The Relation between MLU and Noun Bias in Mandarin

Pei-Chieh, Kan and Yi-Yun, Huang

Abstract

It is believed that children have noun bias as their lexicon is increasing early on. While some see the predominance of nouns as a universal property, others argue that there are verb-friendly languages such as Japanese and Korean that do not exactly exhibit the trend. Cross-linguistic research has been set to investigate whether the phenomenon is a language-specific characteristic. In this article, we aim to look at noun bias in Mandarin in relation with MLU. Measuring the lexical diversity of nouns and verbs, noun bias is observed that Mandarin-speaking children do exhibit noun bias.

Keywords: MLU, noun bias, verbs, language development, Mandarin, lexical diversity

1. Introduction

1.1. Universal noun bias

In early lexical development, children appear to acquire nouns before verbs. It has been shown that even with larger inputs of verbs, the amount of nouns acquired exceeds that of verbs (Taverna & Waxman, 2020). In general, learners have a preference of nouns over verbs (Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. 1999). However, noun bias has been pointed out to be weaker for languages like Mandarin (Waxman et al, 2013). The potential factor would be that infants need to develop the knowledge of the nouns before they are able to make good use of verbs (Waxman et al, 2013).

1.2. Cross-linguistic variation

Earlier research revealed that languages with SOV structures are verb-friendly. For example, children who are native Korean or Japanese speakers have a higher tendency to acquire verbs from their input. While those of SVO structures are noun-friendly. For example, French, English, and Italian in which children have a high preference for acquiring nouns (Waxman et al, 2013).

This finding implies that when acquiring specific languages, children have tendencies to extract different knowledge from given input, displaying how linguistic variations in language input influences children's acquisition of verbs and nouns.

1.3. Noun bias in Mandarin

However, there is an exception in children's acquisitions. Take Mandarin for example, Mandarin is a language whose syntax structure is SVO. But the results turned out to be that these children had higher preference for acquiring verbs from their input. The possible reason is that the objects are often deleted in Mandarin conversation, resulting in SV structures. Thus, in order to testify whether noun bias appears in Mandarin-speaking children, we used the data from these children, and analysed their MLU and the ratio of verbs and nouns in their utterance.

1.4. Lexical diversity of nouns and verbs

The usage of nouns and verbs can be defined through usage percentage, type token ratio (TTR), and other calculations. However, neither is an ideal measure for the scope of this article.

Usage percentages do not take into account the repetition of identical words. As for TTR, based on previous research (Malvern, David D et al, 2004), it is not an accurate measurement of lexical diversity since it is greatly influenced by usage frequency. Instead, we used the D measure (Malvern, D., Richards, B., Chipere, N., Durán, P., 2004) to calculate the lexical diversity of verbs and nouns, and to examine the correlation between MLU and the D value.

2. Method

2.1. Method-corpus

The corpus used is the TCCM (Taiwan Corpus of Child Mandarin) from CHILDES, containing longitudinal studies of ten participants.

2.2. Participants

Table 1: *Speaker names and age intervals.*

speaker	age
xu	106 - 205
yang	105 - 209
wu	107 - 210
pan	109 - 309
chou	201 - 304
jc	202 - 305
wang	205 - 304
wuys	207 - 310
cheng	301 - 311
chw	306 - 403

2.3. Materials and procedure

The materials are from the CHILDES TCCM corpus, the longitudinal study contains a transcription file for every child's different years of age as listed. In total, the 127 transcription files were extracted and run through CLAN.

The FREQ command was used to count the total number words and those of the two categories to calculate usage percentages. The mean length of utterance for each file was obtained with the MLU function. The D value for lexical diversity was calculated using the VOCD function. For verb and noun-specific calculations, commands +sm|v and +sm|n were added in the command line. In terms of statistical analysis, JASP was used to calculate the P values for MLU versus D-v and D-n.

2.4. Equations

Two equations are used to process data. The first is the calculation of the type token ratio of verbs and nouns in terms of D, a measurement of lexical diversity. Whereas the second, derived from the first one, shows the exact calculation used for the D value of verbs and nouns.

2.4.1. Equation 1, type token ratio

$$TTR = \frac{D}{N}\left[\left(1 + 2\frac{N}{D}\right)^{\frac{1}{2}} - 1\right]$$

2.4.2. Equation 2, D value

$$D = \frac{TTR^2 \times N}{2 \times (1 - TTR)}$$

In both equations (Malvern, D., Richards, B., Chipere, N., Durán, P., 2004), TTR is the type token ratio, D is the D measure of lexical diversity, and N denotes total tokens.

3. Results

3.1. The D measure of verbs and nouns

The D value is higher for nouns, indicating an overall higher measurement of lexical diversity across all participants.

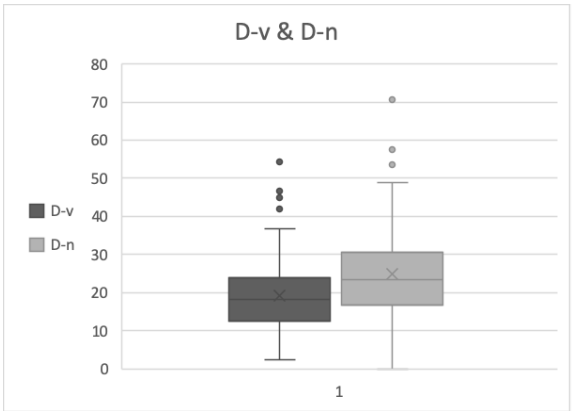

Figure 2: *D value of verbs and nouns*

3.2. Usage percentages of verbs and nouns

The usage percentage of verbs is slightly higher, with verbs and nouns sharing a similar range.

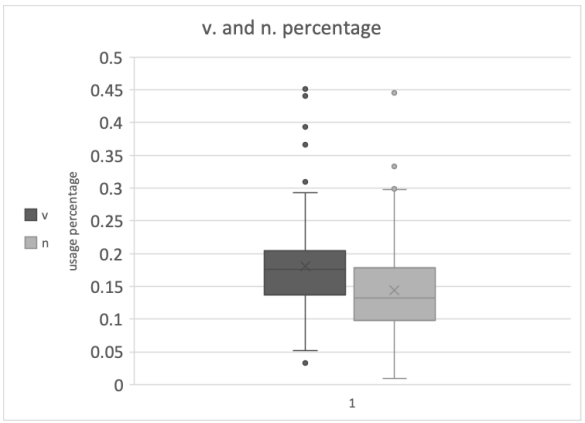

Figure 1: *Verb and noun percentages*

3.3. The Correlations between MLU and D value of verbs

The scatter plot indicates a positive correlation between MLU and the D value of verbs.

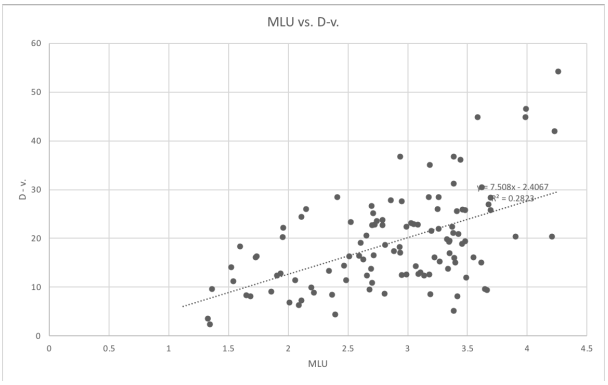

Figure 3: *Scatter plot of MLU and the D value of verbs.*

3.4. The Correlations between MLU and D value of nouns

The scatter plot indicates a positive correlation between MLU and the D value of nouns.

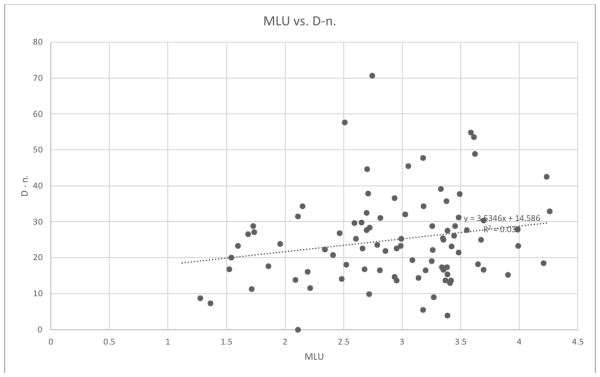

Figure 4: *Scatter plot MLU and the D value of nouns.*

3.5. The Correlations between MLU and the D value of verbs for longer utterances

The scatter plot shows us that the correlation between longer MLU and the D value of verbs is much higher than the overall one.

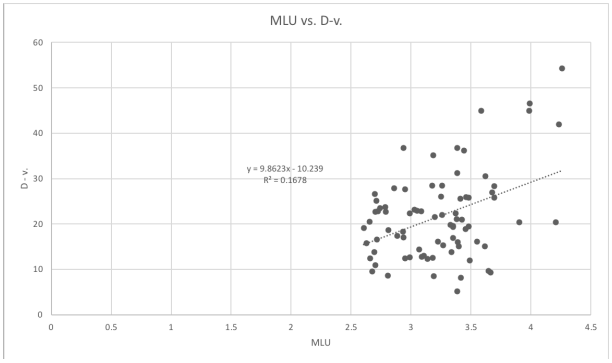

Figure 5: *The correlation between MLU and D value of verbs when MLU ≥ 2.6*

3.6. The Correlations between MLU and D value of nouns for longer utterances

The scatter plot shows us the correlation between longer MLU and the D value of nouns is much lower than the overall one, approaching 0.

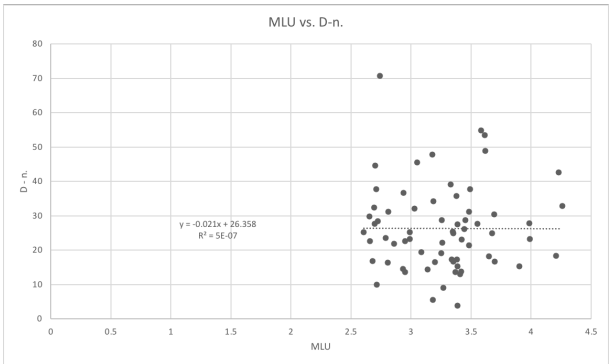

Figure 6: *The correlation between MLU and D value of verbs when MLU ≥ 2.6*

4. Discussion

4.1. The overall correlation between MLU and D-v, D-n

The two complete scatter plots both increase in diversity as utterances increase in length. However, the positive relation is not as significant for nouns, since the diversity of nouns start out higher than verbs, implying that nouns are learnt and therefore used more in early lexical development.

4.2. Focusing on longer MLU

The trend of a more diverse base of nouns early on is even more obvious when the longer half of MLU data points are extracted and plotted against D-v and D-n. As children enter later stages of syntactic and morphological development, nouns are already steadily high in diversity, while the amount of verbs is still growing to reach the same level of lexical diversity.

In other words, since verbs are better learned after nouns, when the rate of noun acquisition slows down, the relative rate of verb acquisition is high, as observed in both sets of graphs and especially with higher MLU values.

4.3. The D measure and usage percentages

The box plot of D values again verifies the overall higher level of diversity in nouns. Usage percentages not only play the role of verifying that higher lexical diversity does not result from a mere higher use of nouns, it again shows the need to examine the usage of nouns not just based on frequency. Less conclusive implications would have been made if results were solely based on usage percentages.

5. Conclusion

In conclusion, in terms of lexical diversity, nouns have an overall higher D value than verbs. Looking at usage percentages, the factor of the difference in lexical diversity resulting from a mere larger portion of use is ruled out. The lower correlation between MLU and nouns is the children's knowledge of nouns reaching a sufficient amount earlier on, leading children to acquire more verbs than nouns in the later stages of syntactic and morphological development. The early trend of high lexical diversity in nouns verifies the presence of noun bias in Mandarin.

6. References

[1] Taverna, A., & Waxman, S. (2020). Early lexical acquisition in the Wichi language. Journal of Child Language, 47(5), 1052-1072.

[2] Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. Cognition, 73(2), 135–176. https://doi.org/10.1016/S0010-0277(99)00036-0

[3] Waxman, Sandra et al. "Are Nouns Learned Before Verbs? Infants Provide Insight Into a Long-Standing Debate." Child development perspectives 7.3 (2013): 155–159. Web.

[4] Malvern, David D et al. Lexical Diversity and Language Development: Quantification and Assessment. London: Palgrave Macmillan Limited, 2004. Web

[5] Malvern, D., Richards, B., Chipere, N., Durán, P. (2004). A Mathematical Model of Lexical Diversity. In: Lexical Diversity and Language Development. Palgrave Macmillan, London. https://doi.org/10.1057/9780230511804_3

[6] CHILDES. childes.talkbank.org.