in the way of identification. When hearing the verbal stimulus including *ge*, most of our children showed hesitation in deciding their answers. Some children even first pointed to one picture and then suddenly switched to another one. All this indicates that for children *ge* is, in a way, an all-purpose classifier. When asked to describe the picture by using the classifier phrase, they were very likely to produce *ge* in their speech. But when asked to look for the referent with the provided classifier, *ge*, they had a difficult time finding the right corresponding entity in that it seemed that *ge* could go with nouns with any semantic properties for them.

On the whole, it was found that there was a significant difference between our subjects' comprehension and production ($p<.05$). Our subjects aged three to five showed better abilities in comprehension than in production of the tested classifiers. Meanwhile, in spite of the fact that they overused *ge* in their production, they were still aware of other classifiers, able to link the classifier and its corresponding semantic properties, showing that there as a general gap between our subjects' comprehension and production of Chinese classifiers.

4.4 The Hierarchy of Difficulty of Count and Mass Classifiers

This section aims to address the fourth research question about the hierarchy of difficulty in the subjects' acquisition of count and mass classifiers. The classifier use of the subjects will be unfolded first, followed by a summary of the hierarchy of difficulty of classifiers. At the end of the section, discussion will be presented.

The correct responses of all the subjects to each classifier are shown in Table 4-5:

Table 4-5: Subjects' Mean Scores of Correct Count and Mass Classifiers

Count Classifier				Mass Classifier			
Type		mean	SD	Type		mean	SD
C1-1-1	C1 *ge*	**0.60**	0.20	M1	M1 *gongjin*	**0.20**	0.27
C1-1-2	C2 *wei*	**0.30**	0.25	M1	M2 *gongfen*	**0.17**	0.24
C1-2-1	C3 *zhi*	**0.57**	0.36	M2	M3 *wan*	**0.54**	0.38
C1-2-2	C4 *tou*	**0.30**	0.27	M2	M4 *ping*	**0.58**	0.41
C2-1-1	C5 *tiao*	**0.47**	0.22	M3	M5 *kuai*	**0.39**	0.28
C2-1-2	C6 *zhi*	**0.53**	0.33	M3	M6 *pian*	**0.50**	0.34
C2-2-1	C7 *zhang*	**0.53**	0.25	M4	M7 *qun*	**0.30**	0.25
C2-2-2	C8 *mian*	**0.31**	0.25	M4	M8 *shuang*	**0.50**	0.30
C2-3-1	C9 *ge*	**0.69**	0.25				
C2-3-2	C10 *ke*	**0.50**	0.30				

Table 4-5 illustrates the results of the 45 subjects' use of classifiers. As far as each individual classifier is concerned, the rankings were (in descending order of total mean scores): 3D *ge* (0.69), human *ge* (0.60), *ping* (0.58), animal *zhi* (0.57), wan (0.54), 1D *zhi* (0.53), *zhang* (0.53), *ke* (0.50), *pian* (0.50), *shuang* (0.50), *tiao* (0.47). *kuai* (0.39), *mian* (0.31), *wei* (0.30), *tou* (0.30), *qun* (0.30), *gongjin* (0.20), and *gongfen* (0.17).

Thus far, we have presented the mean scores of each classifier. The results of each individual classifier and their corresponding hierarchy of difficulty are illustrated in Figure 4-3:

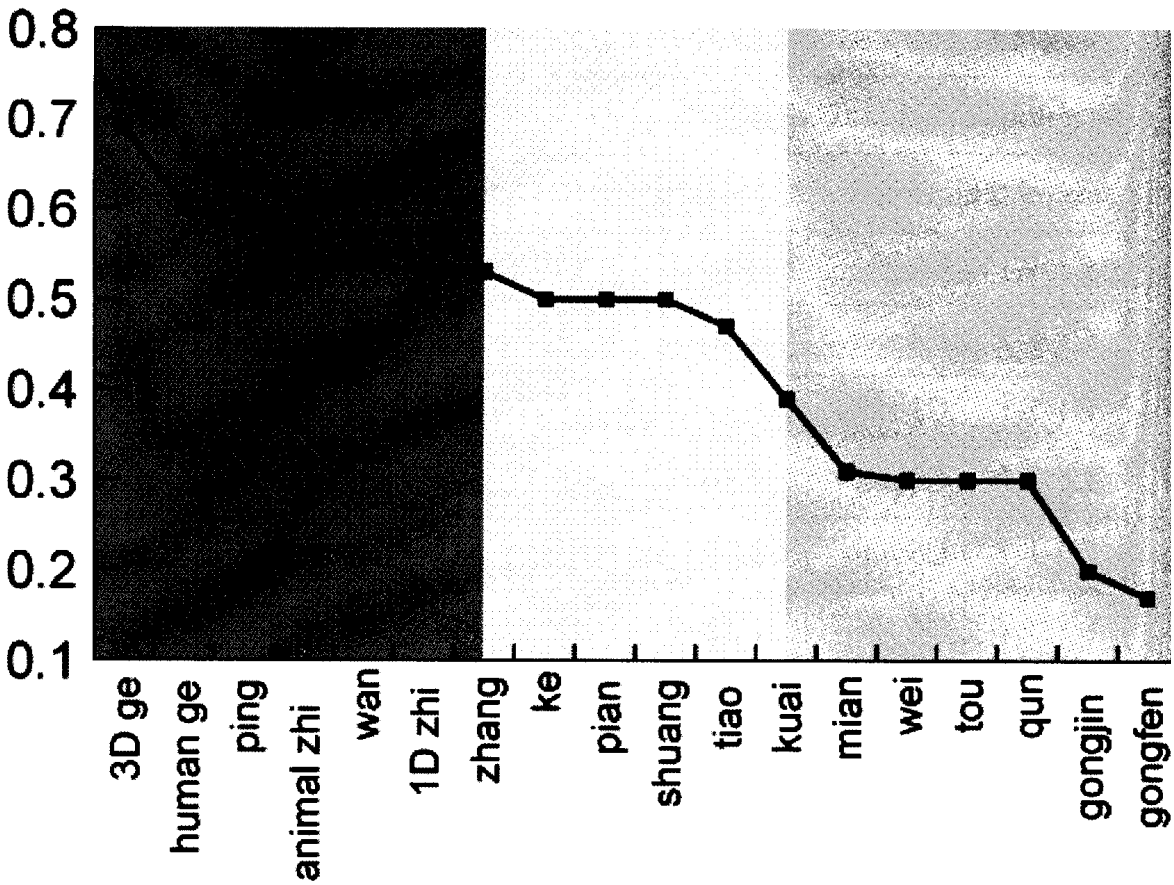


Figure 4-3: Distribution of the Subjects' Correct Responses to Each Classifier

As can be seen, the classifiers used by the subjects indicated a general emergence order among the children aged between three and five. Among the classifiers tested, *ge* indeed was the earliest acquired classifier. Children used *ge* initially and preponderantly as a "general classifier" for all objects, which contributed to its highest frequency in their speech. Similarly, although some previous studies have shown different results concerning the order of classifier acquisition, they all reached the conclusion that children first acquired the classifier *ge* (Ahrens 1990, Chien et al. 2003, Erbaugh 1986, Fang 1985, Hu 1993, Loke and Harrison 1986, Loke 1991, Ying 1983). The present findings can be another piece of evidence for the *ge*-prominent phenomenon in Chinese children's classifier acquisition.

In addition to the human classifier *ge,* the animal classifier *zhi* was among the early acquired classifiers as well, showing that our children's use of the classifiers was constrained by ANIMACY (MacWhinney and Fletcher 1995). This observation is comparable with Fang's (1985), Erbaugh's (1986) and Hu's (1993) findings. In an experimental study of 36 four to six-year-old 36 Chinese children, Fang (1985) found that the animal classifier *zhi* was the most correctly used classifier. The animacy classifier *zhi* was also found among the most frequently used classifiers by two of Erbaugh's (1986) four children. In Hu's (1993) study, *zhi* (the animacy classifier) was the earliest acquired classifier as well. Hu further pointed out that *shuang* (the arrangement classifier), and *zhang* (the shape classifier denoting two dimensions and thinness) were also the most frequently used among her twenty-four Mandarin-speaking children from ages three to six years. The present findings also support Hu's results in that *zhang* (0.53) and *shuang* (0.50) were found among the early acquired classifiers.

Furthermore, for our subjects, container measures *ping* and *wan* were relatively easy both in the comprehension and production tasks. This is because containers are easy to find in children's immediate environment and they have to eat and drink every day, hence children are familiar with the classifiers *ping* and *wan.* On the other hand, patritive measures, such as *pian* and *kuai,* were challenging for our children, which indicates that the children aged three to five have not fully established the meaning and use of partitive measures, which designate partitions or parts of objects rather than groups of them. Our subjects' unfamiliarization of *kuai* and *pian* supports the findings of Loke and Harrison (1986) and Loke (1991). In their attempt to investigate young children's use of seven shape classifiers[8] in Mandarin Chinese, it was found

[8] In Loke and Harrison's (1986) study, the seven classifiers under investigation were *li, zhi, tiao, zhang, pian, kuai,* and *ge.* Their subjects were aged between 5;0 and 6;8.

that *kuai* and *pian* were used for the least number of objects and had the lowest frequency of use among the tested classifiers. They were the least salient classifiers and therefore among the late acquired classifiers.

Among the late acquired classifiers, standard measures *gongjin* and *gongfen* received the lowest mean scores. In other words, our children encountered the most difficulty with standard units of measurement. Such findings meet our expectations in that most preschoolers are not familiar with the concept of standard measuring system, which is often taught in the curriculum of elementary school after children acquire the basic notion of mathematics.

It is worth mentioning that though some classifiers belong to the same semantically-based subgroup, one was relatively easy for our children while the other was just the opposite. Classifiers such as human *ge* versus *wei*, *zhang* versus *mian*, *zhi* versus *tou*, and *shuang* versus *qun* were these cases.

To begin with, both *ge* and *wei* can be used to denote the human referents. *Wei*, which exhibits a higher social status of the human referents, posed considerable difficulty for our subjects aged from three and five. In the comprehension task, ten of the subjects chose the noun referent *xiaoyinger* (baby) rather than *laoshi* (teacher) to co-occur with the classifier *wei*. *Wei* reflects the social and cultural prominence of the human referents but *ge* simply denotes common humans who are not in the category "people to be polite to." Hence, while *wei* has the semantic feature [+honorable], *ge* does not. At this stage of development, our children had difficulty differentiating "people to be polite to" from "the general public." For them, detecting whether a person is honorable and respected and thus required the use of the honorific classifier *wei* is closely associated with their cognitive development. With the increase of their age, they gradually understand and discern the social and cultural subtlety. This not only corresponds to Hu's (1993) study, where the honorific classifier *wei* was one of

the late acquired classifiers, but also matches the findings of a number of studies (Chang 1983, Fang1985, Ying and Chen 1983) that found Chinese children's proficiency in classifiers to be closely related to their educational leveld and cognitive development.

Similarly, while *zhang* proved to be simple for our children, *mian* remained an obstacle. Several subjects could not tell the differences between the two-dimensional classifiers *zhang* and *mian*. They chose the noun referent *jingzi* (mirror) rather than *tiezhi* (sticker) to co-occur with the classifier *zhang*. Both *zhang* and *mian* share the physical attribute of flatness. Since *jingzi* (mirror) is a flat object, our children's misuse of *zhang* is understandable. Nonetheless, if we take a closer look, *zhang* denotes flat, thin and flexible entities such as in *yi zhang congyoubing* (a deep-fried scallion pancake), but *mian* refers to hard or nonflexible objects with flat surfaces such as *qiang* (wall). In other words, the following semantic features can be assigned to *zhang*: [+2D, +thin, +flexible]; and to *mian*: [+2D, ±thin, −flexible]. In other words, *zhang* is much more salient than *mian* in the feature 'flexibility.' In addition to the difference in flexibility, *mian* is only salient on the front side or the face (Tai and Chao 1994). It mainly classifies objects which have a front side or face. As a consequence, a mirror is often the most prototypical object. For our children, such subtle differences were not detected and applied until they acquired the adults' usage, hence resulting in the difficulty of *mian*.

Both *zhi* and *tou* are classifiers associated with non-human animates. The former is used with any animals whereas the latter is based on a part-whole relationship between the animal enumerated and its head. That is, one prominent attribute of the referent is used to refer to the whole referent. Such part-whole notion, again, posed great difficulty for our subjects.

Shuang and *qun* are group measures, designating a group or collection of

individuals. For our children, *shuang* is comparatively easy in that it is a classifier to refer to the object that is made from two similar parts that are joined together. Noun collates of *shuang* such as *xiezi* "shoes" and *kuaizi* "chopsticks" are readily available and easy to detect in their immediate environment, hence resulting in their familiarization. *Qun*, on the other hand, designates several people, animals or things that are all together in the same place. The classifier *qun* provides a collective reference for separate entities. This collective notion involving large quantities caused our children difficulties. Despite the fact they might comprehend the meaning of *qun* (27 subjects comprehended its meaning in the PI task), none of the 45 subjects were able to use it in the production task. Most of them used the phrase *hen duo* (a large quantity) rather than the classifier *qun* to describe the picture where lots of monkeys were presented.

4.5 Subjects' Misuse of *Ge*

This section is to answer the fifth research question, attempting to investigate what semantic feature governed our subjects' misuse of count and mass classifiers. The subjects' misuse of the classifier *ge* in the two tasks will be presented first, and the relationship between the misused classifier and semantic features of the entities will be discussed.

In the production task, all young children aged three to five used the general classifier *ge* more than half of the time when an object was presented to prompt a classifier. They predominately used the general classifier *ge* to classify objects. No other classifier could compete with *ge* in terms of frequency of use. They generalized use of the general classifier widely and used the general classifier without paying any attention to the semantic properties denoted by the referents of nouns. To put it differently, they overused *ge* with nouns belonging to different semantic classes (e.g.,

*si ge yang, *yi ge pao-mian, *san ge shu-ye and *san ge ka-pian). *Ge* was, in a way, an all-purpose classifier for them. Their strategy was to use *ge* whenever they did not know which classifier to use.

This overwhelming use of *ge* in place of a specific classifier was also addressed in a great abundance of literature (Ying 1983, Fang 1985, Erbaugh 1986, Loke and Harrison 1986, Ahrens 1990, Loke 1991, Hu 1993, Myers 2000, Chien et al. 2003). In these studies, children before the age of four predominantly used the general classifier *ge*, and by the age of six or seven they still were very likely to use *ge* rather than other specific classifiers. Erbaugh (1986), in particular, pointed out that her children at 2;6 used only the general classifier for all nouns. This is consistent with our findings with the three-year-olds. Almost everyone in this age group used *ge* exclusively for all the nouns tested in their production of speech.

Ge was the classifier most frequently used and earliest acquired, perhaps because adults use it very often in their daily speech. In addition, the important feature for application of this classifier is that the object is countable. In children's immediate environment, most of the objects they come into contact are mostly countable. Hence, *ge* is well-established in children's vocabulary, naturally becoming their first and most useful and practical choice.

The subjects' preponderant use of *ge* might confirm its default status and indicate that it has no semantic meaning for them and therefore can be used to classify or re-classify all nouns regardless of different semantic properties denoted by the referents of nouns (Loke 1994). Also, this general classifier is particularly useful for objects that are not very familiar to them. In consequence, *ge* can be considered UNMARKED.

As mentioned in section 4.2, it was found that the older the subjects were, the less frequently they used the general classifier. Such **decrease** in the use of *ge* and the

proportionate **increase** in both the range and frequency of other count classifiers used also provided a clear piece of evidence that our children first used *ge* as an unmarked classifier and later used other classifiers as marked ones.

In the comprehension task, on the other hand, *ge* was the most troublesome classifier. Hearing the general classifier *ge* did not make our subjects look for a particular referent, hence yielding its lowest score among all the classifiers. This finding corresponds precisely to the result of Chien et al. (2003). In their comprehension experiments, where fourteen count classifiers and four mass classifiers (all container measure words) were tested, the general classifier *ge* posed the biggest problem as well. The "*ge*-preference" prevented most of their children from choosing the appropriate classifier and therefore they allowed *ge* to co-occur with almost any nouns which were considered by them to be inappropriate or linguistically unsophisticated in adult language.

To sum up, our children's misuse of classifiers exhibited overgeneralization. They tended to extend their use of classifiers from prototypes to peripheral exemplars[9] (Erbaugh 1986). The young children particularly generalized the use of the classifier *ge* widely. Meanwhile, our children's use of classifiers against adults' expectation or usage might reflect their own semantic system of classification[10], which might be different from that of adults. To put it differently, our children's classification was based on their own perception of the objects classified. And they might consciously or

---

[9] There are other misuses particularly deserving of note. One of the earlier acquired classifiers, the animal classifier *zhi*, was another problem of overgeneralization. The classifier *zhi* is supposed to be used to specify animals being enumerated. However, several subjects of the present study associated *zhi* with *laoshi* (teacher). They overgeneralized the animal feature to human beings. This *zhi*-overgeneralization echoed Hu's(1993) study. In her study, children not only applied *zhi* to appropriate class members such as *gou* 'dog' and *mao* 'cat,' but they also overextended it to inappropriate class members such as *ma* 'horse' or human beings. Some even overgeneralized *zhi* to movable objects such as *qiche* 'cars' and *jiuhuche* 'ambulances,' which often co-occur with the vehicle classifier *liang* in adult language.

[10] In the PI task, one subject described the picture of chopsticks as *liang zhi kuaizi* (two chopsticks) instead of *yi shuang kuaizi* (a set of chopsticks). He was not capable of using the group measure *shuang* to designate a collection of two chopsticks. Instead, the child perceived the cylindrical shape of the long, rigid chopstick and thus associated *kuaizi* with *zhi*.

subconsciously aware of the perceptible physical attributes of the objects counted or a quantitative measurement of the objects classified. Accordingly, their use, or misuse from the adults' perspective, definitely shed new light on their classifier acquisition.

4.6 Summary of Chapter Four

This chapter has presented and discussed the major findings of the present study. In section 4.1, the subjects' performance on the count-mass distinction was explored. It was found the forty-five young children of the present study, at their early stages of language acquisition, responded significantly differently to the count-mass distinction. They performed better on count classifiers than on mass classifiers and, for most cases, they used count classifiers to enumerate individuated entities and used mass classifiers to quantify substances. In section 4.2, age effects were found in our subjects' performance, which cohered with the previous linguistic analyses that age plays a crucial role in Chinese classifier acquisition. Most importantly, the age from three to four was found to be a critical stage for our children's rapid development in their acquisition of classifiers. Additionally, our children's predominant use of the general classifier *ge* was found to decrease with their increase in classifier development. In section 4.3, the formats of both tasks were found to influence the performance of the subjects. They showed better abilities in comprehension than in production of the tested classifiers. Furthermore, although they overused *ge* in their production, they were still aware of other classifiers, able to link the classifier and its corresponding semantic properties. In section 4.4, the hierarchy of difficulty in the subjects' acquisition of count and mass classifiers was presented. It was found that *ge* was the classifier acquired earliest whereas the standard measures pose the major difficulty. Finally, in section 4.5, the subjects' misuse of the *ge* classifier was under investigation. Our subjects generalized use of *ge* widely and used it without paying

any attention to the semantic properties denoted by the referents of nouns. The preponderant use of *ge* confirmed its default status and indicated that it had vacuous semantic meanings for our children. As a consequence, *ge* can be considered an UNMARKED classifier by our subjects. In Chapter Five, a summary of the main points of the present study, and suggestions for the further research will be given.