INITIAL LEXICAL ACQUISITION AND NOUN BIAS HYPOTHESIS VERIFICATION

Aquisição lexical inicial e verificação da hipótese do viés nominal

Letícia Arruda Nóro(1), Débora Durigon da Silva(1), Fernanda Marafiga Wiethan(2), Helena Bolli Mota(2)

ABSTRACT

Purpose: verifying how the initial lexical acquisition occurs in children with typical development, regarding to types and tokens of the lexical items. Furthermore, one wants to verify if the noun bias hypothesis occurs, and in what version, strong or weak. Methods: the sample consisted of 20 children, male and female, with typical language development. This research covered ages from 1:0 to 1:11 (years: months) divided in three age groups (1:0 – 1:3;29, 1:4 – 1:7;29, 1:8 – 1:11;29). Audio data from spontaneous speech were collected, and after, lexical analysis was performed regarding to types and tokens produced. The Statistical tests Mann Whitney; Kruskal – Wallis and Wilcoxon were used, with significance level p< 0.05. Results: no statistical significance was found to the variables regarding to sex. However, statistical difference was found between the age group 1 in relation to 2 and 3 to the majority of variables. Furthermore, one verified prevalence of content words in the age groups 2 and 3. The prevalence of nouns over verbs in all age groups was observed. Conclusion: the initial lexical acquisition in children with typical development occurs gradually according to the increase of age. In this period the sex variable doesn’t influence in the linguistic performance. Furthermore, the noun bias hypothesis was confirmed in its weak version, confirming the thesis that inspired this research.

KEYWORDS: Child; Language Development; Vocabulary

INTRODUCTION

In order to exist a communication between people it is necessary some sort of language. Therefore, when there is not an organic or psychic impediment for it, oral language is used. In this manner, the communication ability is one of humans differential traits, presenting different complexity levels, also being likely to irregular inadequacies and productions, which can be or not significant to the speech intelligibility. Language is composed by the lexical heap still acquired as a child, these initial acquisition steps. This communicative ability can be explained through the individual capability and fulfillment in receiving, elaborating and transmitting messages as long as they have an informative content1,2.

The lexical acquisition is one of the first remarkable manifestations in the language development and it is related to the comprehension capacity, also to the production of several types of meanings3,4. In this complementary manner, the vocabulary is the initial feature that distinguishes the acquisition from a specific language and allows the analysis plus the entire development of the features: phonology, morphology, syntax, pragmatics, semantics and fluency5.

In this study the term lexicon was chosen instead of the term vocabulary, because the first refers to lexical items inserted in the speech and the second refers to the language terms, which can be considered separately. Besides this, it is not intended to just consider nouns, verbs and adjectives, like in another studies about the theme3,6,7, but as well as other elements that compounds the

---

(1) Universidade Federal de Santa Maria – UFSM, Santa Maria, RS, Brasil.
(2) Programa de Pós-graduação em Distúrbios da Comunicação Humana da Universidade Federal de Santa Maria – UFSM, Santa Maria, RS, Brasil.

Source of research support: CAPES, CNPq

This research was performed at Language and speech study center (Centro de estudos em linguagem e fala – CELF) – Speech Pathology and Audiology graduation course; Post-graduation program in Human Communication Disorders, Federal University of Santa Maria -UFSM – Santa Maria, RS, Brazil.

Conflict of interest: non-existent
language grammar, such as, pronouns, conjunctions, prepositions, numerals, articles, interjections and adverbs apart from the ones already mentioned, nouns, verbs and adjectives

The lexicon is a phenomenon in continuous growth as more knowledge is acquired, it is an open system, in constant improvement and enlargement. The contact between people, as a group, in society, at work and in several settings that offer human communication, also lead to their lexical heap increase, through an individual and heterogenic process.

Still, lexicon is defined as a unit set, without the origin in the grammatical rules, but in the internal language. From what it was referred to, it becomes comprehensible the difficulty in lexical analysis accomplishment, through cultural implicatures or through the dynamic and mutant characteristics of the language observations, that seeks to follow the communication needs. The lexicon is dynamic and precise, it also results from the settings we attend to. So there it can follow the several new nomenclatures, new objects, new situations that happen in our peculiar quotidian; the gradual heap increase oh each person is necessary.

The first items from this heap appear when the child is about one year old. This universal phenomenon is explained by the fact that in this age the child reaches certain neuropsychological maturity. The beginning of the standard lexical acquisition easily happens through recognition and word repetition that are similar in its phonology, followed by a fast increase in word number, which is characterized by the vocabulary explosion when the child is around 18 months old. This would explain itself by the initial codification system and attributions of the characteristics. When the child is around 2 years old, it is noticed an acquisition of 50 to 600 words in a speed of 10 words a day.

To check the lexical variety or the different varieties of spoken words by the child, a calculus is done through the rate in relation to the type/token – number of the several lexical items produced, divided by the total of lexical items, in other words, it is a measure of the linguistic production to estimate the lexical proficiency. In this way the type (kind) is each different lexical item spoken by the child and token (occurrences) refer to the repetitions of each type in the same talk.

Still the examined variables in the lexical study, some papers propose the noun bias hypothesis, in which the names (nouns) are the word prevalent categories during the initial lexical acquisition. There are two versions from this hypothesis, the first being the strongest version, referring the noun acquisition, then the verb acquisition and then the remaining parts of speech; the second and weaker version in which the names appear simultaneously to the verbs, however, still in a high-priority manner.

Based on what was exposed, the objective of this article is to verify in to what extent the initial lexical acquisition happens in children with typical development, in terms of types and tokens plus the lexical items occurrences. Besides this, it seeks to verify if the noun bias hypothesis really occurs, and in which version, stronger or weaker, related to the age group.

**METHODS**

The current research is from a transversal and quantitative nature and it is attached to a previously approved project by the Research Ethics Committee from the Federal University of Santa Maria under the registration 0219.0.243.000-11. The sample was composed by 20 children from both genders, with a typical language development, Portuguese speakers from the south, no bilingual record and from a low economical class. The number of individuals was obtained through a calculus sample over the enrolment number from the child education in public child schools from three regions of Santa Maria – RS. This study covered the age group of 1:0 to 1:11 (years old: months) resulting in a sample number of 20 children. Whereas the sample was divided as it follows: three age groups and genders. In age group 1 are 1:0 to 1:3; 29; in age group 2 are 1:4 to 1:7; 29 and in age group 3 are 1:8 to 1:11; 29. The first two age groups were composed by six subjects and the last one by eight subjects.

The guardians for these subjects accepted participating in the research after they received a complete explanation about the research nature, its procedures, risks, benefits and secrecy about their identities. After everything, they signed a Free and Clarified Consent Term and filled a questionnaire involving pre and postnatal exams.

The inclusion criteria adopted for the subjects of the study participation were the following: to be between 1:0 and 1:11; 29 days; be a member of a Portuguese speaking family; present a typical language development of both genders, and to be from a low economical class.

The established exclusion criteria were: to present any level of hearing loss; neurological, emotional and/or cognitive limitation; the presence of alteration in the motor or organic origin; have done speech therapy, or be doing during the research; to present speech alterations that damage the language and speech development.

The subjects from the sample did the following evaluations: Behavioral Observation Protocol.
bias hypothesis really occurs and if it presents the stronger or weaker version, besides comparing the data found with other studies about the theme, in national and international literature.

The study data were submitted to Mann Whitney statistical analysis; Kruskal – Wallis and Wilcoxon. The significance level adopted for the statistical tests was 5% (p<0.05).

RESULTS

Table 1 presents the analysis between the linguistic variables in each subject productions in relation to gender, through comparative averages. There was not a significant difference to none of the variables.

Table 2 exposes the results as far as the comparisons between the age groups through the Kruskal – Wallis statistical test. It was verified that there is a significant statistical difference between the age group 1 in relation to age group 2 and 3 for most of the variables.

Table 3 presents the grammatical word analysis and the content produced in each age group, through comparative average. It was checked the predominance of the content words in the age groups 2 and 3.

Table 4 presents the analysis with the noun and verb class comparisons in each age group. It was checked the noun predominance over the verbs in all age groups.

Table 1 – Numerical variable average between genders

<table>
<thead>
<tr>
<th>Variables</th>
<th>Male</th>
<th>Female</th>
<th>P* Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Types produced</td>
<td>23,33</td>
<td>21,57</td>
<td>P=0,563</td>
</tr>
<tr>
<td>Tokens produced</td>
<td>34,00</td>
<td>33,71</td>
<td>P=0,773</td>
</tr>
<tr>
<td>Grammatical Words</td>
<td>7,50</td>
<td>8,71</td>
<td>P=0,901</td>
</tr>
<tr>
<td>Nouns</td>
<td>21,33</td>
<td>21,43</td>
<td>P=0,869</td>
</tr>
<tr>
<td>Verbs</td>
<td>5,00</td>
<td>3,57</td>
<td>P=0,380</td>
</tr>
<tr>
<td>Content Words</td>
<td>26,50</td>
<td>25,0</td>
<td>P=0,772</td>
</tr>
</tbody>
</table>

* Statistical test used: Mann Whitney. Considered statistically significant, p<0.05 value.
Table 2 - Numerical variable average between age groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>Age Group 1 (1:0 to 1:3;29)</th>
<th>Age Group 2 (1:4 to 1:7;29)</th>
<th>Age Group 3 (1:8 to 1:11;29)</th>
<th>P* Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Types produced</td>
<td>7,00a</td>
<td>28,00b</td>
<td>29,00b</td>
<td>P=0,037*</td>
</tr>
<tr>
<td>Tokens produced</td>
<td>8,17a</td>
<td>43,33b</td>
<td>45,88b</td>
<td>P=0,023*</td>
</tr>
<tr>
<td>Grammatical Words</td>
<td>2,00a</td>
<td>10,67b</td>
<td>11,38b</td>
<td>P=0,017*</td>
</tr>
<tr>
<td>Nouns</td>
<td>5,17a</td>
<td>28,17b</td>
<td>28,50b</td>
<td>P=0,024*</td>
</tr>
<tr>
<td>Verbs</td>
<td>1,00a</td>
<td>4,50b</td>
<td>5,88a</td>
<td>P=0,175</td>
</tr>
<tr>
<td>Content Words</td>
<td>6,17a</td>
<td>32,67b</td>
<td>34,50b</td>
<td>P=0,030*</td>
</tr>
</tbody>
</table>

*Statistical test used: Kruskal – Wallis. Considered statistically significant, p< 0,05 value. Different letters overwritten between themselves, indicate that there is a statistical significance between the values.

Table 3 – Comparative analysis of grammatical and content words in each age group

<table>
<thead>
<tr>
<th>Age Groups</th>
<th>Grammatical Words</th>
<th>Content Words</th>
<th>P* Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age Group 1</td>
<td>2,00</td>
<td>6,17</td>
<td>P=0,093</td>
</tr>
<tr>
<td>Age Group 2</td>
<td>10,67</td>
<td>32,67</td>
<td>P=0,028*</td>
</tr>
<tr>
<td>Age Group 3</td>
<td>11,38</td>
<td>34,50</td>
<td>P=0,042*</td>
</tr>
</tbody>
</table>

* Statistical test used: Wilcoxon. Considered statistically significant, p< 0,05 value.

Table 4 - Comparative analysis of nouns and verbs in each age group

<table>
<thead>
<tr>
<th>Age Groups</th>
<th>Nouns</th>
<th>Verbs</th>
<th>P* Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age Group 1</td>
<td>5,17</td>
<td>1,00</td>
<td>P=0,027*</td>
</tr>
<tr>
<td>Age Group 2</td>
<td>28,17</td>
<td>4,50</td>
<td>P=0,028*</td>
</tr>
<tr>
<td>Age Group 3</td>
<td>28,50</td>
<td>5,88</td>
<td>P=0,012*</td>
</tr>
</tbody>
</table>

*Statistical test used: Wilcoxon. Considered statistically significant, p< 0,05 value.

DISCUSSION

In Table 1 it was possible to observe that there was not a significant statistical difference of the variables studied between boys and girls. These data confirm a study, which had as an objective to analyze the comparison in shifting between types and tokens also the type/token rate in children, from both genders, Brazilian Portuguese speakers, as far as the parts of speech plus the total and segmental measure. The study authors concluded the fact that there was not a difference between genders, which shows a balance in initial lexical acquisition between these two groups.

However, according with another study in which grammatical and lexical development measures were made including the average extension of the statement in relation to type/token, there is a gender variable influence in language acquisition. The study statistical result revealed a general gender effect, showing a small advantage in language production for the girls over the boys until 36 months of age. This difference between the studies can be related to the language, since the mentioned study was done with French children, same ages, for being premature.

Researching about lexical and morphological coda acquisition it was found that female gender as favoring the correct production. This fact reinforces the findings that highlights the female superiority in the tasks related to language and speech abilities. Nevertheless, the same did not occur in the current study, because there was not a variation between the genders, probably for being aimed to the initial lexical heap not containing the phonetic analyses.

In the nouns and verbs comparative analysis between the genders, the current study confirms with a research that demonstrates the relation between the nouns and verbs usage and their classification, in a spontaneous speech situation in preschool with typical language development. Likewise it was here...
found, the study concluded that the genders did not influence in the verb and noun production.

In the comparative analysis between the age groups, it was verified the difference statistically significant for: types, tokens, grammatical words, nouns, content words. This data agrees with authors24 who checked that the numbers of types and tokens occurrences of a language sample in a fixed extension increases due to age, being classified as “linguistic facility index”, in which it reflects several factors, such as speech maturation; to produce a minimal syntactic organization like a nominal and verbal phrase and even a possible clause which demand a higher syntactic and lexical knowledge, in other words, with the highest number of conjunctions, pronouns and articles, among others25.

As the child grows, his or her lexical heap increases. If, at the beginning of the analysis the child uses a reduced number of words, that belongs to a few parts of speech, with aging, the number of words increases also the variety of the parts of speech26.

International studies confirm that, after a small vocabulary growth approximately from 12 to 24 months of age, the child goes through a period called vocabulary explosion, demonstrating the age effect over the produced lexical items, in the same manner that occurred in the current study27-30.

Still related to the age group, the data found here confirm an international study in which it places the rising of grammatical words as slow20, since there was a certain content word predominance.

According to other study, the first children’s words could be stuck to the context, being produced only in limited or specific situations. This context is an event that occurs with certain regularity for the children, however, there are words contextually flexible that are used in a reference manner to indicate classes of objects, proper names, individualized objects, people/animals, or actions. Initially, words are acquired in a slow velocity (around one, two or three new words per week), the statements are reduced to one word each time. That explains the fact that in this study the significant differences are between the age group 1 and 2; also 1 and 3, not between 2 and 3, because these last two groups a bigger stability was verified31.

By analyzing the age group studied in relation to the parts of speech and content words it is possible to observe that the age groups 2 and 3 present statistical significance, something that did not occur in age group 1. This result can occur due to the reduced number of produced words by the children in age group 1. Still, it confirms a literature finding, in which the 18 month age group, the nouns are highest in the lexical set in this children’s group. In the age groups from 24 to 32 months of age, there is a verb tendency to match or overcome the nouns. And finally, around the 32 months of age, which in this study was not approached, the parts of speech and the content words should be more balanced19.

A study done31 with a sample of 8 months of age 2:6 also observed that nouns are predominant, fulfilling a 55% average in the children’s lexicon with a vocabulary between 100 to 200 words, while the content words were less than 15%.

Table 4 can confirm the noun bias hypothesis in its weaker version, because the analysis results showed the noun numbers was higher than the verb numbers during the study lexical acquisition period, however the noun production it was not exclusive even in this initial period of the language acquisition.

This contributes with the Natural Partitions Theory22, in which the noun prevalence in relation to verbs during the initial lexical acquisition is the result of a cognitive tendency, for being the first to be understood by the child, because the noun is more concrete than the verb. It is known, the verbs are relational terms, that refer to the most abstract concepts and less cohesive, therefore the limits that differentiate one verb from the other are less clear and harder in the acquisition33,34.

So, even though there are limitations with the reduced time in the sample speech recordings plus the fact that the child interacted with the examiner and not someone who his or her is used to, it is believed, this study can contribute with the speech clinic also with the early diagnose in language alterations in children from low socioeconomic class and for it to be considered in the therapeutic planning, an adequate lexical heap.

CONCLUSION

After the data analysis from this research, it was verified that the initial lexical acquisition in children with typical development happens in a progressive manner as the age increases and in this period the gender variable do not influence in the linguistic production.

The existence of the noun bias hypothesis was confirmed in its weaker version, agreeing with the thesis that justifies this research. Based on the results therapeutic sessions can improve, according to the words used by children in their initial language acquisition phase, helping the nomination techniques, used in therapy, for example. Besides, it is possible to early detect the risk of children developing language alterations and from this point on, perform strategies for prevention, guidance for the mothers and early stimulation.
From the results found, new researches are suggested about the theme, using a wider age group, comparing genders, analyzing children from different social classes, as well as from other cities.

ACKNOWLEDGEMENT

We would like to thank CNPq and CAPES for their support for the accomplishment of this research.

RESUMO

Objetivo: verificar de que modo se dá a aquisição lexical inicial de crianças com desenvolvimento típico, em termos de tipos e ocorrências dos itens lexicais e se a hipótese do viés nominal realmente ocorre, e em qual versão, forte ou fraca. Métodos: a amostra foi composta por 20 crianças de ambos os sexos, com desenvolvimento típico de linguagem. Este estudo abrangeu a faixa etária de 1:0 a 1:11 (ano : meses) dividida em três subfaixas (1:0 – 1:3;29, 1:4 – 1:7;29, 1:8 – 1:11;29). Foram realizadas gravações da fala espontânea de cada sujeito, e após, análise lexical quanto aos tipos e ocorrências dos itens lexicais produzidos. Empregou-se a estatística Mann Whitney; Kruskal – Wallis e Wilcoxon, com nível de significância p< 0.05. Resultados: não houve diferença estatística para nenhuma das variáveis em relação ao sexo, porém, há diferença entre as subfaixas etárias 1 em relação a 2 e a 3 para a maioria das variáveis, além disso, verificou-se o predomínio das palavras de conteúdo nas subfaixas etárias 2 e 3 e por fim, verificou-se o predomínio de substantivos sobre os verbos em todas as faixas etárias. Conclusão: a aquisição lexical inicial em crianças com desenvolvimento típico dá-se de forma progressiva de acordo com o aumento da faixa etária e neste período, a variável sexo não influencia na produção linguística. Além disso, a existência da hipótese do viés nominal foi confirmada em sua versão fraca, corroborando a tese que inspirou essa pesquisa.

DESCRITORES: Criança; Desenvolvimento da Linguagem; Vocabulário

REFERENCES


Received on: March 10, 2013
Accepted on: June 07, 2013

Mailing address:
Letícia Arruda Nóro
Rua Felipe de Oliveira, 601/201 – Centro
Santa Maria – RS – Brasil
CEP: 97015-250
E-mail: titamecmat@gmail.com

Rev. CEFAC. 2015; 17(Supl1):52-58