Use of Markers Observed in the Spoken Language Lexical Corpora of Children in Kannada Language

B. A. Mahalakshmi Prasad, M.A.
Dr. K. S. Prema, Ph.D.

Abstract

Research in the area of child language thrives on the availability of naturalistic language data. However, Indian endeavours are meagre and confined to the language data acquired by the Speech - Language Pathologists which is limited to the purview of their study. The present study is part of a post-doctoral work carried out at All India Institute of Speech and Hearing (AIISH).

The study established a spoken language lexical corpus of children between the ages of 6 and 8 in Kannada language. The language data of 240 children, living in the city of Mysore was collected through description of standardised pictures, story narration and narration of daily activities of the child. The present paper looks into the use of markers by children while highlighting the need to encourage further research in the field of child language.

Keywords: spoken language corpus, language acquisition, markers,
Indian Languages (CIIL), Mysore, Deccan College, Pune, Indian Institute of Applied Language Science, Bhubaneshwar, Sampurnanada Sanskrit University, Varanasi, and Aligarh Muslim University, Aligarh. However, the project was discontinued and later revived in the 2003. (See Dash, 2005)

The corpora generated in different languages and presently available are listed by the Open Language Archives Community (OLAC), which lists corpora that are available to the public as well as corpora that is specialised with its use restricted to certain agencies. OLAC lists the available corpora such as the corpora to study language development such as CHILDES, ESL/EFL learner corpora such as International Corpus of Learner English (ICLE), etc., monolingual corpora such as Modern Chinese Language Corpus (MCLC) and comparative corpora such as COMPARA and many others.

**Sporadic Child Language Corpuses in Indian Languages**

Child language corpuses that exist are created using reports by parents, audio and video recording of child speech. The major corpus in the area of language acquisition is CHILDES and has been expanded to document the intricate processes of language acquisition by children in different language settings such as Hebrew, Japanese, French, etc.

However, in the Indian scenario endeavours in the area of child language acquisition have been sporadic and is largely restricted to the purview of conducted studies such as a study of the different stages of acquisition or the lacunae that might occur during the process of language acquisition like the Tamil audio recording by R. Narasimhan (Tata Institute of Fundamental Research) and R. Vaidyanathan (Audiology and Speech Therapy School, Nair Hospital, Mumbai) which led to CHILDES database in 1984.

The audio recordings are interactions of a child with her parents in unstructured caretaking situation in her home from the age of 9 months to the age of 33 months. The interactions were recorded over a period of 24 months with biweekly intervals.

Although, various agencies in India such as AIISH have begun to collate and establish a corpus of child language, there still exists a dearth of normative data of language acquisition that can be of use in the area of speech - language pathology.
**Present Project**

The present project is a beginning in the gargantuan endeavour of documenting and describing the language use seen in typically developing children in various contexts such as spontaneous speech settings, picture elicitation, picture description, etc. The project’s primary endeavour is to establish a lexical corpus of spoken language (Kannada) of children between the ages of six to eight years.

**Previous Studies**

Table 1.1 gives a brief overview of the studies in the area of child language acquisition in India.

**Table 1.1**

*Studies in Language Acquisition in Children in India*

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Study Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sailaja (1994)</td>
<td>Investigated the role of syntax in the acquisition of Telugu.</td>
</tr>
<tr>
<td>Shukla &amp; Mohanty (1995)</td>
<td>Studied the influence of maternal speech styles on language acquisition showed a significant correlation between the two.</td>
</tr>
<tr>
<td>Devaki (1995)</td>
<td>Studied the development of past tense in Kannada children</td>
</tr>
<tr>
<td>Khokle (1995)</td>
<td>Studied the acquisition of aspirated /g/ segment in Marathi</td>
</tr>
<tr>
<td>Lakshmi Bai &amp; Vasanta</td>
<td>Contributed extensively to the study of language acquisition in children by studying the acquisition of different classed phonemes of Tamil and Telugu languages in different positions. Her observation of language acquisition in children supported the view that “the development of phonology cannot be studied meaningfully without considering the lexical items that contain the speech segments, which are affected by the phonological process operating at particular stages of development” (Misra, 2009).</td>
</tr>
</tbody>
</table>

**Language in India**  [www.languageinindia.com](http://www.languageinindia.com) **ISSN 1930-2940 13:7 July 2013**

B. A. Mahalakshmi Prasad, M.A., and Dr. K. S. Prema, Ph.D.
Use of Markers Observed in the Spoken Language Lexical Corpora of Children in Kannada Language
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Study Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vasanta, Sastry &amp; Maruth (1995, as cited in Misra, 2009)</td>
<td>Studied the development of metalinguistic ability in children, awareness of metalinguistic skills in Telugu speaking children between the ages of 4.5 to 8.5 years concluded that metalinguistic ability improved with age. Children between the ages of 6.5 to 8.5 years were able to perform well in tasks such as sentence correction.</td>
</tr>
<tr>
<td>Prakash &amp; Mohanty (1995)</td>
<td>Studied the development of metalinguistic ability and its effect of children from grades 1 to 5 and concluded that phonemic awareness does not play a crucial role in learning to read.</td>
</tr>
<tr>
<td>Prema (1979)</td>
<td>Reported a gradual improvement in the performance of Kannada speaking children on phonological tasks (rhyme identification, phoneme oddity, phoneme segmentation, phoneme identification), which reached a ceiling level by grade V.</td>
</tr>
<tr>
<td>Narasimhan (1998)</td>
<td>Studied language acquisition in children and concluded that acquisition of language is based on pragmatic consideration of language use.</td>
</tr>
<tr>
<td>Nayak (2002)</td>
<td>Reported on the comparative study of developmental patterns in the acquisition of phonological awareness in Marathi speaking children with normal and impaired hearing studying in junior kindergarten and concluded that children in both groups obtained lower scores on phonological awareness tasks than children in grade I.</td>
</tr>
<tr>
<td>Shyamala &amp; Devi (2003)</td>
<td>Reported the developmental milestones of language acquisition in Kannada and Hindi, which revealed that children acquired Kannada and Hindi languages in almost similar manner with a few differences in the acquisition of verb inflections in Kannada speaking children, was attained at (42-48 months). Hindi speaking children attained the same at the age of 24-30 months and they inflected nouns, pronouns,</td>
</tr>
</tbody>
</table>
and adjectives for plurality in Hindi by children at the age of 36-42 months while Kannada speaking children had yet to attain the aspirated and non-aspirated contrasts by the age of 54-60 months. The study also reiterated the fact that children learning two languages simultaneously attempt to build a unitary lexical system that draws from both the languages. The study also indexed the two stage of acquisition of phonology in children i.e. stage one being the process wherein the word is paid attention to, in an undifferentiated manner and in stage two the child sorts the articulatory differences that make up the word.

<table>
<thead>
<tr>
<th>Study (Year)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kumudavalli (1973)</td>
<td>Investigated the relationship between articulation and discrimination of Kannada speech sounds in terms of distinctive features in children between the ages of four and eight years. The study discovered that there was a definite pattern in the development of discrimination. Features of voicing and nasality were distinguished at an earlier stage. Children acquired all the distinctions by the age of 8 years.</td>
</tr>
<tr>
<td>Sreedevi (1976)</td>
<td>Studied the aspects of acquisition of Kannada by 2+ years old children found that the additions of /-illa/, /be:da/ are acquired earlier than other type of negative markers with modal auxiliaries.</td>
</tr>
<tr>
<td>Tasneem (1977)</td>
<td>Investigated the acquisition of Kannada phonemes in one hundred and eighty typically developing schoolchildren in the age range of 3 to 6.6 years in the city of Mysore. The results indicated that there was a statistically significant difference in articulation score for different age groups and between genders in the same age group. The study also found a definite pattern in the acquisition of articulation and showed that the socioeconomic status affected the acquisition of phonology. There was a gradual change from age to age and the fricative /h/ was not acquired by the age of 6:6.</td>
</tr>
<tr>
<td>Author (Year)</td>
<td>Study Details</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Subramanya (1978)</td>
<td>Studied the development of some morphological categories in children Kannada: A study of children 6-8 years age range indicated the use of morphosyntactic markers by children and discovered that Morphological markers are used to indicate changes in number, gender and tense. The plural marker /galu/ was predominantly used but, /aru/ and /andiru/ was not acquired. Order of acquisition was /galu/, /aru/ and /andiru/. Gender allomorph /-i/ and /alu/ was acquired. The children had difficulty using /-e/ and used it less frequently. The children had not acquired /-gitti/. The order of acquisition was /-i/, /alu/, /-e/ and /-gitti/. The children showed the ability to use future and past tense. The tense allomorph was more difficult to use than gender and plural.</td>
</tr>
<tr>
<td>Roopa (1980)</td>
<td>Studied some syntactic development in 4-5 years old Hindi speaking children. The study looked into the spontaneous speech, storytelling, description of a picture book, describing view master slides, playing with toy animals, etc. of four children living in the city of Mysore. The study reported that sentence structures of children are similar to that of the adults though certain sentences showed a lack of noun and verb agreement. Word negations were not seen in the samples obtained. The children used both affirmative and imperative sentences in their speech. The 5-year-old children used almost all the structures used by the 4 year old. Differences between boys and girls in the same age group are evident only in structure used rarely by the children.</td>
</tr>
<tr>
<td>Venugopal (1981)</td>
<td>Studied the production of certain syntactic elements like negation, interrogation, imperative, coordination, pronominalisation and relativization in Tamil speaking children between the ages of five</td>
</tr>
</tbody>
</table>
and six. The study reports that the syntax structure of children is similar to that of adults. 6-year-old boys and 5-year-old girls rarely showed disagreement between noun and verb, pronoun and verb in their sentences. Consistent use of negative affixes /–aad/ was not observed. Transformational rules for deriving interrogatives sentences have been acquired by the children.

| B. A. Mahalakshmi Prasad, M.A., and Dr. K. S. Prema, Ph.D. |
| Use of Markers Observed in the Spoken Language Lexical Corpora of Children in Kannada Language |

| Vijayalakshmi (1981) |
| Developed a test for acquisition of syntax in Kannada speaking children up to five years of age. The performance of eighty-five children on TASK has led to the following conclusions that the process of acquisition of syntax shows systematic development in acquiring more and more lexical structures and sentences types of age progresses. Comprehension of language is better than expression when children begin to speak and this difference exists until the children are around 3-6 years of age. Comprehension of language starts early and develops faster than expression until the age of about 3-6 years. The expressive ability picks up faster and competes closely with that of comprehension after the age of 3-6 years. Girls perform better than boys in the age range between 2.0 to 3.0 years. From 3.0 years onwards, boys pick up faster. Around the age of 5 years both, perform similarly. There is universality in the process of syntax acquisition, as seen from the general agreement with the report of other language. |

| Madhuri (1982) |
| Studied some aspects of syntactic development in Marathi speaking children aged 21/2 to 3 years. The study reported that the sentences structure of the children were similar to that of the adult. The children did not consistently maintain concord between gender, number of the noun and the verb in their sentences. Three-year-old children used more abstract nouns, case endings, than two and half- |
year olds. Coordinated speech samples were present in the speech samples while pronominalization within sentences was not observed. A 6-8 months difference in age caused considerable difference in the sentence structures of the children.

| Prakshan (1999) | Developed a Picture Speech Identification Test for Children in Tamil between the ages of 3-6.6 years. It reported that an age related difference in the performance in the speech identification score with a significant statistical difference between 3-3.11 and oldest 6-6.11. |
| Lahl Mangaihi (2009) | Worked on development and standardisation of spondees and phonetically balanced word list in Mizo that can be used to measure the Speech Recognition Test and Speech Identification Score for native speakers of Mizo. The list was created using familiar bisyllabic and monosyllabic randomly selected words from different sources like newspapers, books and telephonic conversations of individuals fluent in the language. |

The above studies have examined the various aspects of language acquisition in typically developing children along with children exhibiting deficiency in language acquisition from the age of two years onwards. However, by the age of six years, children start going to school and their process of language acquisition sees an accelerated growth due to various factors such as interaction with peer group, and so on. The studies also lack normative data on the lexical acquisition of children from the age of 6 years.

**Method**

A total of two hundred and forty typically developing normal children’s speech ability, was ascertained by administering the WHO ten-Question disability screening checklist (Singhi, Kumar, Malhi & Kumar, 2007), as well as using the teacher’s report. The selected participants’ mother tongue was Kannada; they lived in predominantly Kannada speaking areas, and attended
state run Kannada medium schools. The age group of the participants was between the ages of six and eight years with a mean difference in their ages of about six months. The participants were divided into two groups of thirty each according to gender and were assigned to the four age groups namely 6-6.5, 6.6-7.00, 7.1-7.5 and 7.6-8.00 (30 Boys ; 30 Girls) in each group.

Table 1.2

<table>
<thead>
<tr>
<th>Age</th>
<th>Boys</th>
<th>Girls</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-6.5</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>6.6-7.00</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>7.1-7.5</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>7.6-8.00</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>240</strong></td>
<td></td>
</tr>
</tbody>
</table>

Spoken language samples were elicited from the participants through *spontaneous utterances* obtained during the narration of daily activities and *elicited utterances* to standard pictures as well as picture description and story-telling. Each participant was seen individually in a library or in a quiet classroom. The child was made comfortable with a few general questions to build rapport as well as to familiarise the child with recording equipment. Instructions were given to the child to describe the pictures that were shown to him/her. Tasks were administered in the following order:

- The first task administered was Story-telling (Standardised pictures, Nagapoornima, 1990). In this task, the participant was shown a series of pictures and asked to describe the picture by making a story from the sequential pictures presented. The responses were recorded on a digital recorder that was later orthographically transcribed and analysed using SALT software.
• The second task was narration of experiences in their daily routine activities as a part of spontaneous language sample. The responses were recorded on a digital recorder that was later orthographically transcribed, fed into SALT software for analysis.

• The third task was picture description in which Computerized linguistic protocol for screening - CliPS (Anitha & Prema, 2004) was administered. This was administered using a laptop on which a series of line drawing were presented and the participant was asked to name/point to each picture as instructed.

The recorded data was transferred on to a computer, transcribed and each linguistic unit was tagged according to the conventions of SALT keeping the rules of Kannada language.

Results and Discussion

The data was subjected to Univariate Analyses of Variance to examine associations between age group, gender and lexical categories. The results show that with the children of the age group of eight years, the mean percent frequency of lexical categories uttered by the participants is in the following order, the highest being PNG Markers (72.36), Adjectives (59.20), Nativised words (55.86), Verb (53.19), Nouns, (53.07), Numbers (51.00), Negatives (49.16), Comparatives (48.20), Dialectal Varieties (43.71), Prepositions (42.11), Code switches (39.45), Pronouns (39.96), Error words (32.32), Non words (29.50), New words (26.55), Question tags (16.25) and the least being Standard words (12.49). The mean percent frequency for overall developmental pattern was greater (53.23) at eight years compared to other age groups.

The present paper is concentrating on the usage of Markers by the participants and hence concentrates only on that section

Kannada is an agglutinative language wherein the root takes the markers that represent the markers for person, number and gender. The following table represents the Mean and S.D Scores obtained for PNG Markers.

Table 1.3

Mean and S.D of Markers

Language in India www.languageinindia.com ISSN 1930-2940 13:7 July 2013
B. A. Mahalakshmi Prasad, M.A., and Dr. K. S. Prema, Ph.D.
Use of Markers Observed in the Spoken Language Lexical Corpora of Children in Kannada Language
<table>
<thead>
<tr>
<th>Age group</th>
<th>Boys</th>
<th>Girls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D</td>
</tr>
<tr>
<td>6.0-6.5</td>
<td>74.07</td>
<td>4.74</td>
</tr>
<tr>
<td>6.6-6.7</td>
<td>73.16</td>
<td>5.01</td>
</tr>
<tr>
<td>7.1-7.5</td>
<td>73.19</td>
<td>4.75</td>
</tr>
<tr>
<td>7.6-8.0</td>
<td>71.54</td>
<td>4.94</td>
</tr>
</tbody>
</table>

Figure 1.3: Performance of Participants across Gender, Age and Lexical Category of PNG Markers

The mean scores for PNG markers are specified in Table 3.1.12, which illustrated the usage of PNG markers across age group and gender. It can be concluded from Table 3.1.12 and Figure 3.1.12 in the age group of 6.0-6.5 the utterances of boys with a mean of 74.07 (S.D=4.74) contained a higher number of PNG markers than the utterances of girls with a mean of 72.09 (S.D=4.99). In the age group of 6.6-7.0 it was observed that the utterances of boys with a mean of 73.16 (5.01) contained a higher number of PNG markers than girls with a mean of 71.05 (S.D=8.38). In the age group 7.1-7.5 utterances of boys with a mean of 73.19 (S.D=4.75) contained an
almost equal number of PNG markers with the utterances of girls with a mean of 73.15 (S.D = 8.09). In the age group of 7.6-8.0 it was observed that the PNG markers contained in the utterances of boys subject with a mean of 71.54 (S.D = 4.94) was higher than the PNG markers contained in the utterances of girls with a mean of 70.70 (S.D = 8.10). The data of age with four categories and gender for verbs were analysed using univariate ANOVA. The results of the study revealed significant difference in the interaction effect \([F (1, 232) = 0.36, p < 0.78]\). Further there was a significant difference in the age \([F (1, 232) = 1.38, p < 0.24]\). Duncan’s post hoc analysis revealed no significant difference in the age groups.

A sloping trend is observed in the utterances of PNG markers from the age group of 6.0-6.5 to 7.6-8.0. The highest number of PNG markers being uttered by boys in the age group of 6.0-6.5 and the lowest being uttered by boys in the age group of 7.6-8.0. The utterances of PNG markers observed in the expression of boys in the age groups of 6.0-7.0 and 7.6-8.0 was higher than the PNG markers observed in the utterances of girls. However, in the age group of 7.1-7.5 boys uttered approximately the same number of PNG markers as girls. Due to the lack of significant difference in the category of PNG markers, it may be assumed that participants in the age group of 6.0-8.0 have already acquired the PNG markers.

**PNG Markers as Indicators of Socio-Economic Status**

The socio-economic conditions of the participants were calculated using the scale given by (N.I.M.H., 1999). The five-point scale considers the parameters of occupation, highest education score, annual family income, property and per capita income per annum. Analyses of the demographic data revealed that as most of the parents of the participants were unskilled workers, illiterate or have studied less than SSC (Secondary School Certificate), per capita income is below Rs. 15,000, the pecuniary income of the family per year is below Rs. 1 lakh and they posses no property; they are classified as ‘SES ONE’ or participants belonging to the lower socio-economic strata. However, out of the 240 participants, 33 participants belonged to ‘SES TWO’ category.

Table 1.4

*Indicators of Lower Socio-Economic Conditions as Shown by use of PNG Markers*
Mann Whitney U-test was done to check the significance of difference in total developmental pattern between Group I and Group II. The results show that significant difference does not exist between the two groups $|Z|=0.26$, $p>0.79$.

An evaluation was made of the linear relationship between socio economic scale and total developmental pattern using Pearson’s correlation. An analysis of Pearsons correlation coefficient indicates a statistically insignificant ($p>0.52$) linear relationship between socio-economic scale and total developmental pattern. The mean and S.D of group I and group II is given in Table 3.11. The mean and S.D of group I is 52.06 and 32.69. The mean and S.D of group II of SES is 56.04 and 30.54. From the results it can be interpreted that group II has a higher TDP than group I. To determine the dependence between socioeconomic scale and total developmental pattern Spearman’s correlation co-efficient was calculated. The results revealed statistically insignificant dependence ($p>0.79$) between socioeconomic scale and total developmental pattern as given by Figure 1.4.
As the participants in the present study belonged to the lower socio-economic strata, there was rampant use of inconsistent markers in their speech. An over extension as well as under extension of meaning of words were also observed. For example, during administration of CLIPS participants used the word /huDuga/ to depict /magu/ ‘child’, /huDuga/ ‘boy’ and /yuvaka/ ‘youth’.

Sociological constraints on the use of gender markers indicate that the use of PNG markers created certain limitation while tagging the data. Usage of neutral markers [aite] for PNG markers by the lower socio-economic strata in place of gender markers.

Examples:

1. /huDuga/ /ball/ /aDtaite/ (‘boy playing ball’).
2. /appa/ /malgaite/ (‘father sleeping’).

In the above examples /huDuga/ which means ‘boy’ and /appa/ which means ‘father’ respectively should be followed by the gender marker /iddane/ as the gender marker.

However majority of the subjects (as shown in the statistical analysis) have used the neutral
nonhuman marker /aite/ in the colloquial sense rather than the standard /ide/ instead of the human gender marker.

3. /amma/ /aDige/ /maDtaite/ (‘mother is cooking’).
   In the above example /amma/ which means mother should be followed by the gender marker /iddale//le/ as the gender marker. However majority of the subjects (as shown in the statistical analysis) have used the neutral nonhuman marker /aite/ in the colloquial sense rather than the standard /ide/ instead of the human gender marker.

4. /bekku/ /haal/ /kuDitaite/ (‘cat milk drinking’/ ‘cat is drinking milk’).

5. /na:yi/ /jump/ /hoDitaite/ (‘dog jumping’/ ‘dog is jumping’).

   In the above examples /bekku/ which means ‘cat’ and /na:yi/ which means ‘dog’ respectively should be followed by the nonhuman marker /ide/ majority of the subjects (as shown in the statistical analysis) have used the neutral nonhuman marker /aite/ in the colloquial sense rather than the standard /ide/.

   It has been observed that in a particular, irrespective of gender the neuter gender marker [aite] was used to describe both animate and inanimate [human and non human verb form]. While in some schools it was observed that neuter gender marker along with male gender marker was used for animate, inanimate [human and non human forms] wherein the marker for female gender and non human verbs has been replaced by neuter gender [aite] while retaining the male gender marker for human, male [human verbs].

Conclusion

The present paper is a report of the partial results that have been arrived at as part of the Indian endeavour to establish spoken language lexical corpora of children between the ages of 6-8 years living in the city of Mysore. The language samples of two hundred and forty children were elicited using spontaneous and picture narration. It can be inferred from the results that Markers play a very important role in depicting the stages of acquisition as well as its relation to socio-economic variables. However, the present study may be considered as a mere drop in the ocean that is corpus studies. The study of markers highlights or brings out the relevance of corpus studies and language acquisition which need further investigation into the phenomenon.
References


*Language in India* [www.languageinindia.com] ISSN 1930-2940 13:7 July 2013

B. A. Mahalakshmi Prasad, M.A., and Dr. K. S. Prema, Ph.D.
Use of Markers Observed in the Spoken Language Lexical Corpora of Children in Kannada Language


**Language in India** www.languageinindia.com ISSN 1930-2940 13:7 July 2013

B. A. Mahalakshmi Prasad, M.A., and Dr. K. S. Prema, Ph.D.

Use of Markers Observed in the Spoken Language Lexical Corpora of Children in Kannada Language


B. A. Mahalalshmi Prasad  
#2961/75. 2nd Main, 5th Cross  
Saraswathipuram  
Mysore-570009  
Karnataka  
India  
*machiprasad@gmail.com*

A.E.S. National College  
Gouribidanur, Karnataka, India

Prema K.S. Rao, Ph.D. (Speech and Hearing)  
Professor of Language Pathology in Department of Speech Language Sciences  
Head  
Department of Special Education  
All India Institute of Speech and Hearing  
Manasagangotri, Mysore-570 006  
Karnataka  
India  
*prema_rao@yahoo.com*