International Journal of Research in Medical Sciences
Kant AR et al. Int J Res Med Sci. 2015 Dec;3(12):3534-3542
www.msjonline.org
pISSN 2320-6071 | eISSN 2320-6012

DOI: http://dx.doi.org/10.18203/2320-6012.ijrms20151394

Research Article

Analysis of characteristics of semantics of spoken language in normally developing Hindi speaking children

Anjali R. Kant[1]*, Bisma S. Dafadar[2], Arun A. Banik[1]

[1]Department of Speech and Language Pathology, Ali Yavar Jung National Institute for the Hearing Handicapped (AYJNIHH), Mumbai, Maharashtra, India
[2]Department of Child Development, Surya Mother and Child Care Super Speciality Hospital, Mumbai, Maharashtra, India

Received: 29 September 2015
Revised: 10 October 2015
Accepted: 16 November 2015

***Correspondence:**
Mrs. Anjali R. Kant,
E-mail: anjali.kant151@gmail.com

ABSTRACT

Background: There appears to be a lack of database of and dearth of studies focusing on the characteristics of semantics in Hindi speaking school aged children. Such a data base will be useful for building vocabulary for language disordered children and for constructing AAC boards for non-verbal children. Hence, it is essential to study the characteristics of semantics of normally developing children. This paper focuses on describing the semantic characteristics of spoken language in Hindi speaking children.
Methods: 200 normally developing Hindi speaking children within the age group of 3 - 7 years were shown and instructed to describe three validated pictures of daily events. The responses were recorded and transcribed. Analyses included type-token ratio, frequency of occurrence and comparisons between different word classes.
Results: Percentage of nouns is highest followed by verbs, pronouns, adjectives. Frequency of occurrence of words increases with increase in age. The common words with high frequency of occurrence are hɛ̃, hũ, rʌhe, rʌha, rʌhi, dʒa, ɔr, khel, gaḍi, log, pe, ke. There appears to be marked increase in different classes of words, one at 4 yrs of age (after Sr. KG) and other at 6 yrs of age (standard I).
Conclusions: One of the highlighting features of this study is the huge database of semantics (of spoken language) collected from 200 school going children. Creating such a database and utilizing it for assessing language of the disordered population appears to be the need of the hour.

Keywords: Hindi, Semantics, Type-token ratio, Frequency of occurrence

INTRODUCTION

Children acquire the language or languages to which they receive sufficient exposure during childhood. The development of comprehension and use of oral language in children from the first to 36 months has been given in detail.[1] This learning process is referred to as first-language acquisition, since it requires no direct teaching. Vocabulary development is a process by which children acquire words. Babbling shifts towards meaningful speech as infants grow and produce their first words around the age of one year. By the age of 18 months, infants can typically produce about 50 words. Between 18 to 24 months, child recognizes names of many familiar objects and persons. His vocabulary consists of 50-75 words. Between 24-30 months his vocabulary

grows to about 272 words consisting of nouns (38.6%), verbs (21.0%), adverbs (7.1%), pronouns (14.6%) and unclassified (18.7%). Between 30-36 months, his vocabulary consists of 446 words. Sentence structure advances rapidly and complex and compound sentences are used which are 6-8 words in length. He uses many 'how' and 'why' questions in response to speech of others. By the end of five years sentence structure expands rapidly in accuracy and complexity. Mean number of words in his vocabulary is 2072. Between 6 and 7 years, the child develops command of every form of sentence structure, perception and inner language. He comprehends the meaning of 4000 words.[2]

According to the 2001 Indian census, 258 million people in India reported their native language to be "Hindi". After independence, the Government of India worked on standardizing Hindi and the following changes took place: Hindi grammar was standardized. Hindi became the official language of India on January 26, 1965. Hindi was also recognized as the common second language of Mauritius, Fiji, Trinidad, Guyana and Surinam. It is indeed interesting to note that despite Hindi being the national language of India and many of its dialects being used country wide, there appears to be a dearth of scientifically collected and statistically analyzed corpora required to further research the complex process of acquisition of speech and its intermingled sub-parts, by children.

There are few studies which have focused on the development of semantics and syntax in Hindi speaking children. An investigation into the basic Hindi vocabulary of 625 Hindi speaking children of Class V of 25 Hindi medium schools in Greater Mumbai was done with the objective to find out and recommend the basic vocabulary of pupils reading in class V.[3] Vocabulary was elicited by using a Questionnaire, Word List from textbooks, Interview Schedule Oral Word Test, Test for Recognition Vocabulary and Test for Reproduction Vocabulary. In the recognition vocabulary, 70.98 % of the words were nouns, 0.11% was pronouns, 18.6% were adjectives, 7.32% were verbs, 0.7% was adverbs, 2.02% were prepositions, 0.21% conjunctions and 0.06% was interjections. However, the study focused on the vocabulary of only standard V children. In a study, the author[4] has listed the 100 most frequent words for Hindi language.

In the profession of Speech-Language Pathology, children with speech and language disorders constitute a major proportion of our clientele. There are children with delayed/deviant/deficient speech and language due to a number of etiological factors and disorders, for example, hearing impairment, cerebral palsy, pervasive developmental disorder, specific language impairment, mental retardation, etc. Most of the tests we use to assess language, speech production, speech perception, and reading and writing skills are in English. A number of Academic Institutes in the country have developed tests in regional languages to study various aspects of language in terms of semantics and syntax. In Hindi there are a number of tests that are in the form of unpublished dissertation of Master's students at various universities. For example: A test for word finding skills in Hindi;[5] LPT in 7 Indian languages—1992, AYJNIHH and RRTC, Chennai; Test of Writing for Children in Hindi;[6] Boston Naming test in Hindi Mangalore University.[7]

However, most of these tests have selected specific stimuli to test the children depending on the criteria that they select to measure. There is a dearth of studies focusing on characteristics of semantics, syntax, discourse and narrative skills in Hindi speaking school-aged children. Further, there appears to be a lack of database with respect to vocabulary for different age groups.

Hindi is different from English as the word order in the sentence is subject-object-verb type, and also that it uses postpositions instead of prepositions. Hindi verbs are inflected with respect to gender of the subject (masculine, feminine), number of the subject (singular, plural), tense (present, past, future), action (perfect, imperfect, continuous), and degree of respect (intimate, familiar, respect). Verbs are referred to in their infinitive noun form which ends typically with 'na' after the root verb. This makes it definite that the Hindi vocabulary and grammar differs from that of English in more than one way.

It is required that we also know the most frequently used words by children and the use of syntax in spontaneous speech of these children. Such database will prove to be extremely useful while constructing AAC boards for the non-verbal children, building vocabulary for the language disordered children and creating test material as many Western tests cannot be simply translated due to the socio-cultural and ethnic differences. Also knowledge of the pattern of sentences helps detecting delay or disorder of expressive language of the older non-English speaking children. Knowledge of type of discourse (narrative or descriptive) is also useful as descriptive discourse may be simpler to produce.[8] Hence, the need of the hour is to study the development of semantics, syntax and discourse of normally developing children.

This paper focuses on characteristics of semantics of normally developing Hindi speaking children. The aim of this paper was to describe and compare the semantic characteristics of spoken language in Hindi speaking children belonging to the two age groups.

The objective of the present study was to describe and compare the following across two age groups: a) Class of words; and b) Frequency of occurrence of different class of words.

An attempt is made to determine whether the frequencies of occurrence of class of words were dependent on or independent of the age of the children.

METHODS

200 children with no sensory or motor abnormalities were randomly selected from Hindi medium Municipal schools. They were divided into 2 groups according to their chronological ages viz. Group I is comprised of 100 children between 3 yrs to 5 yrs of age (50 children between 3 to 4 years and 50 between 4 to 5 years). Group II comprised of 100 children between 5 yrs to 7 yrs of age (50 children between 5 to 6 years and 50 between 6 to 7 years). Children with intellectual or behavior problems, as per the teacher's and parent's report, were not included. Three pictures illustrating 'market scene', 'street scene' and 'railway station scene', which were validated by four Speech-Language Pathologists;[9] one picture from Balbharati textbook used in schools for the aforementioned age range and spontaneous speech were used as stimuli/material to elicit verbal responses from the participants.

The aims, objectives and procedure were explained to Principals of the schools and their written consent was sought. The three validated pictures and a picture from the Balbharati textbook of standard I was shown to the children one by one and they were asked to describe what they saw in the pictures. Children were then instructed to talk on two topics 'Myself' and 'Their daily routine' for 3 to 5 minutes. The responses of children were recorded on Cenix digital voice recorder (Vr-p2170) and later transcribed in IPA and analyzed.

The results which were obtained after the qualitative and statistical analysis are presented under the following headings:

1. Analysis of class of words by calculating the type token ratio.

2. Descriptive analysis of: a) Class of words; b) Frequency of occurrence of words; c) Class of words and frequency of their occurrence.

3. Appropriate parametric / non-parametric test to compare differences between the groups.

1. Analysis of class of words by calculating the type token ratio: The type-token ratio (TTR) is a measure of vocabulary variation within a written text or a person's speech. It is shown to be a helpful measure of lexical variety within a text. It can be used to monitor changes in children and adults with vocabulary difficulties.

Type-Token Ratio = (number of types / number of tokens) X 100. The term refers to the "relationship between the number of different word forms, or types, and the number of running words, or tokens".10 The TTR is then multiplied by 100 to give the percentage of unique words in a text." According to Biber, a high type/token ratio marks high density of information. The Class of words in terms of the type token ratio (%) across the four age groups is depicted in Table 1.

Table 1: Type Token Ratio in % for various pictures and conversation across four age groups.

Type token ratio in % for class of words / Age groups	3-4 yrs	4-5 yrs	5-6 yrs	6-7 yrs
Adjectives	6.71	7.14	5.86	6.47
Adverbs	1.24	1.55	1.85	1.94
Affirmative	0.49	0.31	0.28	0.25
Auxillary	1.49	0.93	1.28	0.64
Conjunction	1.74	0.93	0.85	0.64
Nouns	42.03	44.56	43.2	45.46
Postposition	1.99	1.55	1.28	1.29
Pre position	2.48	1.7	1.71	1.55
Pronoun	7.21	5.59	5.15	5.56
Verb	34.57	35.71	38.48	36.13

2. Descriptive analysis of:

a) Class of words: The number of classes of words across the four age groups are depicted in Table 2.

Table 2: Number of Class of words for various pictures and conversation across four age groups.

Class of words and their number /Age groups	3-4 yrs	4-5 yrs	5-6 yrs	6-7 yrs
Adjectives	27	46	41	50
Adverbs	5	10	13	15
Affirmative	2	2	2	2
Auxillary	6	6	9	5
Conjunction	7	6	6	5
Nouns	169	287	302	351
Postposition	8	10	9	10
Pre position	10	11	12	12
Pronoun	29	36	36	43
Verb	139	230	269	279

b) Frequency of occurrence of words: It was observed that the total no. of words obtained for the age group 3-4 yrs is 409, for the age group 4-5 yrs is 655, for the age group 5-6 yrs is 706 and for the age group 6-7 is 787. Thus it is apparent that the no. of words increases with increase in age. Given below is a list of all words for each age group, whose frequency of occurrence is above 100.

3 – 4 yrs : hɛ - 753; je – 562; rʌhe – 520; rʌhi -288; khel – 252; dʒa – 243; rʌha – 240; gaḍi – 171; lʌḍki – 110; lʌḍka – 105; pani – 101.

4 – 5 yrs: hɛ - 1827; je – 1254; ; rʌhe – 611; rʌha – 438; rʌhi – 368; dʒa – 303; khel – 267; ɔr – 262; log – 150; pani – 145; gaḍi – 144; inlog – 138; lʌḍki – 132; mẽ – 129; pe – 123; lʌḍka- 103; hũ – 102.

5 – 6 yrs : hɛ - 2075; je – 1252; rʌhe – 750; rʌha – 468; rʌhi – 415; ɔr – 364; dʒa – 343; khel – 314; inlog – 212; gaḍi – 150; log – 142; tʃʌla – 135; pani – 129;; lʌḍki – 132; mẽ – 115; pe – 110; ke – 109; hũ – 102; kha – 105.

6 – 7 yrs : hɛ - 2425; je – 1582; rʌhe – 766; rʌha – 532; rʌhi – 481; ɔr – 368; dʒa – 340; khel – 289; log – 153; gaḍi – 142; lʌḍki – 132; mẽ – 127; pe – 127; kha – 122; pani – 119; ke – 116.

From the above it is observed that there are a number of words which are common across all age groups which are as follows: hƐ, je, rΛhe, rΛha, rΛhi, khel, gaḍi, dʒa, log, hũ, pe, ke.

c) Class of words and frequency of their occurrence: When the various words obtained are classified according to different classes, viz. nouns, verbs, adjectives it is feasible to compare their frequency of occurrence across the age groups. Hence, the class of words and their frequency of occurrence as per the different age groups are put forth in Table 3.

Table 3: Frequency of occurrence of class of words across four age groups.

Frequency of occurrence of class of words/Age groups (yrs)	3-4 yrs	4-5 yrs	5-6 yrs	6-7 yrs
Adjectives	90	332	305	351
Adverbs	26	52	67	98
Affirmative	764	1831	2083	2436
Auxillary	1058	1428	1728	1785
Conjunction	139	348	475	469
Nouns	2045	2733	2900	3269
Post position	183	527	537	635
Pre position	94	203	216	196
Pronoun	733	1660	1650	1866
Verb	1437	2289	2724	2850

Figure 1 illustrates clustered bar chart showing frequencies (of occurrence) of first five classes of words by age group.

Figure 2 illustrates clustered bar chart showing frequencies (of occurrence) of last five classes of words by age group.

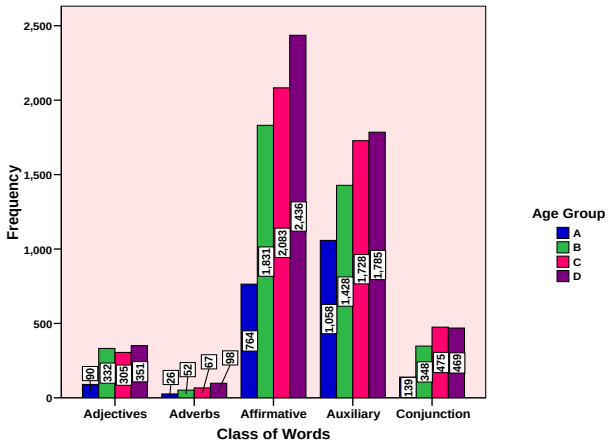

Figure 1: Clustered bar chart showing frequencies (of occurrence) of first five classes of words by age group.

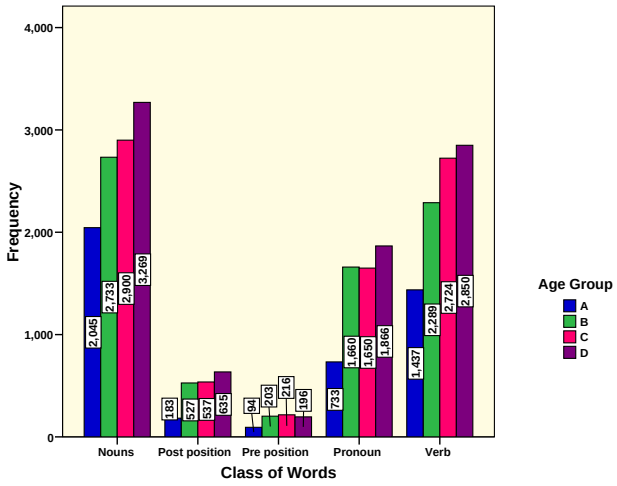

Figure 2: Clustered bar chart showing frequencies (of occurrence) of last five classes of words by age group.

3. Appropriate parametric / non-parametric tests to compare differences between the groups- Comparisons between different classes of words:

Comparisons were done between the following:

a) Adjective vs. adverbs, affirmative, auxiliary, conjunction, nouns, post position, preposition, pronoun and verb.
b) Adverbs vs. affirmative, auxiliary, conjunction, nouns, post position, preposition, pronoun and verb.
c) Affirmative vs. auxiliary, conjunction, nouns, post position, preposition, pronoun and verb.
d) Auxiliary vs. conjunction, nouns, post position, preposition, pronoun and verb.
e) Conjunction vs. nouns, post position, preposition, pronoun and verb.
f) Nouns vs. post position, preposition, pronoun and verb.
g) Post position vs. preposition, pronoun and verb.

h) Preposition vs. pronoun and verb.
i) Pronoun vs. verb and other classes of words.

In the present study only three comparisons will be focused upon and discussed, viz.

A. Affirmative vs. auxiliary, conjunction, nouns, post position, preposition, pronoun and verb.
B. Auxiliary vs. conjunction, nouns, post position, preposition, pronoun and verb.
C. Nouns vs. post position, preposition, pronoun and verb.

A. Comparison between frequencies of "Affirmatives" and frequencies of other classes of words, across all age groups:

The Chi-square test results for comparison of frequencies across the age Groups for the class of word "Affirmatives" with that for the seven other Class of words are as follows- 1. Auxiliary - Chi-square$_{(3)}$ = 139.92*(p-value^{+} = 0.0005); 2. Conjunction - Chi-square$_{(3)}$ = 8.98 (p-value^{+} = 0.03); 3. Nouns - Chi-square$_{(3)}$ = 214.22* (p-value^{+} < 0.005); 4. Postposition - Chi-square$_{(3)}$ = 4.76 (p-value^{+} =0.19); 5. Preposition - Chi-square$_{(3)}$ = 14.46* (p- value^{+} =0.002); 6. Pronoun - Chi-square$_{(3)}$ = 23.47* (p-value^{+} < 0.005); 7. Verb - Chi-square$_{(3)}$ = 84.97* (p-value^{+} < 0.0005). The data presented reveals that five comparison out of the seven are statistically significant (p<.006).

Details of the major differences in the frequencies across different ages with respect to the selected pairs of words are presented below:

1. *Affirmative by Auxiliary:* The auxiliaries are 7% higher than affirmatives for age group A (i.e.3 to 4 yrs). Affirmatives are 5% higher than auxiliaries, for age group D (i.e. 6 to 7yrs) in terms of their frequency of occurrence.

2. *Affirmative by Noun:* The nouns are 8% more than affirmatives, for the age group A (i.e. 3 to 4 yrs). Exactly opposite is the situation for age group D (i.e. 6 to 7 yrs). For age group D, the percent of affirmatives is higher by 5% in comparison to that for the nouns in terms of their frequency of occurrence.

3. *Affirmative by Preposition:* The affirmatives are 7% higher than prepositions for age group D (i.e. 6 to 7 yrs). For other age groups the prepositions percent is higher by 1 to 3% than that for the affirmatives in terms of their frequency of occurrence.

4. *Affirmative by pronoun:* For age group D, the affirmatives are 3% more than pronouns whereas, for the age group B (i.e.4 to 5 yrs), the pronouns are 2% more than affirmatives in terms of their frequency of occurrence.

5. *Affirmative by verb:* There are 5% more verbs than affirmatives for age group A (i.e. 3 to 4 yrs). Quite contrary to this the affirmatives are 6% more than verbs, for age group D (i.e. 6 to 7 yrs) in terms of their frequency of occurrence.

B. Comparison between frequencies of "auxiliaries" and frequencies of other classes of words, across all age groups:

The Chi-square test results for comparison of frequencies across the age Groups for the class of word "auxiliaries" with that for the six other Class of words are as follows- 1. Conjunction - Chi-square$_{(3)}$ = 58.12* (p-value^{+} < 0.0005); 2. Nouns - Chi-square$_{(3)}$ = 12.05 (p-value^{+} = 0.007); 3. Postposition - Chi-square$_{(3)}$ = 74.59* (p-value^{+} < 0.0005); 4. Preposition - Chi-square$_{(3)}$ = 14.72* (p-value^{+} = 0.002); 5. Pronoun - Chi-square$_{(3)}$ = 79.33* (p-value^{+} < 0.0005); 6. Verb - Chi-square$_{(3)}$ = 12.90* (p-value^{+} < = 0.005). The data presented reveals that five comparison out of the six are statistically significant (p<.006).

Details of the major differences in the frequencies across different ages with respect to the selected pairs of words are presented below:

1. Auxiliary by conjunction: Age group A (i.e. 3 to 4 yrs) has almost 8% more auxiliaries than conjunctions whereas, age group C (5 to 6 yrs) and D (i.e. 6 to 7 yrs) has 4% and 3% respectively more conjunctions than auxiliaries.

2. Auxiliary by post position: Like the comparison of auxiliary with conjunction, age group A (i.e. 3 to 4 yrs) has almost 8% more auxiliaries than post positions. However, both age group B (i.e.5 to 6 yrs) and D (i.e. 6 to 7 yrs) has 4% more post positions than auxiliaries.

3. Auxiliary by preposition: Percent of auxiliaries is more by 4% compared to prepositions, in the case of age group A (i.e. 3 to 4 yrs). Contrary to this; percent of preposition is more by 4% compared to auxiliaries, for age group B (i.e.4 to 5 yrs).

4. Auxiliary by pronoun: In the case of age group A (i.e.3 to 4 yrs) auxiliaries are more by 5% than pronouns. However, in the case of age group B (i.e. 4 to 5 yrs) pronouns are more by 4% than auxiliaries.

5. Auxiliary by verb: Except for the age group A (i.e. 3 to 4 yrs) for all other age groups percent of verbs is slightly more than auxiliaries. In the case of age group A auxiliaries are 2% more than verbs.

C. Comparison between frequencies of "noun" and frequencies of other classes of words, across all age groups:

The Chi-square test results for comparison of frequencies across the age Groups for the class of word "Noun" with that for the four other class of words are as follows- 1. Post positions - Chi-square$_{(3)}$ = 90.46* (p-value^{+} < 0.0005); 2. Prepositions - Chi-square$_{(3)}$ = 19.26* (p-value^{+} < 0.005); 3. Pronoun - Chi-square$_{(3)}$ = 112.75* (p-value^{+} < 0.0005) 4. Verb - Chi-square$_{(3)}$ = 45.95* (p-value^{+} < 0.0005). The data presented reveals that all the four comparisons are statistically significant (p<.006).

Details of the major differences in the frequencies across different ages with respect to the selected pairs of words are presented below:

1. *Noun by post position:* Children in age group A (i.e. 3 to 4 yrs) have 9% more nouns than post positions. Exactly opposite is the situation for age group B (i.e. 4 to 5 yrs), C (i.e. 5 to 6 yrs) and D (i.e.6 to 7 yrs). These have respectively 3, 2, and 4% more postpositions than nouns.

2. *Noun by preposition:* Age groups A (i.e. 3 to 4 yrs) and D (i.e.6 to 7 yrs) have respectively 5% and 2% more nouns than prepositions. As against this both age groups C (i.e. 5 to 6 yrs) and D (i.e. 6 to 7 yrs) have 4% more prepositions than nouns.

3. *Noun by pronoun*: Age group A (i.e. 3 to 4 yrs) has 6% more nouns than pronouns. However, the age group B (i.e. 5 to 6 yrs) has 3% more pronouns than nouns.

4. *Noun by verb:* Age group A (i.e.3 to 4 yrs) has 3% more nouns than verbs. However, the age group C (i.e. 5 to 6 yrs) has 3% more verbs than nouns.

DISCUSSION

1. Analysis of class of words by calculating the type token ratio: It is observed from the Table 1 that the percentage of nouns is highest across all age groups, followed by verbs, pronouns and adjectives. The percentage of prepositions, adverbs, conjunctions, auxiliary and affirmatives is very less when compared to the other classes of words. It is also observed that there is no increase in percentage of tokens across age groups.

As per the Longman Grammar of Spoken and Written English (LSWE) corpus study, [10] nouns are the most frequent lexical word class; verbs are less frequent, followed by adjectives and adverbs. In conversation, nouns and verbs are about equally frequent. In relation to the different class of words, the results of the present study are in agreement with the LSWE corpus study. The results of the email analysis used in the study mostly agree with those mentioned in LSWE are given below: the proportion of word classes within lexical words and their frequency- Nouns: 43.3%; Verbs: 33.5%; Adjectives: 13.6%; Adverbs: 9.7%. In relation to the TTR the results of the present study are in agreement with the aforementioned study especially in terms of nouns and verbs. In the present study, nouns range from 42.03% to 45.06% when compared to 43.3 % in their study; verbs range from 34.57% to 38.48% as compared to 37.5% of their study.

In a study, preliminary analysis of the DoE-CIIL (Department of Electronics-Central Institute of Indian Languages, Mysore) corpus developed in part by CIIL is presented. This includes corpora of 13 major Indian languages, each approximately three million words in size viz. Assamese, Bangla, Hindi, Kannada, Kashmiri, Malayalam, Marathi, Oriya, Punjabi, Sindhi, Tamil, Telugu and Urdu. The type token ratio obtained in various languages has been given. For Hindi language, the tokens obtained were 2980968 and the types were 1249932. The TTR was 4.19. However, most of the studies have focused on written language or 'script' and there appears to be a paucity of analysis in terms of frequency of occurrence and TTR in spoken language.

2. Descriptive analysis of:

a) Class of words: From the Table 2 it is observed that in all age groups nouns are most frequent than other class of words. They are followed by verbs, adjectives, pronouns and prepositions. The other classes of words viz. adverbs, auxiliary, conjunction and affirmative are very few in number. As age advances there appears to be a marked increase in the nouns, verbs, adjectives and pronouns and not so in other classes of words.

In a study,[11] the characteristics of vocabulary and grammar development in 512 Slovenian-speaking infants and toddlers aged 0.6 months and 2.6 yrs were examined using the Slovenian adaptation of the MacArthur-Bates Communicative Development Inventories (CDI). The findings suggested that nouns predominate in the vocabularies of infants and toddlers of various ages; as they age and with the increasing size of their vocabularies, the share of interjections decreases and the share of verbs and adjectives increases. The results of the present study are in agreement with the aforementioned study.

The results of a study in Tamil speaking children[12] revealed the following: pronouns, quantity adjectives, case markers and place adverbs are acquired between 2.6 to 3.6 yrs of age; post positions, manner adverbs, time adverbs and tenses are acquired between 3.6 to 4.6 yrs of age; color adjectives are acquired between 4.6 to 5.6 yrs of age. However, the results of the present study cannot be compared with the aforementioned study as different classes of words are used and the vocabulary of children till 5.6yrs of age is studied; whereas in the present study additional classes of words like prepositions, conjunctions, affirmatives etc. were studied.

The acquisition of transitive and intransitive constructions in a Hindi-speaking child was examined.[13] An analysis of verb constructions at each time point showed that from about 28 months of age, the child was

using at least two verbs in both transitive and intransitive forms. However, the main focus of this study was only on one class of words, that is, verbs, acquired by only one single subject. Hence, comparisons between the above study and the present study are not feasible.

b) Frequency of occurrence of words: It was observed that the total number of words obtained for the age group 3-4 yrs is 409, age group 4-5 yrs is 655, age group 5-6 yrs is 706 and age group 6-7 is 787. Thus it is apparent that the no. of words increases with increase in age.

Under a project, the Ministry of Education in 1964 entrusted the work of collecting linguistic information of a statistical nature in Hindi to Deccan College, Pune, with the objective of finding out the relative frequency of the occurrence of words, syllables, phonemes and morphemes in spoken and written Hindi of today. They collected 100,000 words from all sections of modern Hindi, written and printed (newspapers and periodicals, light literature, scientific literature and movie scripts). Under the same project a list of words and their frequencies have been provided in their book.[14] However, the results of the present study cannot be compared with the aforementioned study of CIIL as the material/data collected was from scripts and not from spoken language.

Statistical lexical distribution in electronic media speech for comparative analysis of the vocabularies of Urdu and Hindi was undertaken in 2003. The author[4] has listed the 100 most frequent words for Hindi and Urdu used in the electronic media. The 10 most frequent words for Hindi are listed below: ka - frequency of occurrence- 43098, hona-24149, me - 15141, ne - 10533, kama-9346, ko-9220, se - 8892, jana - 8134, ki - 8029 and yaha - 7329. This appears to be similar to the present study in terms of counting the frequency of occurrence of words. However, comparisons cannot be made as the data collected is from print media and not from spoken language. Basic statistics like word frequency, syllable frequency, word length, sentence length etc was used to compare the corpora of ten Indian languages.[15] In this experiment, CIIL corpus was used for comparative analysis among the ten languages. Also, the percentage distribution of top five high frequency words in the corpus is included in this study, and it is as listed for Hindi language:- ke -- 3.59%, he –3.08%, me – 2.79%, ki—2.355%, se—1.70%. As seen from the aforementioned study postpositions appear to be the most frequently occurring class of words. A similar trend is also seen in the present study. Additionally, the present study also gives the frequency of occurrence of various words across four age groups.

c) Class of words and frequency of their occurrence: The following is observed from Table 3:

Nouns: Nouns are the most frequently occurring class of words. There is marked increase in number of nouns between two groups 3-4yrs and 4-5yrs (i.e. average age of 4 years), there is another spurt of growth of nouns between 5-6yrs and 6-7yrs age group (i.e. average age of 6 years).

Verbs: Verbs are the next frequently occurring class of words. In this study, a marked increase in the number of verbs is observed between 3-4yrs and 4-5yrs of age (i.e. average age of 4 years).

Auxillary: Auxiliaries are the third frequently occurring class of words. A marked increase in auxiliaries is seen at 5yrs of age.

Affirmatives

Affirmatives are the fourth frequently occurring class of words. A marked increase in auxiliaries is seen at 4yrs and 6yrs of age.

From Figure 1 and Figure 2 the following is observed:

1. For six class of words out of a total of ten, the frequencies of words increase systematically with increasing age. These six classes of words are adverbs, affirmative, auxiliary, nouns, post positions and verbs.

2. For two classes of words viz. adjectives and pronouns though the frequencies of words is highest for the age group D (i.e. 6 to 7 yrs) and lowest for age group A (i.e. 3 to 4 yrs), the second highest is the age group B (i.e. 4 to 5 yrs) and not the age group C (i.e. 5 to 6 yrs).

3. For the class of word conjunction the frequencies for age group C (i.e. 5 to 6 yrs) are highest followed by that for age group D (i.e. 6 to 7 yrs). The last two age groups in terms of frequencies are B (i.e. 4 to 5 yrs) and A (i.e.3 to 4 yrs), in that order.

4. For only the class of words 'preposition', the frequencies across the age groups do not increase systematically with age. The frequency is highest for age group B (i.e.4 to 5 yrs) and the second highest is age group C (i.e.5 to 6 yrs), which is followed by the age group D (i.e.6 to 7 yrs). The frequency is however, lowest for age group A (i.e. 3 to 4 yrs).

From Figure 1 and Figure 2, it is also observed that the pattern of frequencies across majority of the words is constant, i.e. it shows increase from lowest age group to highest age group. This is reflected through the results of the Chi-square test. The value of Chi-square is statistically significant ($\text{Chi-square}_{(27)}$) = 469.90, $p < .0005$). This shows that the variable "class of words" depends on the variable "age group".

The frequency of words increases with the size of vocabulary between the ages of 1;6 and 2;6 and enables toddlers to form grammatically more complex utterances.[16] Vocabulary development in Greek children was studied and authors [17] reported that vocabulary size increased markedly with age. The results of the present

study are in agreement with the aforementioned studies. It is also observed that for all other class of words, frequency of occurrence is least for the younger age group (3-4 yrs) compared to the older age group (6-7 yrs).

3. Appropriate parametric / non-parametric tests to compare differences between the groups- Comparisons between different Classes of Words:

Comparisons were done between the following: There is a dearth of studies studying comparisons between different classes of words and their frequency of occurrence across all four age groups.

The conclusions of the present study can be listed as following:

Results obtained after analyzing class of words by calculating type token ratio: Percentage of nouns is highest across all age groups followed by verbs, pronouns, adjectives. The percentage of other class of words is very less. There is increase in the percentage of tokens across age groups.

Nouns occur most frequently than other class of words. They are followed by verbs, adjectives, pronouns and prepositions. The other classes of words are few in number. At higher ages there appears to be marked increase in nouns, verbs, adjectives and pronouns and not so in other class of words. Frequency of occurrence of words increases with increase in age. The common words with high frequency of occurrence are auxiliary hɛ̃, hũ, pronoun je, verbs rʌhe, rʌha, rʌhi, dʒa, conjunction ɔr, Nouns k^hel, gaḍi, log, post positions pe, ke. There appears to be marked increase in different classes of words, one at 4 yrs of age (after Sr. KG) and other at 6 yrs of age (standard I).

Comparing differences between age groups for different classes of words and their frequency of occurrence- Pattern of frequency of occurrence across majority of the words is constant, i.e. it shows an increase from lowest age group to highest age group. The value of Chi-square is statistically significant showing that the variable "class of words" depends on the variable "age group."

From the review of literature, it is apparent that there is a large body of research into various components of Hindi language, including semantics and syntax. However, there appears to be a dearth of studies focusing on description and analyses of semantics of Hindi speaking normally developing pre-school and school going children. This study appears to be amongst one of the recent studies looking into the aforementioned aspects. One of the highlighting features of this study is the huge database of semantics and syntax (of spoken language) collected from 200 pre-school and school going children. For our field there appears to be a need to develop standardized tests in Hindi for assessing the language of the disordered population. Creating such a database and utilizing it for the aforementioned purpose appears to be the need of the hour.

ACKNOWLEDGEMENTS

The authors are grateful to Mr. R. Rangasayee, Ex – Director of AYJNIHH, Mumbai for his support and Dr Ashok Kumar Sinha, Director, AYJNIHH for the encouragement provided.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. Berry MF. In: Language Disorders of Children. Eds. New York NY: Appleton-Century-Crofts; 1969.
2. Kaushik JN. In: Basic Hindi Vocabulary. New Delhi: Lipi Prakashan; 1979.
3. Bhanushali NG. An investigation into the basic Hindi vocabulary of Hindi speaking children of class V in Greater Bombay. In: Research in Language Education a Trend Report: D.P. Pattanayak; 1985. Available at http://www.education.nic.in. Accessed 17 October 2011.
4. Martynyuk S. Statistical approach to the debate on Urdu and Hindi. In: The Annual of Urdu studies. 2003:8.
5. Ojha S. A Test for Word Finding Skills in Hindi. In: Unpublished Masters' dissertation, University of Mysore; 1992.
6. Kiran J. Test of Writing for Children in Hindi. In: Unpublished Masters' dissertation, Mangalore University; 1994.
7. Kumar M. Boston Naming Test in Hindi. In: Unpublished Masters' dissertation, Mangalore University; 2002.
8. Olness GS, Ulatowska HK, Wertz RT, Thompson J L, Auther-Steffan LL. Discourse elicitation with pictorial stimuli in African Americans and Caucasians with and without aphasia. Aphasiology. 2002;16:623-33.
9. Ghosh D, Bhat S, Athaide D, Shah A, Kant A. Analyses of semantic, syntactic and pragmatic characteristics of normally developing hindi speaking children. Paper presented at 42nd Conference of the Indian Speech and Hearing Association; 2010.
10. Biber, Johansson, Leech, Conrad and Finegan. Longman Grammar of Spoken and Written English (LGSWE). London: Longman, Hardcover;1999.
11. Marjanovič-Umek L, Fekonjapeklaj U, Podlesek A. Characteristics of early vocabulary and grammar development in Slovenian-speaking infants and toddlers: a CDI-adaptation study. Journal of Child Language, First view article. 2012.

12. Vedhasorubini K, Chengappa KS. Manual for enhancement of syntax in Tamil for children with language impairments. Language in India, 12; 2012.
13. Budwig N, Narasimhan B, Srivastava S. Interim solutions: The acquisition of early verb constructions in Hindi. In: E. V. Clark and B. Kelly, eds. Constructions in acquisition. Stanford, CA: CSLI Publications;2006:163-183.
14. Ghatage AM. Phonemic and Morphemic frequencies in Hindi. Pune: Poona University and Deccan college publications in Linguistics; 1964.
15. Bharati A, Sangal R. Basic Statistical Analysis of Corpus and Cross Comparison among Corpora, 2002. Available at http://ltrc.iiit.ac.in/MachineTrans/publications/technicalReports/tr022/camera-187.pdf. Accessed 18 February 2013.
16. Conboy BT, Thal DJ. Ties between lexicon and grammar: Cross-sectional and longitudinal studies of bilingual toddlers. Child Development 2006; 77(3):712–35.
17. Papaeliou CF, Rescorla LA. Vocabulary development in Greek children: A cross-linguistic comparison using the language development survey. J Child Lang. 2011;38(4).

Cite this article as: Kant AR, Dafadar BS, Banik AA. Analysis of characteristics of semantics of spoken language in normally developing Hindi speaking children. Int J Res Med Sci. 2015;3:3534-42.