Receptive vocabulary development of children acquiring Chinese, Swedish and Korean as measured by the PPVT-4

Bachelor thesis Taalwetenschappen

200200756

Date: 13 June 2014

Supervisor: Aoju Chen

Second reader: Shalom Zuckerman

Saskia Verstegen, 3581373

Acknowledgements

First, I would like to thank Aoju Chen, my supervisor, who guided me very well through the process of writing this thesis. I would also like to thank Shalom Zuckerman, as the second reader, and Valentijn de Jong who read and commented on my pre-final versions extensively.

I am also grateful to Anna Sara Romøren (Swedish data), and Anqi Yang (Korean and Chinese data) for making the PPVT data available to me, Xiaoli Dong, Zenghui Liu, Soyeon Julie Park, Diantha de Jong, Anna Sara Romøren and Anqi Yang are thanked for helping me out with issues concerning translation and explaining some of the found results. Mattis van den Bergh is thanked for statistical advice and for teaching me how to do the analysis step by step.

This thesis would not have been possible without funding for the collection of the PPVT data in Stockholm, Seoul and Beijing. Therefore, I am grateful for the VIDI grant (276-89-001) awarded to Aoju Chen by the NWO (Netherlands Organisation for Scientific Research).

Abstract

Children's early vocabulary seems to develop in a similar way across languages and cultures regarding rate and route of acquisition, but there is still much individual variation within and across languages. A noun bias has been found in many languages, though there is not yet a consensus on the further composition of vocabularies and whether verbs are more prominent in vocabularies of children speaking languages like Chinese and Korean. In the present study, receptive vocabulary from Swedish-, Korean- and Chinese-speaking children aged 4 to 11 years were examined on the basis of data obtained via the Peabody Picture Vocabulary Test-4 (PPVT-4). It has been found that vocabulary development is not similar in these three languages: the Chinese-speaking children outperformed the Korean- and Swedish-speaking children. Further, the noun bias were replicated for the children from all three languages, differences between verb knowledge between these children were not found. Suggestions are given on how to revise the translated versions of the PPVT-4, in addition to critical remarks on the current study.

1. Introduction

Children start learning words from a very early age (e.g. Bates e al., 1994). Their early vocabulary size is a good predictor of later language skills (Hao et al., 2008). For instance, Marchman and Fernald (2008) found that vocabulary size at 25 months explained unique linguistic and cognitive variance at 8 years of age. Moreover, the rate, style and sequence at which children expand their vocabularies can vary within and across languages (Caselli et al., 1995). For example, British children aged 1 year (1;0) to 2 years and 1 month (2;1) scored lower than American children of the same age on a British vocabulary test. Even though receptive vocabulary seems to develop linearly over the first two years in both British and American children, there is much variation among and between them (Hamilton, Plunkett & Schafer, 2000). Since vocabulary seems related to later general development, it is an important issue to be further looked into.

This study aims to describe the vocabulary development of Chinese-, Korean- and Swedish-speaking children. Furthermore, differences in prominence of word categories in the vocabularies between children speaking different languages will be examined. Also, a view at potential (cultural) biases will be given. All this will be done based on data obtained from translated versions of the PPVT-4, which is a test for receptive vocabulary. These translations will be evaluated and implications for further developing translations will be suggested.

In the remaining part of section 1, important literature concerning vocabulary and its development and composition will be discussed, followed by a discussion on methodological issues.

1.1. **Productive vocabulary development**

Human vocabulary can be split up into a receptive or comprehensive vocabulary and a productive one (Bornstein, & Hendricks, 2012). Many researchers have studied various aspects of productive vocabulary development for children acquiring different languages, such as the stages in lexical composition, the relationship between the lexical composition in adult language and that in child language and the effect of vocabulary size on lexical composition in child language.

Bates and colleagues (1994) examined English-speaking children's productive vocabulary and found three waves of vocabulary organization. In the first wave with a vocabulary size up to 100 words, the vocabulary consists mostly of nouns. In the second wave with a vocabulary size from 100 to 400 words, verbs and adjectives start to enter the vocabulary. Closed categories like prepositions and determiners mainly enter in the third wave with a vocabulary size from 400 to 700 words. Word comprehension seems to develop linearly, whilst production growth looks exponential. These patterns found in English-acquiring children have been replicated for Italian-acquiring children by Caselli and colleagues (1995), and for Swedish-acquiring children Berglund and Eriksson (1994, as cited in Eriksson & Berglund, 1999). The latter did find great individual variation, however. Clark (1973, in Rescorla et al., 2013) also concluded that the lexical composition is similar for children across different languages.

Bleses and colleagues (2008) have compared data obtained from children acquiring 14 different languages in 18 studies, gathered via the MacArthurs' Communicative Development Inventory (CDI), and found a pattern of vocabulary development for children aged 8 months (0;8) to 1 year and 3 months (1;3) in all the languages similar to what has been found in previously mentioned studies. Specifically, a steeper development curve between 10 months (0;10) and 1 year of age (1;0), is present in almost all languages, indicating some kind of language-independent word spurt (Bleses et al., 2008). Slight language differences are, however, present at all ages. For instance, at all ages children speaking some languages (e.g. Danish) scored lower than children speaking other languages. It was nevertheless concluded

that there were more similarities than differences in vocabulary development across languages though (Bleses et al., 2008).

Bornstein and colleagues (2004) have examined slightly older children (1;8) acquiring different languages, using the Early Language Inventory (ELI), which is an earlier version of the CDI (Heilman et al., 2005; Bates et al., 1984). The languages under examination included Spanish, Dutch, French, Hebrew, Italian, Korean and American English. These languages differ in several important characteristics, like saliency and frequency of different word types. Bornstein and colleagues (2004) found a noun bias in all languages. Mothers in every country indicated that their children uttered more nouns than other word types. Korean mothers specifically mentioned that their children also uttered more verbs than adjectives. This difference in uttering verbs and adjectives did not reach significance however. Further, they found that the different word types correlated within a language, meaning that if a child had more nouns in his or her vocabulary, he or she also had more verbs, adjectives and closed classes. Moreover, they found that the size of the vocabulary could predict the composition of the vocabulary. Children with very small vocabularies, 0 to 50 words, did not show a noun bias: Nouns and verbs were equally often uttered. Children with a larger vocabulary (larger than 51 words) did show a noun bias. Children with a vocabulary of 101 to 500 words produced more verbs than adjectives. It was concluded that all word types thus developed concurrently with each other and that differentiation increases as vocabulary grows. Like Berglund and Eriksson (1994), Bornstein and colleagues (2004) also found variations between children and also between children speaking different languages.

In sum, productive vocabulary grows with age, seemingly exponentially, and with much variation between children. This pattern has been found at least for English-, Italian- and Swedish-speaking children. Also, for many languages a noun bias was found and differentiation across word types seems to increase with age. This all seems to indicate similar development of productive vocabulary across languages, albeit with limited variation within and across languages.

1.2. **The relation between productive and receptive vocabulary development**

The development of productive and receptive (or comprehensive) vocabulary seems to be related to each other (e.g. Eriksson & Berglund 1999). Berglund and Eriksson (2000) have tried to describe the communicative development, including lexical development, of Swedish-speaking children aged 16 to 28 months. Vocabulary was measured via a parental

checklist of 710 words. They found a word spurt when children had acquired 50 to 100 words in production. After 22 months, productive vocabulary development increased steeply for the children in the lowest 25th percentile. For the children in the top 25th percentile, vocabulary development was linear. In previous research on Swedish-speaking children, Eriksson and Berglund (1999) found that verbs were less common in productive vocabulary in comparison to comprehensive vocabulary and that comprehensive vocabulary was strongly related to productive vocabulary. Their results are in agreement with research on English-speaking children of Fenson and colleagues (1993, 1994, in Eriksson & Berglund, 1999). They both found that comprehensive vocabulary seems to precede productive vocabulary as well.

Hao and colleagues (2008) also found that receptive vocabulary was larger than productive vocabulary and that it grew faster than productive vocabulary. They have attempted to describe the vocabulary development of Chinese children. They adapted the CDI to a Mandarin Chinese version and collected data from 884 families with children aged 12 to 30 months. They also found that there were more individual differences in receptive vocabulary than in productive vocabulary, especially during the period between 12 and 16 months of age. This is consistent with data collected via the English CDI from American children (Hao et al., 2008). Growth rate was limited during this period, however. During the period between 17 and 30 months of age, vocabulary develops substantially and the large individual variation remains.

For both Swedish- and Chinese-speaking children it has thus been found that receptive vocabulary is related to productive vocabulary with receptive vocabulary preceding production.

1.3. **Later vocabulary development**

When looking at later vocabulary development, Gathercole and colleagues (1992) have tried to describe the receptive vocabulary development of children aged 4 to 8 years. They found that receptive vocabulary still develops linearly at these ages, albeit a less rapid development. Bornstein and Hendricks (2012) also looked at vocabulary development at older ages. They have compared 101,250 children aged 2 years (2;0) to 9 years and 11 months (9;11) and found that comprehensive vocabulary is only slightly larger than productive vocabulary at these ages but there was still a positive correlation between the two types of vocabulary. Wassenberg and colleagues (2008) studied later language development as well and found that language comprehension develops at least until the seventh grade.

Thus, receptive vocabulary still grow linearly at later ages, but it does not differ much from productive vocabulary anymore. It has been found that vocabulary develops at least until seventh grade, and perhaps even longer. All this emphasizes the need to look at vocabulary development not only at early ages, but also in early adolescence and beyond.

1.4. **Differences between word types**

An important characteristic of vocabulary development that has been found is that children acquire nouns earlier than verbs and that nouns outnumber verbs in early vocabulary (Gentner, 1982). Gentner (1982) gives as an explanation the natural partition hypothesis, which says that verbs are more difficult to map to real world observations than nouns, making them harder to learn. However, several researchers have argued that the noun bias is not a universal phenomenon (Gopnik & Choi, 1990, 1995; Tardif, 1996). Gopnik and Choi (1990) mentioned that the bias for nouns being learned earlier than verbs could be the result of methodological issues. More specifically, most research concerning this issue of lexical development has been conducted with English-speaking children and in cultures where object-naming is found to be relatively important (Gopnik & Choi in Caselli et al., 1995). In these cultures mothers often play word-naming games, where in most of the cases the object is a noun, emphasizing the importance of nouns (Kim, McGregor & Thompson, 2000).

However, not all cultures have games like object-naming and not all languages have the same typological characteristics as the English language. This makes it hard to generalize findings from English-speaking children to children in other cultures and languages.

Gopnik and Choi (1995) have for instance found that Korean-speaking children show more verbs in their speech than do English-speaking children of the same age. These results are in contrast to results like those of Bates and colleagues (1994) but are in agreement with the remark that their children produced more verbs made by Korean mothers in the study of Bornstein and colleagues (2004). Xuan and Dollaghan (2013) investigated the noun bias in bilingual English- and Mandarin-speaking children. When studying bilingual children, individual differences are minimized. Both the English and Chinese vocabulary size had to be between 50 en 300 words for this study to prevent noun biases due to different vocabulary sizes. The results were that nouns outnumbered verbs in both languages but in the vocabularies of the English-speaking children 54% of the words are nouns, while in the vocabularies of the Chinese-speaking children only 38% of the words are nouns. Xuan and Dollaghan (2013) also found that verbs were significantly more uttered in Chinese than in

English. Most of commonly used verbs were semantically heavy verbs, for example ‘eat’, ‘sit’ and ‘draw’, while the most used English verb was ‘go’, a semantically light and versatile verb. It was suggested that light verbs are more in common in languages like English, while semantically heavier verbs are more common in other languages (Xuan & Dollaghan, 2013). However, when looking at verbs that were uttered by at least 50% of the children, more semantically heavy verbs appear in the English vocabularies. The Chinese verbs in the children’s’ vocabularies still largely outnumber the English verbs, though: 33 verbs were found in the children’s’ Chinese vocabularies, and only 5 in their English ones.

An explanation for the found difference in proportion of nouns and verbs is that in Korean SOV order is used in and Chinese SOV order can be used (Li & Bates, 1993; Li & Thompson, 1974). This word order places the verb in a more salient position in contrast to SVO languages like English. Additionally, in some languages such as Chinese and Korean, arguments like subjects and objects are often dropped. This leaves a sentence with only a verb. In these languages, verbs are not only more salient, but also more frequent in maternal input than nouns (Imai et al., 2005). This too might be an explanation of children speaking those languages having more verbs in their vocabularies than children speaking other languages (Tardif, 1996). However, Chinese is markedly different from other languages like Korean, in the fact that Chinese verbs are not inflected (Imai et al., 2005), which makes it harder to the distinguish verbs from nouns. Korean verbs are practically always inflected and can thus be more easily recognized as being verbs (Kim et al., 2000). An explanation for verbs being more prominent in vocabularies of Chinese-speaking children might thus be that those verbs are not known to be clear verbs, but as verbs that seem similar to nouns.

It should be noted that other studies on Korean- and Chinese-speaking children did not find a verb bias. For instance, Kim and colleagues (2000) found that the Korean-speaking children studied had more nouns than verbs in their vocabularies, although they do learn verbs more easily than English-speaking children. Imai and colleagues (2005) and Rescorla and colleagues (2013) found that there is no difference at all between verb learning for Chinese- and English-acquiring children. Tardif and colleagues (1996) only found more verbs than nouns in the vocabulary of Mandarin Chinese-speaking adults, but not in the vocabularies of children. These inconsistencies might too be a result of methodological issues like data collection (Imai et al., 2005). Tardif and colleagues (1996) already pointed out that the context in which vocabulary was measured was related to the found number of nouns uttered: When looking at picture books, children used more nouns than verbs, while during

playing with toys equal number of verbs and nouns are uttered. Xuan and Dollaghan (2013) also claimed that it is difficult to compare studies about vocabulary composition, considering most studies differ on age groups, vocabulary sizes and ways that nouns are counted and identified.

The noun bias that seems to be universal may thus not be so universal after all. At least, there is not yet a consensus on the universality of this noun bias. Explanations for some children having more verbs in their vocabularies than other children might be that it is because of language typological features, or that the differences found were based on methodological issues.

1.5. **Methodological issues**

A few methodological issues have already been pointed out in the preceding sections, but there are more, which will be discussed in this sub-section. Several methods of measuring vocabulary and vocabulary development can be used. Frequently used methods are parental reports based on diaries or a checklist of words like the commonly used MacArthurs Communicative Development Inventory (CDI). Also audio recordings can be used, as well as real time observations (Berglund, & Eriksson, 2000). Parental reports, however, have the disadvantage of likely being biased. Parents often portray their children as better than how they actually are, potentially giving a wrong idea of, in this case, actual vocabulary (Hamilton, Plunkett, & Schafer, 2000). The CDI is a widely used measurement, and is suitable to gather information about vocabulary of children aged 8 months to 2 years and 4 months (Hao et al., 2008). But Pine, Lieven and Rowland (1996) raised a methodological topic about this test. They have reviewed studies using different measurements to measure vocabulary and found that especially checklists like the CDI find higher proportions of nouns in vocabularies than do observational data. But observational data, however, is hardly ever controlled, which makes it hard to distinguish in what kind of context an utterance was made (Tardif et al., 1999). For example, a noun bias in observed production data might have been evoked by maternal responses or a playing game. Thus, most types of measurements have their limitations.

The Peabody Picture Vocabulary Test (PPVT), first published in 1959 by Dunn, is another widely used measurement for testing receptive vocabulary in both children and adults and is not based on parental reports, which makes it a preferable instrument (Dunn, & Dunn, 2007). The PPVT-4 is the latest edition and is a revised version of earlier editions. The test is

normed on an American population consisting of 3,540 inhabitants of the United States, aged 2 years and 6 months (2;6) to 90 years and older. This norm makes it possible to compare individual scores to a representative sample of US citizens of the same age group.

The goal of every test that intents to measure human behavior is to be unbiased, meaning that every group should have equivalent chances to perform well, if it is expected that all groups have the same abilities. No group preferences should thus be present if no differences are expected (Stockman, 2000). If the norm group from the test matches to the actual group to be tested, this should be the case. However, several studies found that the PPVT does not meet this norm for all cultural and ethnical groups and is thus not unbiased, even for within the US population. This is, considering one is not expecting differences based on culture or ethnicity (Haitana et al., 2010; Stockman, 2000). A reason for this is that the first version of the PPVT did not include minority populations in the norm group at all and thus the norm did not apply well to the actual, culturally different population. In the second version, the PPVT-R, the norm group consisted of 14.6% people from minority cultures (Stockman, 2000) but biases in scores on the PPVT-R were still found. African American children scored around one standard deviation below the mean scores of White American children (Rock, & Jenner, 2005). With the coming of the PPVT-III, the percentage of people from minority groups in the norm sample was heightened to 34%, to accurately represent the minority groups in the actual total population (Stockman, 2000). Significant differences were not found any more between the two groups, indicating that earlier found differences are of methodological ground and not because the African American population actually has a significant lower receptive vocabulary. However, the norm was reduced by including more minority people, so actual differences between the groups might still be present. This seems to be the case, since African American preschool children did still score 1,5 standard deviation below average, but it did not reach statistical significance anymore (Washington, & Craig, 1999). Washington and Craig (1999) thus concluded that the test was suitable to measure this population's receptive vocabulary and that the test was unbiased for ethnicity, based on the criterion as mentioned by Stockman (2000) that all group means should be equivalent.

Qi and colleagues (2006) also studied the African American population, but focused on people with low socioeconomic status (SES) in specific. They too found that the African American children scored 1,5 standard deviation below average but in this study it did reach significance. The test also indicated a larger group of children as having language delays than

other tests. Still, they did conclude that the test was suitable for the population, but test results must be interpreted with caution: It could be possible that this group actually has a lower receptive vocabulary, but differences could also occur because of methodological inequalities. Restrepo and colleagues (2006) looked at the same population and they compared the PPVT-III to another test to measure vocabulary, the Expressive Vocabulary Test (EVT). African American preschool children scored lower than average on both tests, but the difference was larger on the PPVT-III. SES was not a significant predictor in this study, in contrast to those of Washington and Craig (2000) and Qi and colleagues (2006). Restrepo and colleagues (2006) concluded that the EVT was a better indicator for vocabulary. However, the EVT measures expressive vocabulary, while the PPVT-III measures receptive vocabulary, which makes it hard to compare the tests. It was again suggested that results from the PPVT-III should be interpreted with caution, meaning that it should be considered it is possible that the score does not reflect the actual vocabulary size. Preferably a different test that does not have this potential cultural bias should be used to assess vocabulary in African American children

As for other populations outside the USA, Haitana and colleagues (2010) looked at data gathered via the PPVT-III among Maori children living in New Zealand. The test has been administered to the New Zealand population and it turned out to be a valid test according to Stockman's (2000) criterion of equal group means. But for the specific minority group of Maori children, it turned out that the PPVT-III did give equal scores for Maori children who went to mainstream schools in comparison to non-Maori children, but not for Maori children going to Maori schools. Children going to Maori schools scored around one standard deviation below children going to mainstream schools. This could indicate that they really do have a lower receptive vocabulary or that the test items do not reflect words used and taught in the Maori culture. Either way, the criterion of equal group means is not met. Haitana and colleagues (2010) also found that all children, regardless of school or age, had difficulties with some of the same items. These items seemed to characterize concepts known in the USA, but not in New Zealand, for example animal types that do not live in the Pacific area. As a suggestion, the researchers advised to replace such items for items more commonly known in the country where the testing will take place (Haitana et al., 2010).

As can be noticed, most of the research concerning this topic has focused on the PPVT-III or older versions and not the PPVT-4, because of the relatively short time the PPVT-4 has been in use. Several researchers, like Haitana and colleagues (2010)

acknowledge that in the PPVT-4 some of the cultural issues might already have been solved. For example, in the PPVT-4 the item pool has been adapted based on suggestions made by professional users of the PPVT. Easier items have been added in the earlier sets, so that even vocabulary in children functioning at low levels can be measured and items for children aged 10 to 13 years in particular have been revised (Dunn, & Dunn, 2007). Also, the number of items in the category verbs has been reduced, because they turned out to be too hard for (young) children. All PPVT-III items have been reviewed by Canadian, British and Australian professional users of the PPVT to identify words that were considered inherently American or could not be used across countries and cultures for other reasons (Dunn, & Dunn, 2007). Based on these reviews, modifications to the items were then made. In total, 75% of the items of the PPVT-III have been used (with or without modifications) in the PPVT-4. The other 25% are new items, which were also thoroughly tested. Few studies with actual data on the PPVT-4 have been conducted however, so the applicability of the newest version has not been systematically examined yet. Also, no unofficial translated versions without norms have been reviewed, which is what will be done in this study.

2. **Current research**

It can be concluded from the research discussed above that the size and manners of (early) vocabulary seem to develop similarly across languages, since there are more similarities than differences (Bleses et al., 2008). There is much individual variation within and across languages, though. Also, there is not yet a consensus about the composition of the vocabularies and whether or not verbs are more prominent in vocabularies of children speaking languages like Chinese and Korean. A noun bias for all languages has been found in most studies. There are also some methodological issues like the cultural biases of measurements of vocabulary concerning the topic. Another striking point is that the majority of the research has been done with young children, mostly aged younger than two years. Few studies have focused on vocabulary development at later ages, like that of children in primary schools even though it turned out to develop at least until seventh grade (Wassenberg et al., 2008). Another important point to mention is that in most methodological studies the PPVT-III was examined and there is limited availability of studies on the PPVT-4. To complement the existing research, this study will try to take these points into account. It will concentrate on an older group of children and will use data of the PPVT-4. A view on potential cultural biasing of the versions used in this study will also be considered.

The issues discussed above have led to the following research questions: (1) 'Is receptive vocabulary development of children speaking different languages across ages comparable?' It is hypothesized that development looks indeed the same. According to Stockman's (2000) criterion the test should lead to equal group means, and thus to similar developmental patterns. If the test is usable and unbiased across languages and cultures, large group differences should not be expected, since vocabulary development seems to develop similarly in different cultures (Bleses et al., 2008; Bornstein, et al., 2004). Differences between individual children can be expected, and also some small differences between children speaking different languages, but large group differences are not expected. (2) 'Do children speaking different languages perform equally well across different word types?'. If one type of words is more common in one language related to another language, it is to be expected that children speaking those languages will score higher on those items, because those word types are more frequent in their vocabularies. What is expected for the languages specifically studied in this research is that all children show a noun bias, regardless of language. Chinese- and Korean-speaking children will perform better on verbs than Swedish-speaking children, because of the SOV style of Chinese and Korean. Korean-speaking children will perform even better on verbs than Chinese-speaking children because of the fact that verbs are non-inflected in Chinese and thus harder to identify as such.

3. **Method**

3.1. **Participants**

The group of participants consisted of 112 children from three different countries: China (N = 23), South-Korea (N = 34) and Sweden (N = 55), all native speakers of their languages. Of the total group of children 49.7% is male (N = 56). The percentages of sex per country can be seen in Table 1 below. Data from children aged 4 to 11 years have been collected. The number of participants per age group per language can be found in Table 2. Mean age of the total group is 87.7 months (SD = 27.8) or 7;3 years. Mean age per language is 72,1 months (SD = 15,8) or 6;0 years for the Chinese-speaking children, 94,4 months (SD = 29,3) or 7;10 years for the Korean-speaking children and 91,6 months (SD = 26,6) or 7;7 years for the Swedish-speaking ones.

Table 1

Distribution of Sexes Across Languages

	Male (N)	Male (%)	Female (N)	Female (%)

Chinese	12	52,2	11	47,8
Korean	20	58,8	14	41,2
Swedish	24	43,6	31	56,4
Total	56	50,0	56	50,0

Table 2

Number of Participants (N) per Age Group in Years per Language

	Age							
	4	5	6	7	8	9	10	11
Chinese	4	10	5	2	0	1	1	0
Korean	7	8	0	4	4	0	7	4
Swedish	8	10	5	4	13	0	11	4
Total	19	28	10	10	17	1	19	8

The 4 and 5-year old Chinese- and Swedish-speaking children attended kindergartens. All other children attended elementary schools in their own country at the time of testing. One Chinese-speaking boy was excluded from further analysis because of missing information about the date on which he was tested. One Korean-speaking boy was excluded for the same reason. Total number of participants for further analyses is 110.

For additional analyses, data from adults will also be analyzed. Data from 15 native Chinese-speaking adults and 20 Korean-speaking adults will be examined. Unfortunately, data from Swedish-speaking adults is not yet available and can therefore not be analyzed. Mean age from the Chinese-speaking adults is 19;3 (SD = 7,3) and mean age for the Korean-speaking adults is 23;7 (SD = 31,1). Of the Chinese-speaking group 40.0% is male, 55.0% of the Korean-speaking group is male. All Chinese-speaking adults were students at Beijing Forestry University at the time of testing and the Korean-speaking participants were studying at Hanyang University in Seoul, South-Korea.

3.2. **Measures**

The PPVT-4 is a test to measure receptive vocabulary of English-speaking children aged 2 years and 6 months to adults of 90 years and older (Dunn, & Dunn, 2007). The test was translated to Chinese, Korean and Swedish by linguistically trained native speakers. The Chinese version has been translated by two native speakers of Mandarin Chinese. Between-

translator checks have been done and online dictionaries were consulted. According to the translators, there were two types of difficult cases. 1) Some words had more than one possible translation: A formal one and an informal one, where the formal version is used more in written language while the informal word is more commonly used in spoken language. 2) Unclear or confusing words in English. In case 1 the informal option was chosen for items in the earlier sets, the formal one for the later sets. In the case of the second problem the translation that best matched the accompanying picture was used. A list of difficult items and the possible answer options can be found in Appendix A. The Korean version has been translated by one native speaker of Korean who used dictionaries as well but also checked word frequencies. Difficult cases were words where word frequency in Korean did not match word frequency in English. This was the case for a lot of loanwords in English. Those low frequency words were chosen anyway, because it was not possible to correctly translate them otherwise. The Swedish version has been translated by native Swedish-speakers at the speech therapy education in Goteborg but no further information is available about the translation. These translated versions were administered to the participants described above. No norms are available. There are two forms, A and B, which are similar in construction, but have different items. Form A has been administered to all children participating in this study.

The items of the PPVT-4 are based on previous versions of the PPVT. As mentioned before, 75% of the words match to those of the PPVT-III, the other 25% percent are new words. The words are selected from the Webster's New Collegiate Dictionary (1953; 1981), The American Heritage Word Frequency Book (1971), the EDL Core Vocabularies in Reading, Mathematics, Science and Social Studies (1989), Basic Reading Vocabularies (1982), Educator's Word Frequency Guide (1995), Merriam-Webster's Collegiate Dictionary (2003), The Reading Teacher's Book of Lists (1993), and Merriam-Webster's Collegiate Thesaurus (1988) (Dunn, & Dunn, 2007). In the fourth edition of the PPVT, a balanced representation of different word categories was attempted. Categories like body parts, clothing, emotions, fruit and vegetables and musical instruments are represented (Dunn, & Dunn, 2007).

The PPVT-4 consists of 228 items, divided into 19 sets of 12 items. Per item, there are four possible answer options, shown as pictures drawn by hand on a computer screen. Children are given the target stimuli orally and asked to indicate the picture that represents the heard item. There are four possible answer options: The correct item and three distracter items that are based on semantic resemblance to the correct word, not on phonological

resemblance. Each distracter is supposed to be equally attractive, hence, no purposely misleading distracters are present (Dunn, 1997). The answer is correct if the right item has been indicated and only one answer per item is correct.

Each child starts with the set that matches his or her age at the date of testing. First the basal set is established. This is the set in which the child makes one or zero mistakes. For most children, this is the first set that is administered. If the child makes more than one mistake in this first administered set, the preceding sets are administered, until the child makes one or zero mistakes in the same set. After establishing this basal set the child continues with the next sets to establish the ceiling set, which is reached if a child makes eight or more mistakes in the same set. The total testing score is the item number of the last administered item from the last set, minus the false items.

However, most statistical tests assume independent measures. This assumption means that all observations are independent from each other and thus not correlated. When using the total testing score, this assumption is not met, since answers on items correlate within a person and items also correlate with other items within and between sets. So for the analyses done in this study, all answers of the children have been marked as either correct or incorrect, which makes the data binary and gives a maximum of 228 items or observations per child. With binary data, analyses can be done that do not have the assumption of independent measures (Quené & Van den Bergh, 2008). Total testing score is thus not used in this research. Also, using the binary data gives more useful information in this case than total score because now individual items can be further looked into.

Since the question remains as to whether or not language is related to receptive knowledge of word categories, a categorization based on word categories has been made. The test consists of three types of words: nouns, verbs and adjectives. Of the 228 items, 149 have been categorized as nouns, 42 as verbs and 37 as adjectives. This categorization has been made in consult with a bachelor student of English language and cultures, since the manual of the PPVT-4 itself does not specify the word type of every item. Both the author and the bachelor student of English categorized the words. There was an agreement of 98.7 percent between the two raters. They differed on only three items: item 70, 78 and 186, respectively the words ‘uniform’, ‘squash’ and ‘ascending’. These items have been categorized as two nouns and an adjective, in that order, corresponding to the categorization made by the author. Those categorizations seem to match better with the accompanying pictures. For example, the word ‘squash’ could be a verb, but in this case the vegetable ‘squash’ was pictured. Therefore

it was chosen to be categorized as a noun and not as a verb. For a list of all words and their categories, see Appendix B.

3.3. **Procedure**

All tests were administered individually by experienced testers in each language. It took about 20 to 30 minutes to administer the test on average per child and this was done during classes, to make sure it would not cost spare time of the children, their parents or teachers. All the children were tested once for this purpose. Administration of the test has been video recorded.

During the testing phase in China, it has been decided that all children would complete the entire test, even if they had made eight or more mistakes in the same set already. This was done because even the youngest children seem to make few mistakes overall, which was taken as an indication that the translated test might be easy for the Chinese children. The whole test was administered, so that all possible data could be collected. For the children in the other countries the rules for aborting the test were followed as indicated in the PPVT-4 manual.

However, in the Swedish data collection the basal and ceiling set rules of PPVT-III have been used by accident in all cases, instead of those from PPVT-4. In the PPVT-III, four mistakes are allowed in the starting set and nine mistakes in the ceiling set, while in the PPVT-4 maximum one and eight mistakes are allowed respectively. The PPVT-III gives thus more room for errors within a set than the PPVT-4 and this has led to wrongly administering more sets than would have been done if the PPVT-4 rules would have been used. In total, 5,103 items, or 31.7%, of the total of 16,105 items have been wrongly administered. This includes the errors in the Swedish data gathering, the extra data gathered from the Chinese children and other, accidental errors in the rest of the data. If the rules from the manual were correctly followed, all these items should not have been administered. How many items should have been administered to establish correct basal sets for the Swedish-speaking children cannot be estimated because this data is simply not gathered. If the child made between one and four mistakes in the first administered set, the administration continued whilst earlier sets should have been administered if the PPVT-4 rules were followed. If the basal set has not been established correctly, all items were marked as wrongly administered. If all the collected data would be used, it is possible it does not reflect actual receptive vocabulary anymore. When children made more sets than they were supposed to, the chance

of getting a better total score is higher, but they probably score lower on the later individual items if one assumes that these harder items do not yet exist in their vocabularies. The same goes for children whose basal set has not been established correctly. Their total score could be higher than it would have been if the basal set has been correctly established and now they probably have higher scores at item level for these items. For the purpose of this study, several analyses have been made, either with or without the scores of the items that were wrongly administered. This will be further discussed in the result section.

Tests to adults have been administered according to the rules as mentioned in the PPVT-4 manual.

3.4. **Analysis**

It is assumed here that the items in the PPVT-4 are correlated amongst each other, at least within a set, because items within a set are selected based on the fact that they are supposed to be of the same level of difficulty (Dunn, & Dunn, 2007). Answers to different items within an individual are also correlated to each other. This leads to a hierarchy in the data on two levels, on the level of the individual and on test level because of the correlation of individual items. Quené and Van den Bergh (2008) suggested this type of analysis for binary data, because binary data violates the assumption of normally distributed data of an ANOVA, but can be analyzed with multilevel analyses. Multilevel analysis also reduces the chance of making a Type I error, which means that the chance of falsely rejecting the null-hypothesis that there are no differences between groups, is reduced. This makes it a more balanced and conservative method (Quené & Van den Bergh, 2008). Based on this multilevel analysis a growth curve will be given.

An additional analysis was done to see if there are items that are perhaps culturally biased. Items that have a chance less than fifteen percent to be answered correctly by children in every language will be sorted out and be looked into further. Previous studies did not mention using such criteria; therefore this new criterion is proposed. In order to see if these items are just not known by children or not known by adults either, test results of the adults will be examined as well. Items that less than twenty percent of the adults answered correctly will be listed. Because the group of adults is much smaller than those of children, it is chosen to use a boundary of twenty percent instead of fifteen percent so that any possible tendencies are not missed because of a too harsh cut-off point.

Besides the expected language differences, a gender difference might be found. Amongst others, Rescorla and colleagues (2013) found significant gender differences in vocabulary and Platt (2010) already found a gender difference in scores on the PPVT-III in specific, with girls outperforming boys in both studies. The finding of Platt (2010) did not reach significance, though. With scores from the PPVT-4, it will be tried to replicate these results.

For all analyses, an alpha level of .05 will be used. How well children performed will be measured based on the probability correct scores. These scores depict the chance of answering a random item correctly. The higher the probability correct chance, the better the children performed.

4. **Results**

4.1. **Development across ages**

When looking at all the data, including the wrongly administered items, age turned out to be related to testing score as expected ($p < .001$). As one gets older, vocabulary increases. Additionally, language also turned out to be a predictor for testing score ($p = .013$). A significant interaction between age and language was also found ($p = .001$). When looking closer at which languages differed from each other, it can be seen that children speaking all languages differed significantly from each other. The Chinese-speaking children differed from the Korean- ($p < .001$), and Swedish-speaking children ($p < .001$) and the Swedish- and Korean-speaking children also differed from each other ($p = .008$). The Chinese-speaking children performed better than the Korean- and Swedish-speaking children, and Korean-speaking children performed better than the Swedish-speaking children. The interaction effect can be seen in that the Korean-speaking children fall behind initially in comparison to both the Swedish- and Chinese-speaking children, but soon exceed the Swedish-speaking ones.

If the wrongly administered items are not included in the analysis, age still had a significant main effect ($p < .001$). As can be seen in Figure 1, the older one got, the better the probability correct score was. Important to mention here is that there was only one participant aged 9, which could explain the unexpected peak at age 9. Also the groups of participants aged 6, 7 and 11 were relatively small. So this pictured tendency is perhaps not completely accurate and might show some incidental fluctuations that would not have occurred if group size was larger. Language has been found to be even more significant than in the previous analysis ($p = .005$) and there was also still an interaction effect between age and language (p

= .009). The Chinese-speaking children differed significantly from both the Korean- (p < .001) and Swedish-speaking children (p <.001), but the Swedish- and Korean-speaking children did not differ from each other anymore (p = .334). This can be seen in Figure 2. Thus the Chinese-speaking children performed better than Korean- and Swedish-speaking children, but now Swedish- and Korean-speaking children performed statistically equal. Sex did not turn out to be significant, either with (p = .413) or without (p = .996) the wrongly administered data.

Using all data, including the wrongly administered items thus does not affect which variables are related to the probability correct scores. Only the level of significance of these variables changes somewhat.

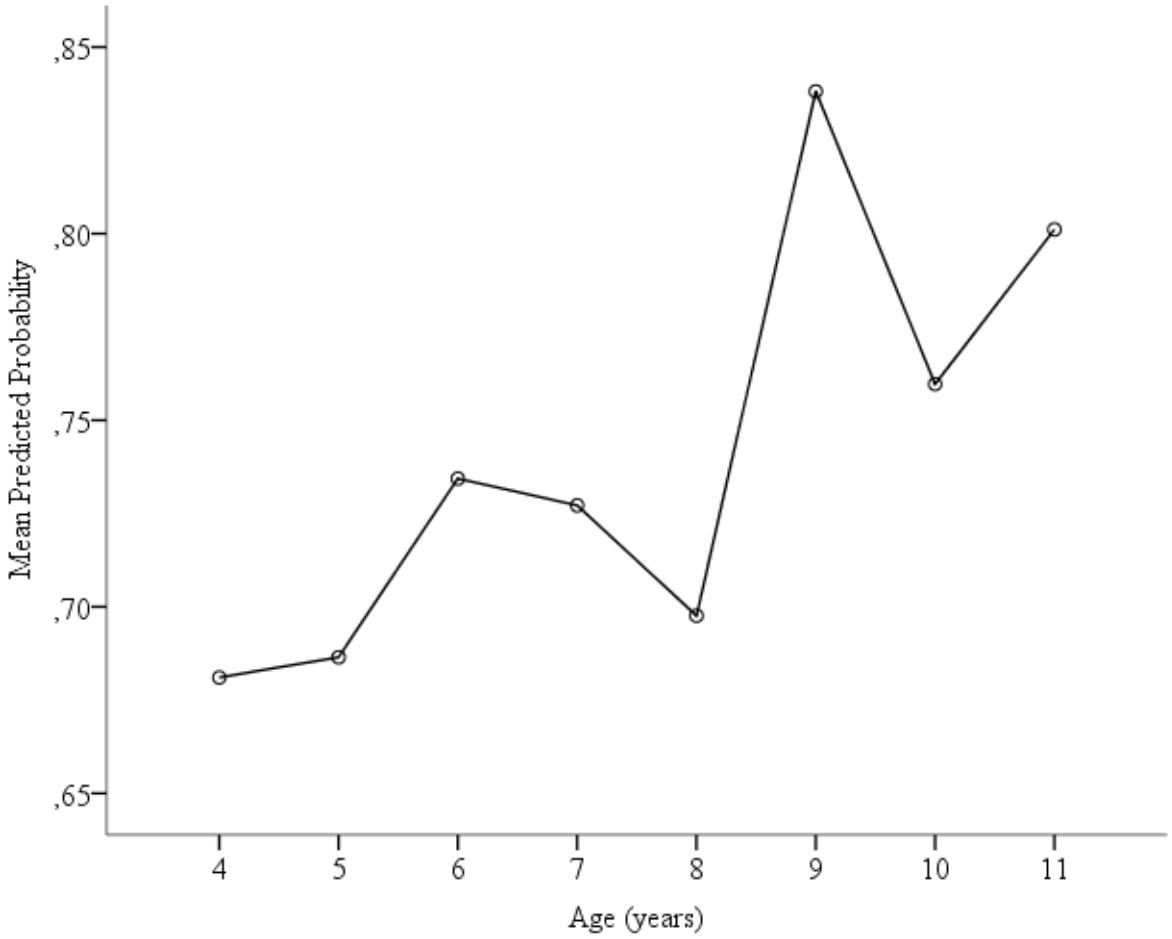

Figure 1. Main effect of age in years, on the mean predicted probability for probability correct chance.

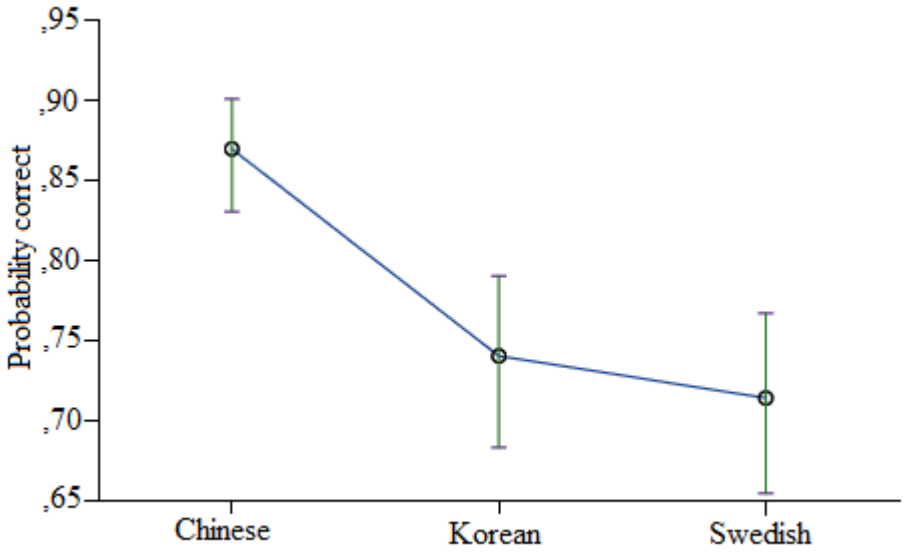

Figure 2. Main effect of language on the probability correct chance.

In Figure 3, a graph can be found of the probability correct chance per language and across ages, solely based on the well administered data. As can be seen, the Chinese-speaking children approach a probability correct of almost 1 at an early age (51 months or 4;3 year). For the Korean-speaking children this is 84 months or 7;0 and for the Swedish-speaking children 90 months or 7;6. This score means that the chance of getting a random item correct is approaching a probability of 1,00. As mentioned before, the higher this score is, the better the children performed. Also, the interaction effect between age and language is visible in this figure. The Korean-speaking children surpass the Swedish-speaking children at an age of 71 months or 5;11.

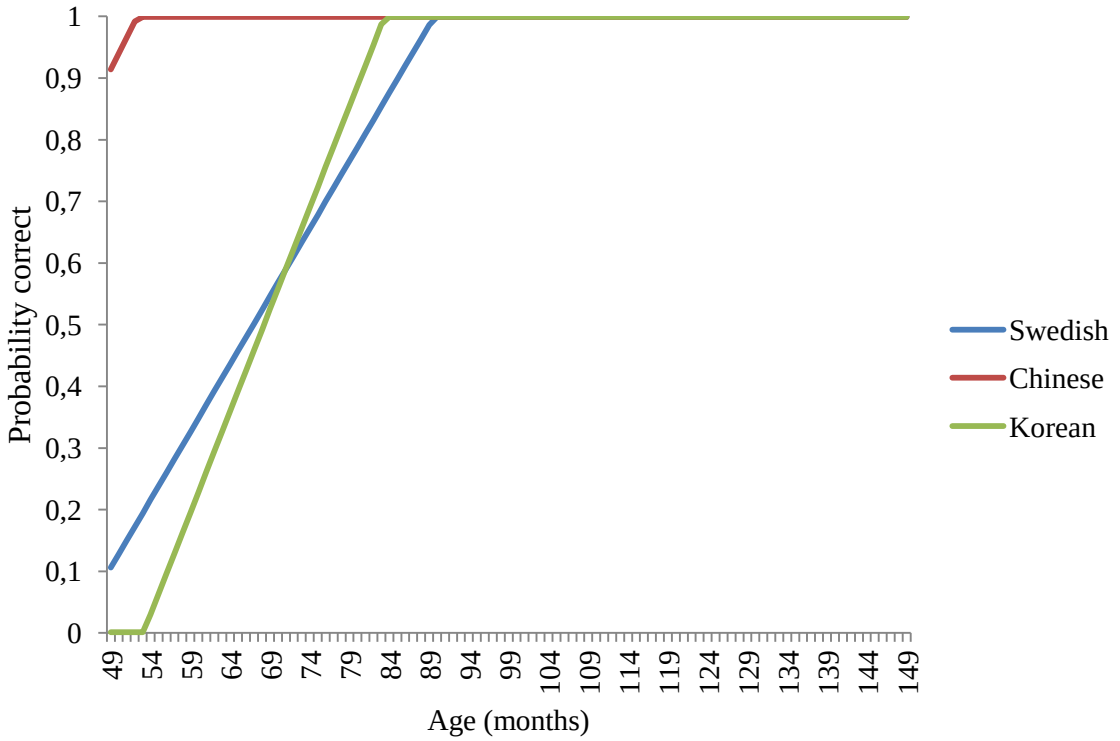

Figure 3. Probability of getting a random item correct, across ages and languages.

4.2. **Differences between languages in word categories**

To see whether or not item type is related to the testing score, this variable has been added to the model. It turned out to be significant ($p < .001$). The children scores significantly different on verbs than on nouns ($p = .010$) and adjectives ($p = .001$), although their scores on nouns and adjectives did not significantly differ from each other ($p = .173$). Verbs were performed worst on. There was an interaction effects found with item type and age ($p = .029$) but not with language ($p = .246$) or both age and language ($p = .253$). For this analysis, all available data has been used.

When only the correctly administered data are used, the results are similar. Item type is still significant ($p < .001$), as can be seen in Figure 4, and there is still an interaction effect between item type and age ($p = .040$). Verbs are still different from nouns ($p = .024$) and adjectives ($p = .016$). Performance on adjectives still does not differ from nouns ($p = .494$). Again, verbs are performed worst on. As for the interaction effect, it turned out that as children get older, their performance on adjectives gets better, but not on nouns and verbs.

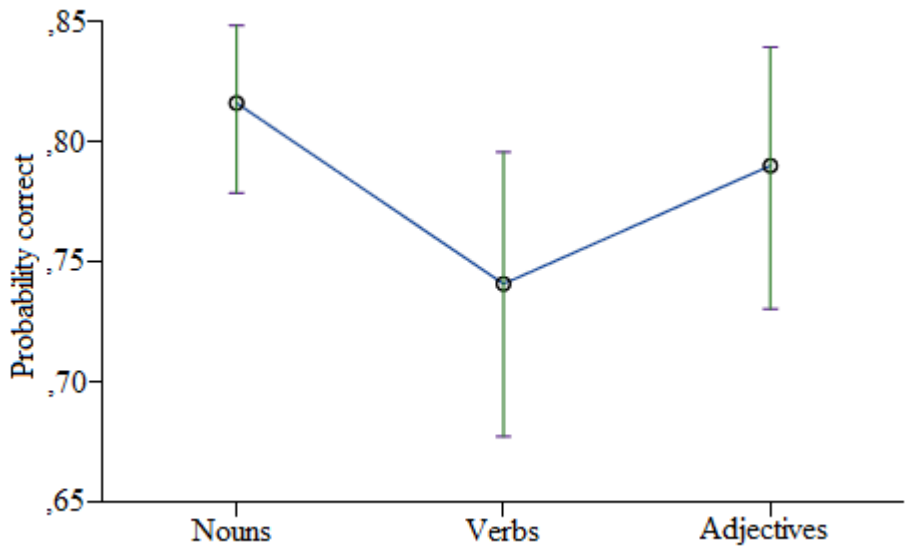

Figure 4. Main effect of item type on the probability correct chance.

4.3. **Differences between languages at item level**

A supplementary analysis has been done on the item level to see which items were harder to answer for children speaking one language compared to another language. It turned out that for the Swedish-speaking children fourteen items had a chance less than 15 percent to be answered correctly. For the Korean-speaking children, only four items have a chance this low and for the Chinese-speaking children only three items. A list of these items and the corresponding words can be found in Table 3. All data has been taken into account for this analysis, including the wrongly administered data, to gather as much information at item level as possible.

Table 3

Items With the Lowest Chance on Getting It Correct per Language.

	Swedish			Korean			Chinese	
Item	Word	%	Item	Word	%	Item	Word	%
138	Injecting	.10	56	Buckle	.14	126	Snarling	.09
157	Primate	.13	100	Wrench	.15	185	Wedge	.15
161	Transparent	.02	214	Embossed	.12			
162	Sedan	.11						
177	Cerebral	.07						

185	Wedge	.08
196	Upholstery	.11
199	Trajectory	.14
201	Barb	.06
213	Convex	.12
217	Legume	.04
221	Vitreous	.11
223	Caster	.04
228	Tonsorial	.14

To see if the items in the table above are also performed worse on by adults, adult data has been examined as well. However, the youngest Chinese-speaking adult participant was 18;4 and the youngest Korean-speaking one was 19;0 so the first administered test was set 13, beginning with item number 145 for the Chinese-speaking group and set 14, starting with item number 157 for the Korean-speakers. Data from previous sets is not available and therefore items before number 145 and 157 cannot be compared. The items made worst by the Swedish-speaking children cannot be compared at all, since data of Swedish-speaking adults does not exist yet. Items answered correctly by less than twenty percent are listed in Table 4.

For the Korean speakers, this leaves only item 214, ‘embossed’ to be compared. This item has been answered correctly by only fourteen percent of the adults. This is also the only item that less than twenty percent of the adults answered correctly. For Chinese speakers, only item number 185 , ‘wedge’ can be compared, because item number 126 has not been administered to adults. Of the Chinese-speaking adults, 80% of the group answered item 185 correctly. Item number 214, ‘embossed’, is the only item that has been answered correctly by less than twenty percent of the Chinese-speaking adults.

In Table 4 a list of all items with a chance of being answered correctly by adults less than 50% has been given. These items were initially not considered to be notable, but since the group sizes were relatively small, it is possible that some of these items would have been answered correctly by less than twenty percent if the group size would have been bigger. Therefore, these items are mentioned anyway. Additionally to the above discussed items, none of the other items match to those performed worst on by children. As for the Korean words, the translator indicated these words all as loanwords. As was mentioned before, direct

translations from these loanwords are very low in frequency. This could explain the difficulties the adults seem to have had.

Table 4.

Items with the lowest chance of being answered correctly by adults.

Chinese			Korean		
Item	Word	%	Item	Word	%
191	Dejected	.47	172	Poultry	.48
199	Trajectory	.47	185	Wedge	.29
211	Cupola	.27	196	Upholstery	.38
212	Derrick	.40	214	Embossed	.14
214	Embossed	.20	220	Supine	.38

Since only item 214 for the Korean speakers and item 185 for the Chinese speakers can be compared, not much can be said about the similarities between the children and adults participating. Item 214 has been performed worse on by both the Korean adults and children, indicating there is something in this item that makes it hard. Item 185 was only performed worse on by children, not by adults. These items will be further discussed in the discussion.

4.4. **Final results**

In the final model, age, language, the interaction between age and language, item type and the interaction between item type and age are included. This final model is based solely on the well administered data and is different to the model discussed in 4.1 since item type has been added. The total model is significant for predicting the chance of getting an item correct, $F(9, 11.934) = 29.57$, $p < .001$. Age is as expected a significant predictor, with $F(1, 11.934) = 80.99$, $p < .001$. The older the children are, the better they performed. Language is also a significant predictor ($F(2, 11.934) = 6.24$, $p = .002$): The group mean for the Chinese-speaking children is significantly higher than for the Swedish- and Korean-speaking children. Also the interaction effect between age and language is significant, with $F(2, 11.934) = 4.21$, $p = .015$. The rate at which the performance grows with age, is not the same for children speaking different languages. The Korean-speaking children show a steeper developmental pattern than the Swedish- and Chinese-speaking children. Also item type is a significant predictor for performance ($F(2, 11.934) = 4.22$, $p = .015$). Verbs turned out to be harder than

nouns and adjectives. Lastly, the interaction between item type and age $F(2, 11.934) = 3.22$, $p = .040$) is a significant predictor. As the children get older they get better at adjective but their performance on nouns and verbs does not improve.

In Figure 5, the estimated means of getting an item correct per language are displayed, based on this model.

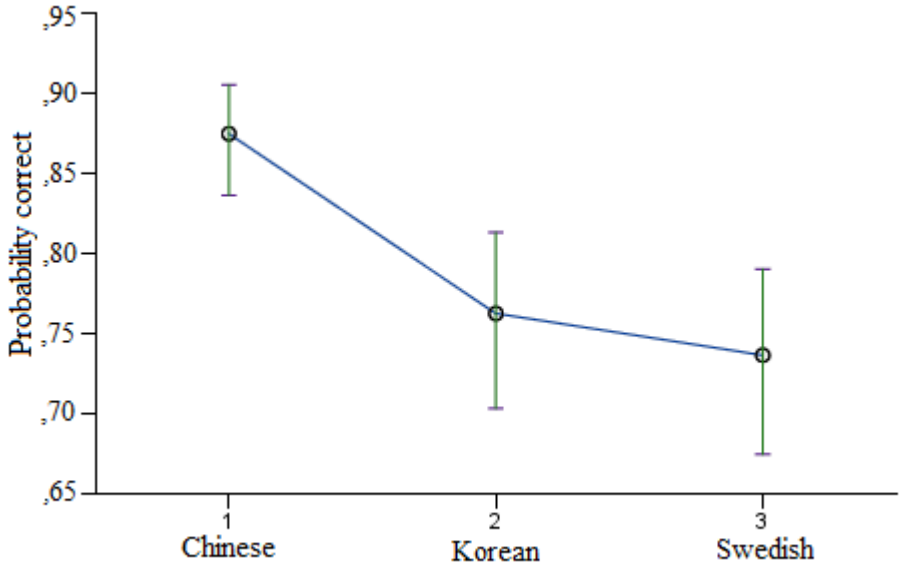

Figure 5. Estimated means of the probability correct per language, based on the final model.

5. **Discussion**

5.1. **General conclusions**

Two research questions were raised at the beginning of this study, namely 'is receptive vocabulary development as measured by the PPVT-4 of children speaking different languages across ages comparable?' and 'do children speaking different languages perform equally well across different word types?'. These questions can now be answered.

With respect to the first question, receptive vocabulary development as measured by the PPVT-4 is not the same across languages. The developmental pattern is similar, but the age at which a child has a good chance of answering an item correctly is very different, as can be seen in figure 1. The Chinese-speaking children outperform the other two groups of children significantly and have reached the point of having a high chance to answer correctly on a random test item already by the age of 4;3, while the Korean- and Swedish-speaking children reach this point at 7;0 and 7;6 respectively. The Korean- and Swedish-speaking children have statistically comparable scores, but the developmental patterns differ. The

Korean-speaking children seem to fall behind the Swedish-speaking children initially, but soon exceed them at an age of 71 months or 5;11. This can also be seen in figure 3. This is in contrast to the hypothesis that was made, namely that all children would perform equally well and have the same developmental pattern.

The second question can also be answered. The children speaking different languages perform indeed equally well across word types. No interaction between item type and language was found. This is in contradiction to the hypothesis that the Chinese- and Korean-speaking children would outperform the Swedish-speaking children on verbs, and that the Chinese-speaking children would even outperform the Korean-speaking children. Verbs were significantly harder for all children, regardless of language, and there was no significant difference between performance on nouns and adjectives, although performance on adjectives was worse than on nouns. The noun bias that was expected was indeed found for the children all languages. This result is in agreement with those of Imai and colleagues (2005). An unexpected interaction effect with item type and age was also found for the children speaking all languages. That is, as a child gets older, the chance of getting a noun or verb correct gets lower. This seems to be against expectations, but could be explained by the fact that testing items get more difficult in the later sets (Dunn, & Dunn, 2007). It might be the case that the difficulty level of the items got harder in comparison to the development the children made, leading to a decrease in chance of getting those items correct.

5.2. **Conclusions about the translations**

To evaluate the translations that were made, additional analyses at item level have been conducted. Items with a very low chance of being answered correctly have been determined for all three languages. What is evident is that there are more of such words in Swedish than in Chinese or Korean. When looking closer at these items it can be seen that in Swedish, most of the items with a low chance of getting answered correctly are from the later sets. These sets should be harder (Dunn, & Dunn, 2007), thus this result is as expected. The items that are performed worst on by Korean speakers are, however, mostly items from the lower or middle sets. For Chinese, only two such words were found, one from the middle of the test and one of the last testing items.

Remarkable here is the fact that in Chinese and Korean only a few items have a low probability correct chance and almost none of them are from the last few sets. Unfortunately, there is no adult data available on the early items, so it is not known if the adults would have

problems with those words as well. The words performed worse on in the earlier sets are ‘buckle’ and ‘wrench’ in Korean and ‘snarling’ in Chinese. The latter word is debatable however. This was an item from the middle of the test, number 100, and is harder to picture than some other words and might not have anything to do with cultural biasing. The Chinese translators also marked this word as difficult to translate because it has both a formal and an informal version in Chinese. One of the translators suggested the low probability correct score for this item could be due to low frequency of the translated word in spoken language. Both words from Korean are well to imagine, however, and should thus be not that hard to be answered correctly, unless these concepts and/or their words are divergent from the Korean concepts and/or words. The Korean translator indicated these words as loanwords, and the Korean translations for those words are low in frequency. So the concepts are not divergent between cultures, but the words are not comparable in frequency and therefore not in difficulty. This explains the unexpected difficulty with these items. The Korean translator also indicated adults would probably have problems with these words as well, but there is no adult data for these items so this cannot be checked. She suggested these items would be moved to later sets, because they are too hard in comparison to items in the surrounding sets. In Chinese, there seems to be less of a problem since the words with a low chance of being answered correctly are more from later sets and are thus supposed to be harder to answer, especially for younger children.

But this immediately leads to the second remarkable result: Both in Chinese and Korean the words that are supposed to be the hardest were not found to be the hardest ones by the children. This is an indication that the translation of the test might have made these words easier and therefore not comparable to the level of difficulty the items were supposed to have in the original English test or the Swedish one for that matter. Since in the Swedish version the last items were indeed performed worst on, this indicates that in the Swedish translation the original level of difficulty was better copied. It is possible though, that it is simply not possible to correctly translate the words to Chinese and Korean and maintain level of difficulty at the same time, while the Swedish language seems better suited to do this. However, when looking at the Chinese and Korean adult data, it can be seen that the later items are found relatively more difficult by the adults. So information loss of level of difficulty due to translation does not seem plausible anymore. Also, the Korean translator indicated that when translating loanwords, difficulty increases a lot due to low frequency. Level of difficulty is thus not lost due to translation, at least not for this language.

Another explanation for the fact that it has not been found that children deem the last few sets as hard, is that the group of children answering these later items was too small to find any significant low probability correct scores. Perhaps similar results for adults and children would have been found if the group of children answering these items was larger. However, the group of adults was small too. Therefore few items were statistically found to be hard, only some items that were approaching a low probability correct score have been found (see Table 4). To be better able to judge adult performance on these later items, bigger participant groups are needed.

The next issue to address is about the cultural biasing, which was already briefly mentioned above. With the translation of the PPVT-4, cultural biases might have gotten bigger. The development across languages is not the same across ages, which it should have been if the test was unbiased and if one expects that vocabularies develop similarly across cultures, which has by found my for example Bleses and colleagues 2008. In this case, almost all Chinese children have a better score than Korean and Swedish children, and Korean children perform better than Swedish children, albeit not significantly. As Stockman (2000) mentioned, a test should give equal scores to all groups, which in this study did not turn out to be the case. Group differences were found to be large. Furthermore, at the item level there are probably biases as well. It is remarkable that especially the Korean-speaking children have difficulties with items which are supposed to be relatively easy. But this could be explained by the fact that these specific words were loanwords in English and, as the translator indicated, translations for these words have a very low word frequency and are therefore harder. In Chinese this problem seems less evident, since the items marked as difficult to translate were not the items the children performed worst on. The differences between languages could also be ascribed to actual differences in vocabulary development, but since the difference between the Chinese-speaking and the other children is so large, it is not very likely that this is only due to actual vocabulary differences. However, cross-linguistic differences have been reported before (Eriksson & Berglund, 1999). So this explanation cannot be ruled out completely.

What can be concluded is that the development of receptive vocabulary as measured by the PPVT-4 is not the same for the Chinese-, Korean- and Swedish-speaking children participating in this study. It could be the case that the Chinese-speaking children actually have better receptive vocabularies or that the translated versions of the PPVT-4 are not unbiased. If the latter was the case, this could be due to the fact that when translating words

from English to Chinese or Korean, level of difficulty was harder to maintain than when translating English to Swedish. This favors the Chinese- and Korean-speaking children over the Swedish-speaking ones. Especially in the later sets this seems evident, since the Chinese- and Korean-speaking children do not perform worse on the later sets than on earlier sets. This would have been expected since level of difficulty gets harder toward the end of the test (Dunn, & Dunn, 2007). The adults do, however, perform worse on these last few sets, indicating these sets actually do contain harder items.

5.3. **Implications for future research**

For future research, there are several issues that must be taken into account. The most important aspect to look at is the translation of the original test. As can be seen in this study, in translation, information like the level of difficulty can get lost. Especially the Chinese and Korean version should be given a second look, since from the child data it did not seem that the level of difficulty in the last few sets was translated from the original version to the translated ones. However, the adult data did show an increase in difficulty towards the later sets, but the group size was too small to show actual significant effects. It is also possible that the group of children that last few sets were administered to was too small as well to find any significant differences. If it turns out that it is not possible to translate the words and maintain the level of difficulty at the same time, an alternative solution could be to take one of the distracter pictures and use that word as testing item, as has been done in the Dutch translation of the PPVT-III (Dunn & Dunn, 2005). This should be done with much consideration though, since these items are not tested and approved by standards as used by Dunn and Dunn (2007). This would also only be an option if the translation of those words would be comparable in difficulty to the original English target word. If this is not the case, selecting another picture would not solve the problem. It might then be considered to leave out the item entirely, but this endangers the validity of the entire test and therefore would not be recommended. Another possibility was suggested by the Korean translator: To move up difficult items to later sets. In each case, translation, substitution, displacement or omission should be done very carefully to not undermine the validity of the test, as Peña also suggests (2007). She did not evaluate the PPVT in specific, but gave some general remarks about translating linguistic tests. Translations should be linguistically, functionally, culturally and metrically equivalent to the original test items. As guidelines for maintaining level of difficulty, Peña (2007) suggests to look at proportion or probability correct scores for each item and to examine word

frequency and age of acquisition. If those match the features in the original language, the translation seems to be done right. In this study, probability correct scores have been established and the Korean translation is based on word frequency. Further research should look at the other factors if one wants to improve the translations.

Another point to mention is the categorization of word types. The manual of the PPVT-4 does not specify the word type of each item, which is why a self-made categorization has been proposed. This categorization does have its limitations, however. First of all, several items have been indicated as having the property as possibly being both a verb and an adjective. For example the word ‘floating’ which could be a verb in the sense of ‘the boat is floating’ or an adjective as in ‘the floating boat’. In agreement with the second reviewer of the categorization, in this case the word was categorized as a verb. Similar items are ‘flaming’ and ‘ascending’. For the latter two it was chosen to identify them as an adjective, because this seemed to better match the accompanying picture. The complete list of words and the category they were assigned to can be found in Appendix B. When continuing this type of analysis, it is recommended to reexamine this categorization. Secondly, since only two people have taken a look at the categorization so far. Preferably, native speakers of English should check and possibly adjust the current categorization.

Furthermore, in future research it is desirable to be able to use all data gathered. In this study almost a third of the data collected (31.7%) has not been taken into account in the final analysis because of wrong administration of those items. This is unnecessary waste of data and should be prevented in future research. However, in the Chinese testing phase, it was deliberately chosen to administer all test items to be able to evaluate test items. This data has been used for such analyses and therefore was not wasted, but for all other cases the wrongly administered data has not been used for the final analyses. This was done especially since for the Swedish-speaking children not all basal sets have been established right. Because it is not known how many items should have been administered but were not, this data was not used.

To be able to use the PPVT-4 for Chinese-, Korean- and Swedish-speaking children, it is necessary to take research on this topic to the next level and take the points mentioned above into consideration. If the PPVT-4 is going to be translated to other languages, all these points should be taken into account as well. Also, it is recommended to test more adults, especially if one wants to study potential biases at item level. It would be best to administer all test items with the adults, so that also possible confusing items in the early and middle sets can be established. If there are items which even the adults perform worse on, it can be

expected that those items are (culturally or linguistically) biased. If this is not the case and the adults do perform well on those items, it might be true that those items are just not prominent in child vocabularies but are known in the culture or country. This is important to know if one wants to further develop the translated versions. Few adult data was available for this study and data from Swedish-speaking adults was not available at all so this point should definitely be put into practice.

References

Bates, E., Beeghly, M., Bretherton, I., McNew, S., O'Connell, B., Reznick, J. S., et al. (1984). *Early Language Inventory.* Unpublished manuscript, Center for Research on Language, University of California, San Diego.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., …Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language, 21,* 85-123. doi:10.1017/S0305000900008680

Berglund, E., & Eriksson, M. (2000). Communicative development in Swedish children 16 28 months old: The Swedish early communicative development inventory – words and sentences. *Scandinavian Journal of Psychology, 41,* 133-144. doi:10.1111/1467-9450.00181

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomson, P., Madsen, T., & Basbøll, H. (2008). Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language, 35,* 619-650. doi:10.1017/S0305000908008714

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S. Y., Pascual, L., … Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development, 75,* 1115-1139. doi:10.1111/j.1467-8624.2004.00729.x

Bornstein, M. H., & Hendricks, C. (2012). Basic language comprehension and production in >100,000 young children from sixteen developing nations. *Journal of Child Language, 39,* 899-918. Doi: 10.1017/S0305000911000407

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl. L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development, 10*, 159-199. doi:10.1016/0885-2014(95)90008-X

Rescproa. L., Lee, Y. M. C., Oh, K. J., & Kim, Y. A. (2013). Lexical development in Korean: Vocabulary size, lexical composition, and late talking. *Journal of Speech Language and Hearing Research, 56,* 735-747. doi:10.1044/1092-4388(2012/11-0329

Dunn, L. M., & Dunn, L. M. (2005). *Peabody Picture Vocabulary Test-III-NL, Nederlandse versie door Liesbeth Schlichting.* Amsterdam: Harcourt Assessment B.V.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test.* Minneapolis: Pearson Assessments

Gathercole, S. E., Willis, C. S., Emslie, J., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28,* 887-898. oi:10.1037/0012-1649.28.5.887

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language, 27,* 689-705. doi:10.1017/S0305000900004414

Haitana, T., Pitama, S., & Rucklidge, J. J. (2010). Cultural biases in the Peabody Picture Vocabulary Test-III: Testing Tamariki in a New Zealand sample. *New Zealand Journal of Psychology, 39,* 24-34. Retrieved from: http://www.psychology.org.nz/NZ_Journal

Hao, M., Shu, H., Xing, A., & Li, P. (2008). Early vocabulary inventory for Mandarin Chinese. *Behavior Research Methods, 40,* 728-733. doi:10.3758/BRM.40.3.728

Heilman, J., Weismer, S. E., Evans, J., & Hollar, C. (2005). Utility of the MacArthur–Bates Communicative Development Inventory in identifying language abilities of late talking and typically developing toddlers. *American Journal of Speech-Language Patholog, 14,* 40-51. doi:10.1044/1058-0360

Kim, M., McGregor, K. K., & Thompson, C. K. (2000). Early lexical development in English- and Korean speaking children: Language-general and language-specific patterns. *Journal of Child Language, 27,* 225-254. Retrieved from: http://journals.cambridge.org/abstract_S0305000900004104

Li, P., Bates, E., & MacWhinney, B. (1993). Processing a language without inflections: A reaction time study of sentence interpretation in Chinese. *Journal of Memory and Language, 32,* 169-192. doi:10.1006/jmla.1993.1010

Li, C. N., & Thompson, S. A. (1974). An explanation of word order change SVO → SVO. *Foundations of Language, 12,* 201-214. From: http://www.jstor.org/stable/25000832

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11,* 9-16. doi:10.1111/j.1467-7687.2008.00671.x

Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development, 78,* 1255-1264. doi:0009-3920/2007/7804-0014

Pine, J. M., Lieven, E. V. M., & Rowland, C. (1996). Observational and checklist measures of vocabulary composition: What do they mean? *Journal of Child Language, 23,* 573-590. doi:10.1017/S0305000900008953

Platt, A. E. (2010). *Examining the Appropriateness of the PPVT-III as a Measure of Vocabulary Development in Linguistically Diverse Kindergarten Children* (Doctoral dissertation, University of Toronto)

Qi, C. H., Kaiser, A. P., Milan, S., & Hancock, T. (2006). Language performance of low income African American and European American preschool children on the PPVT-III. *Language, Speech, and Hearing Services in Schools, 37,* 5-16. doi:10.1044/0161-1461(2006/002)

Quené, H., & Van den Bergh, H. (2008). Examples of mixed effects modelling with crossed random effects and with binomial data. *Journal of Memory and Language, 24,* 413 425. doi:10.1016/j.jml.2008.02.002

Restrepo, M. A., Schwanenflugel, P. J., Blake, J., Neuharth-Pritchett, S. Cramer, S. E., &

Ruston, H. P. (2006). Performance on the PPVT-III and the EVT: Applicability of the measures with African American and European American preschool children. *Language, Speech, and Hearing Services in Schools, 37,* 17-27. doi:0.1044/0161-1461(2006/003)

Rock, D. A., & Stenner, A. J. (2005). Assessment issues in the testing of children at school entry. *The Future of Children, 15,* 15-34. doi:10.1353/foc.2005.0009

Stockman, I. J. (2000). The new Peabody Picture Vocabulary Test-III: An illusion of unbiased assessment? *Language, Speech and Hearing Services in Schools, 31,* 340-353. doi:10.1044/0161-1461.3104.340

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology, 32,* 492-504. doi:10.1037/0012-1649.32.3.492

Tardif, T., Gelman, S. A, & Xu, F. (1999). Putting the "noun bias" in context: A comparison of English and Mandarin. *Child Development, 70,* 620-635. doi:10.1111/1467-8624.00045

Washington, J. A., & Craig, H. K. (1999). Performances of at-risk, African American preschoolers on the Peabody Picture Vocabulary Test-III. *Language Speech and Hearing Services in Schools, 30,* 75-82. doi:10.1044/0161-1461.3001.75

Wassenberg, R., Hurks, P. P. M., Hendrisken, J. G. M., Feron, F. J. M., Meijs, C. J. C., Vles,

J. S. H., & Jolles, J. (2008). Age-related improvement in complex language comprehension: Results of a cross-sectional study with 361 children aged 5 to 15. *Journal of Clinical and Experimental Neuropsychology,* 30, 435-448. doi:10.1080/13803390701523091.

Appendix A

Table A1.

List of items the Chinese translators had difficulties with and their alternatives to the chosen translation.

Item nr.	Original word	Chosen translation	Alternative 1	Alternative 2	Alternative 3	Alternative 4
12	mouth	嘴巴	嘴			
30	fence	栅栏	围栏	篱笆		
48	roof	房顶	屋顶			
56	buckle	扣子	带扣			
58	panda	熊猫	大熊猫			
72	furry	毛的	毛毛的	绒的	绒毛的	毛绒的
75	globe	地球	地球仪			
126	snarling	咆哮	嚎叫	狗叫		
144	dilapidated	破败	残破			

Appendix B

Table B1.

List of test items of the PPVT-4 and the word type categorization made.

Item number	Word	Word type
1.	Ball	Noun
2.	Dog	Noun
3.	Spoon	Noun
4.	Foot	Noun
5.	Duck	Noun
6.	Banana	Noun
7.	Shoe	Noun
8.	Cup	Noun
9.	Eating	Verb
10.	Bus	Noun
11.	Flower	Noun
12.	Mouth	Noun
13.	Pencil	Noun
14.	Cookie	Noun
15.	Drum	Noun
16.	Turtle	Noun
17.	Red	Adjective
18.	Jumping	Verb
19.	Carrot	Noun
20.	Reading	Verb
21.	Toe	Noun
22.	Belt	Noun
23.	Fly	Noun
24.	Painting	Verb
25.	Dancing	Verb
26.	Whistle	Noun
27.	Kicking	Verb
28.	Lamp	Noun

29.	Square	Noun
30.	Fence	Noun
31.	Empty	Adjective
32.	Happy	Adjective
33.	Fire	Noun
34.	Castle	Noun
35.	Squirrel	Noun
36.	Throwing	Verb
37.	Farm	Noun
38.	Penguin	Noun
39.	Gift	Noun
40.	Feather	Noun
41.	Cobweb	Noun
42.	Elbow	Noun
43.	Juggling	Verb
44.	Fountain	Noun
45.	Net	Noun
46.	Shoulder	Noun
47.	Dressing	Verb
48.	Roof	Noun
49.	Peeking	Verb
50.	Ruler	Noun
51.	Tunnel	Noun
52.	Brand	Noun
53.	Envelope	Noun
54.	Diamond	Noun
55.	Calender	Noun
56.	Buckle	Noun
57.	Sawing	Verb
58.	Panda	Noun
59.	Vest	Noun
60.	Arrow	Noun
61.	Picking	Verb

62.	Target	Noun
63.	Dripping	Verb
64.	Knight	Noun
65.	Delivering	Verb
66.	Cactus	Noun
67.	Dentis	Noun
68.	Floating	Verb
69.	Claw	Noun
70.	Uniform	Noun
71.	Gigantic	Adjective
72.	Furry	Adjective
73.	Violin	Noun
74.	Group	Noun
75.	Globe	Noun
76.	Vehicle	Noun
77.	Chef	Noun
78.	Squash	Noun
79.	Ax	Noun
80.	Flamingo	Noun
81.	Chimney	Noun
82.	Sorting	Noun
83.	Waist	Noun
84.	Vegetable	Noun
85.	Hyena	Noun
86.	Plumber	Noun
87.	River	Noun
88.	Timer	Noun
89.	Catching	Verb
90.	Trunk	Noun
91.	Vase	Noun
92.	Harp	Noun
93.	Bloom	Noun
94.	Horrified	Adjective

95.	Swamp	Noun
96.	Heart	Noun
97.	Pigeon	Noun
98.	Ankle	Noun
99.	Flaming	Adjective
100.	Wrench	Noun
101.	Aquarium	Noun
102.	Refueling	Verb
103.	Safe	Noun
104.	Boulder	Noun
105.	Reptile	Noun
106.	Canoe	Noun
107.	Athlete	Noun
108.	Towing	Verb
109.	Luggage	Noun
110.	Directing	Verb
111.	Vine	Noun
112.	Digital	Adjective
113.	Dissecting	Verb
114.	Predatory	Adjective
115.	Hydrant	Noun
116.	Surprised	Adjective
117.	Palm	Noun
118.	Clarinet	Noun
119.	Valley	Noun
120.	Kiwi	Noun
121.	Interviewing	Verb
122.	Pastry	Noun
123.	Assisting	Verb
124.	Fragile	Adjective
125.	Solo	Noun
126.	Snarling	Verb
127.	Puzzled	Adjective

128.	Beverage	Noun
129.	Inflated	Adjective
130.	Tusk	Noun
131.	Trumpet	Noun
132.	Rodent	Noun
133.	Inhaling	Verb
134.	Links	Noun
135.	Polluting	Verb
136.	Archaeologist	Noun
137.	Coast	Noun
138.	Injecting	Verb
139.	Fern	Noun
140.	Mammal	Noun
141.	Demolishing	Verb
142.	Isolation	Noun
143.	Clamp	Noun
144.	Dilapidated	Adjective
145.	Pedestrian	Noun
146.	Interior	Noun
147.	Garment	Noun
148.	Departting	Verb
149.	Feline	Noun
150.	Hedge	Noun
151.	Citrus	Noun
152.	Florist	Noun
153.	Hovering	Verb
154.	Aquatic	Adjective
155.	Reprimanding	Verb
156.	Carpenter	Noun
157.	Primate	Noun
158.	Glider	Noun
159.	Weary	Adjective
160.	Hatchet	Noun

161.	Transparent	Adjective
162.	Sedan	Noun
163.	Constrrained	Adjective
164.	Valve	Noun
165.	Parallelogram	Noun
166.	Pillar	Noun
167.	Consuming	Verb
168.	Currency	Noun
169.	Hazardous	Adjective
170.	Pentagon	Noun
171.	Appliance	Noun
172.	Poultry	Noun
173.	Cornea	Noun
174.	Pensinsula	Noun
175.	Porcelain	Noun
176.	Detonation	Noun
177.	Cerebral	Adjective
178.	Perpendicular	Adjective
179.	Submerging	Verb
180.	Syringe	Noun
181.	Lever	Noun
182.	Apparel	Noun
183.	Talon	Noun
184.	Cultivating	Verb
185.	Wedge	Noun
186.	Ascending	Adjective
187.	Depleted	Adjective
188.	Sternum	Noun
189.	Maritime	Adjective
190.	Incarcerating	Verb
191.	Dejected	Adjective
192.	Quintet	Noun
193.	Incandesecent	Adjective

194.	Confiding	Adjective
195.	Mercantile	Adjective
196.	Upholstery	Noun
197.	Filtration	Noun
198.	Replesnishing	Verb
199.	Trajectory	Noun
200.	Perusing	Verb
201.	Barb	Noun
202.	Convergeging	Verb
203.	Honing	Verb
204.	Angler	Noun
205.	Wildebeest	Noun
206.	Confierous	Adjective
207.	Timpani	Noun
208.	Pilfering	Verb
209.	Pestle	Noun
210.	Reposing	Verb
211.	Cupola	Noun
212.	Derrick	Noun
213.	Convex	Adjective
214.	Embossed	Adjective
215.	Torrent	Noun
216.	Dromedary	Noun
217.	Legume	Noun
218.	Carin	Noun
219.	Arable	Adjective
220.	Supine	Adjective
221.	Vitreous	Adjective
222.	Lugubrious	Adjective
223.	Caster	Noun
224.	Terspichorean	Adjective
225.	Cenotapgh	Noun
226.	Calyx	Noun

227.	Osculating	Verb
228.	Tonsorial	Adjective