Quantitative corpus analysis on the social and developmental trends in young children's spontaneous polite language usage in Japanese

Franklin Chang (Kobe City University of Foreign Studies)

Tomoko Tatsumi（Kobe University）

1. Introduction

Almost every Japanese verb is marked for politeness in ways that match the addressee. For example, the verb 'eat' can be either in polite form *tabemasu* or plain form *taberu*. Japanese speakers need to choose and use the appropriate expressions of politeness according to their addressee and the social context. However how young children learn polite language is intriguing because parents and children mainly use plain language with each other. To better understand this development, we provide a broad quantitative characterization across a relatively large corpus of how this language develops in Japanese children as well as the parental input.

2. Method

The current study looks at how plain and polite linguistic forms are used by both parents and children during the early stages of language acquisition in order to understand how children's polite language is guided by their input. Our quantitative analysis was performed on eight recorded Japanese corpora in CHILDES database (MacWhinney, 2014; Hamasaki, 2002; Ishii, 1999; Miyata, 2000, 2012; Okayama, 1973). In order to look at whether linguistic forms in the input are distinguished by addressee and whether children are sensitive to these distinctions in the input, we encoded the speaker of the subsequent turn as a proxy for intended addressee of each given turn using four addressee categories: Target Child, Other Children, Parents, Other Adults. For example, if the mother produces an utterance followed by an utterance produced by the child, we treat the child as the intended addressee of the mother's utterance. We used the addressee and children's age as a variable for analyzing children's and adults' use of plain and polite language forms. Utterances were coded in terms of plain and polite (i.e., whether they have politeness marker for verbs (*-des*, *-mas*) and parent name nouns (polite address *-san*, and plain address *–chan* as in *okaasan* and *okaachan*). By using this automatic method to encode addressee, we can do a quantitative analysis of the 807,272 parent and child utterances in these corpora.

3. Results

Linear regression was used to fit the proportion of polite forms (forms with *masu, masen, mashita, desu, deshita* or *kudasai*) with centered age and helmert-coded addressee crossed. Overall, only 4% of the parent's utterances were in teineigo form and this is consistent with the social norm that parents and children talk to each other with plain speech. The parents used teineigo more with other adults compared to each other, $z=-3.1$, $p=0.002$, and this is consistent with the idea that teineigo should be used towards adult strangers. The parents also used teineigo more with adults than children, $z=-7.8$, $p<0.001$, which is consistent with the norm that plain speech is used with children, regardless of their closeness to the speaker. The parents did not use teineigo with the target child initially

compared to the other addressees, but they increased the use of teineigo with this child over development, $z=3.7$, $p<0.001$.

Regarding children's utterances in the corpus, teineigo forms were produced more as the child grew older, $z=3.1$, $p=0.0021$ (age correlation=0.19). Teineigo forms were used more with the Parents than with Other Adults, $z=4.9$, $p<0.001$, and this did not change with age. There was no difference in the use of teineigo between adults and Other Children and this did not change with age.

Figure 1. The proportion of the polite form utterances from Parent and Target Child speakers to different addressees.
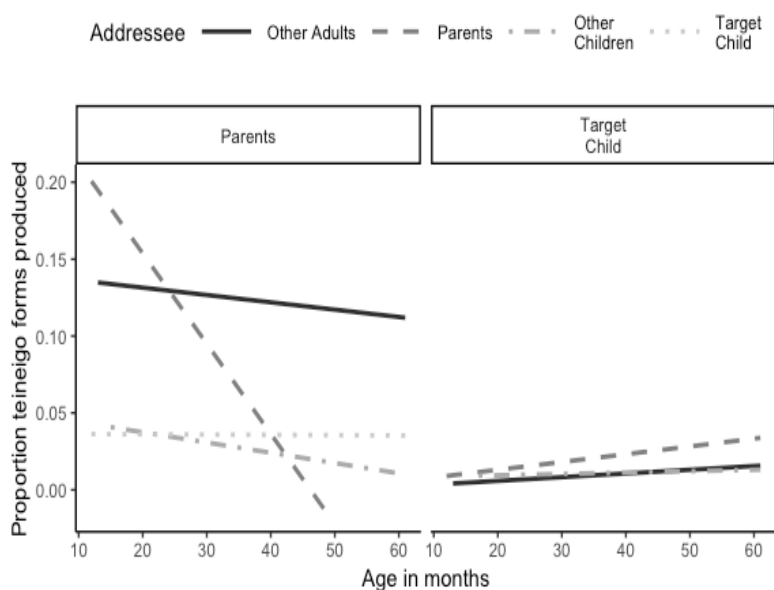
The second analysis looks at nouns, specifically the formal title -san and informal titles -chan and -kun. We first looked at how these titles were used with parent names of the mother (*okaa-san*, *kaa-san*, *mama-san*, *okaa-chan*, *kaa-chan*, *mama-chan*) and the father (*otoo-san*, *too-san*, *papa-san*, *otoo-chan*, *too-chan*, *papa-chan*). In contrast with verb forms where plain forms dominated the input, parents preferred to use polite *san* forms rather than *chan* for parent names (83% were *san* parental names).　Overall, the parental use of *san* forms did not increase with age ($p=0.405$). Utterances to Other Adults had more parental *san* forms than those to Other Children, $t=-2.5$, $p=0.011$, but this difference reduced with the age of the child, $t=2.7$, $p=0.0072$. *San* forms became fewer as the Target child got older, $t=-3.3$, $p<0.001$.

For the Target Child utterances (right panel of Figure 2), they preferred to use *san* forms for parent names in general (60% were *san* forms) and used *san* parent names more with other children as they got older, $t=-3.3$, $p<0.001$.

Figure 2. The proportion of the -san parent names from Parent and Target Child speakers to different addressees.
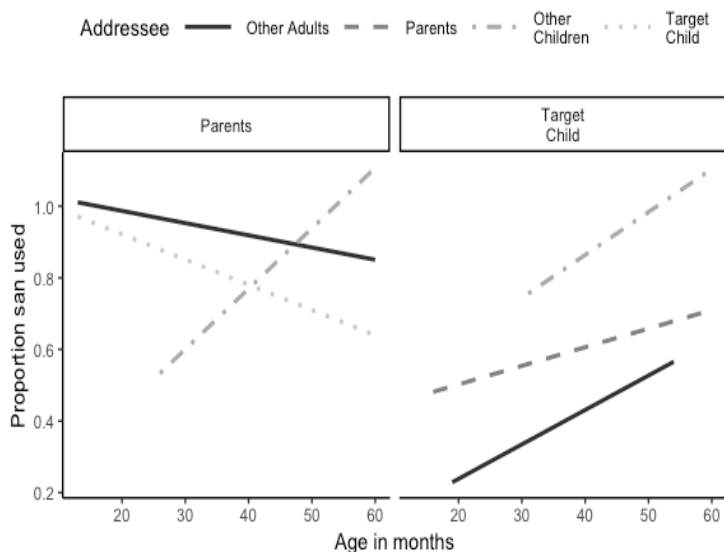
Japanese often use title for a wide range of non-parental words such as animals or vegetables (raionsan 'Mr Lion', ninjinsan 'Mr Carrot') and children/pets (nekochan 'Mr Cat'). While parent names are almost exclusively used with *san*, non-parent-name words are biased for *chan* (72% were *chan* forms). The proportion of *san* forms decreased over age, t=-4, p<0.001. Utterances to Other Children had more *san* form than those to Other Adults, t=2, p=0.047. Utterances to the Target Child contained more *san* forms when compared against the other addressees, t=3.2, p=0.0015, but this reduced over development, t=-3.2, p=0.0016.

Figure 3. The proportion of the -san nouns from Parent and Target Child speakers to different addressees.
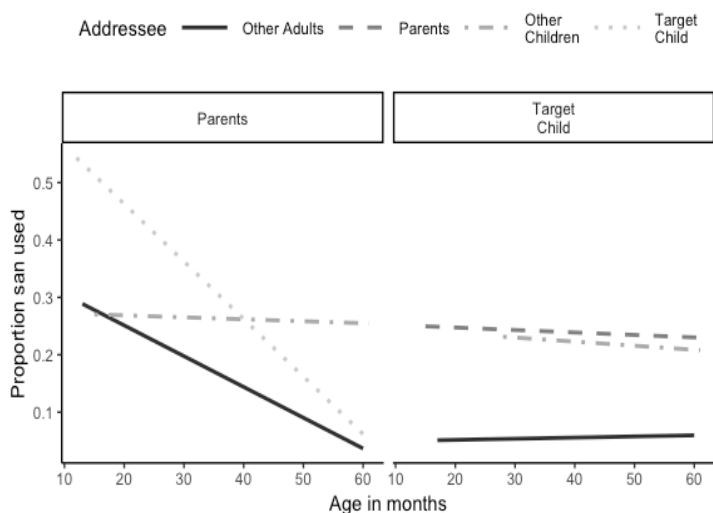
4. Discussion

Previous work has suggested that Japanese children have acquired distinctions related to polite language early in development (Nakamura, 2001). But since there is a social norm for parents to use plain speech with their children,

it was not clear how children could acquire this system so early in development. Here, we found that the parental input provided a substantial volume of polite speech spread across a range of verbs and nouns, although this was only a small percentage of the total input. Secondly, we used the next speaker as a probabilistic measure of the intended addressee and we found that parental speech distinguished their use of polite language for adults compared to children (teineigo verb endings and polite parent names *san* titles). If children are sensitive to this objective feature of the input, then they could use it to learn that polite and plain speech are distinguish by the social context.

References

Hamasaki, N. (2002). The Timing Shift of Two-year-old's Responses to Caretakers' Yes/No Questions. In Y. Shirai, H. Kobayashi, S. Miyata, K. Nakamura, T. Ogura, & H. Sirai (Eds.), Studies in Language Sciences (2) - Papers from the Second Annual Conference of the Japanese Society for Language Sciences (pp. 193–206).

Ishii, T. (1999). The JUN Corpus. Unpublished.

MacWhinney, B. (2014). The CHILDES project: Tools for analyzing talk, Volume II: The database. Psychology Press.

Miyata, S. (2000). The TAI corpus: Longitudinal speech data of a Japanese boy aged 1;5.20-3; 1. Bulletin of Shukutoku Junior College 39, 77–85.

Miyata, S. (2012). Nihongo MLU (heikin hatsuwachō) no gaidorain: Jiritsugo MLU oyobi keitaiso MLU no keisanhō [Guideline for Japanese MLU: How to compute MLUw and MLUm]. Kenkō Iryō Kagaku 2, 1–15. Kenkō Iryō Kagaku 2, 1–15.

Nakamura, K. (2001). Gender and language in Japanese preschool children. Research on Language and Social Interaction, 34(1), 15–43.

Okayama, Y. (1973). Haha to ko no taiwa no shuurooku [A collection of mother-child conversations]. Osaka: Yasuda Seimei Shakaijigyoodan.