ABSTRACT
We investigate 13-month-old Brazilian Portuguese infants’ ability to segment phrases and categorize words using phonologically strong function words. Using a visual fixation procedure, we conducted two experiments. In Experiment 1, infants were familiarized with two pseudo-nouns (tofe, bape). In the test phase, one group heard bape + real determiners and tofe + pseudo-determiners and the other group heard the reverse pattern. Infants segmented the DPs, preferring the ones containing real determiners. In the second experiment, infants heard determiners (Group 1) or subject pronouns (Group 2) + pseudo-words. All infants were tested with non-familiarized determiner or pronoun + pseudo words. Infants identified different functional items as belonging to determiner or pronoun classes and used this information to categorize novel words into noun or verb categories. These results suggest that, despite their phonological properties, BP function words are perceived early by infants and used in DP segmentation and word categorization.

KEYWORDS: word segmentation; word categorization; determiners; pronouns.

RESUMO
Investigamos a habilidade de bebês de 13 meses adquirindo o português brasileiro na segmentação e categorização de palavras, usando itens funcionais fonologicamente fortes. Utilizando a técnica do Olhar Preferencial, conduzimos dois experimentos. No Experimento 1, bebês foram familiarizados com dois pseudonomes (tofe, bape). No teste, um grupo ouviu bape + determinantes e tofe + pseudodeterminantes; o outro grupo ouviu o padrão inverso. As crianças segmentaram o DP, preferindo a condição contendo determinantes. No segundo experimento, bebês foram familiarizados com determinantes (Grupo 1) ou pronomes pessoais (Grupo 2) + pseudopalavras. No teste, todas ouviram determinante ou pronome não familiarizados + pseudopalavras. As crianças atribuíram os itens funcionais às diferentes classes D e Pron, e categorizaram novas palavras como nome ou verbo. Os resultados sugerem que, apesar de suas propriedades fonolóxicas, itens funcionais do PB são percebidos precocemente pelas crianças e usados na segmentação DP e na categorização de novas palavras.

PALAVRAS-CHAVE: segmentação de DP; categorização; determinantes; pronomes.

1. This research was funded by grants from CNPq and CAPES and was approved by the Federal Committee Ethics in Research (no. 100/2011). A preliminary version of this study was presented in the 39th BUCLD (Name, Uchôa & Teixeira, 2014).
1. INTRODUCTION

Language acquisition studies have emphasized the important role of functional elements on lexical and syntactic learning by young children. Functional elements belong to closed classes. They could be free morphemes, as articles and pronouns, or bound morphemes, as verb affixes. These items are very frequent in the speech and appear at the edges of prosodic phrases (Gout & Christophe, 2006). Moreover, according to the minimalist conception of language (Chomsky 1995-2007), they are constituted by a set of formal features that encode grammatical information related to reference (to entities and events) and to illocutionary force. Therefore, parametric variation would be restricted to the possible values that formal features may assume and would be manifested by morphophonological patterns of functional elements (Corrêa, 2009; 2011). Thus, the recognition of functional elements could allow infants to learn something about the lexicon and the syntax of their language.

Indeed, the phonological or prosodic bootstrapping hypothesis postulates that acoustic cues in the speech stream might be perceived by infants and might give them some information about the structure of the language they are acquiring (Gleitman et al., 1988; Morgan & Demuth, 1996; Christophe et al., 1997; 2008). Based on this hypothesis, Christophe and collaborators (1997; 2008) have proposed a model of the first stages of language acquisition in which early infants’ language processing rely on prosodic boundaries and function words.

Experimental evidence suggests that newborns and infants distinguish function from lexical words based on their phonological features (Shady, 1996; Shaffer et al., 1998; Shi, Werker & Morgan, 1999). In several languages, function words are phonologically weak, monosyllabic and unstressed (Shi, Morgan & Allopenna, 1998; Cutler, 1993), whereas lexical words are stressed and show a variable number of syllables.

However, in Brazilian Portuguese (BP) function words have up to 3 syllables and may be stressed. For example, definite (o (masc.) and a (fem.)) and masculine indefinite articles (um) are monosyllabic and unstressed, but feminine indefinite article and demonstrative pronouns have two or three syllables and are stressed (uma (a, fem.), esse, essa (this, masc. and fem.), aquele, aquela (that, masc. and fem.)). Subject pronouns are also stressed and have one or two syllables (eu, nós – 1st pers., sg. and pl., respectively; você, vocês – 2nd pers., sg. and pl., respectively; ele, ela; eles, elas – 3rd pers., masc. and fem., sg. and pl., respectively). Consequently, it could be difficult to BP learning infants to recognize these items and use them as cues on language acquisition if they had to rely only on their phonological features.

In this study, we seek to better understand the role of these properties of function words on language acquisition. More specifically, we ask whether phonologically strong function words (articles, demonstrative and subject pronouns) could help children on word segmentation and categorization, despite their phonological, lexical-like features. For this, we conducted two experiments in order to examine whether 13-month-old Brazilian infants are able to use function words as cues to segment Determiner Phrases (DP) and identify a new word as noun or verb based on its preceded function word.

2. Nominative pronouns would be a more appropriated label, since in Brazilian Portuguese these pronouns may also appear in object function (Eu vi ele – I saw him). However, as the literature on word categorization employs subject pronoun, we made use of the same term.
We chose this age since previous studies showed that German and French-Canadian infants around 14-month-olds are able to segment DPs and categorize nouns using function words (Höhle & Weissenborn, 2000; Höhle et al., 2004; Shi, Werker & Cutler, 2003; Shi & Melançon, 2010). If phonological properties of BP function words do not disrupt word segmentation and categorization, we predicted that Brazilian infants at 13 months will be successful in their tasks. However, if infants find it hard to distinguish function words from lexical words because of their phonological similarity, they should need more time to be sensitive to these items and then, they would fail to segment and categorize new words at this age.

Next section presents a brief characterization of function words on its phonological and grammatical aspects and reviews some studies focusing on infants’ segmentation and categorization abilities using these items.

2. FUNCTION WORDS ON LANGUAGE ACQUISITION

Function words are closed-class elements that belong to functional categories. Within the minimalist framework (Chomsky 1995-2007), they are constituted by a set of formal features that encode grammatical information related to reference (to entities and events) and to illocutionary force. As Corrêa pointed out, “formal features can be thought as representing in the lexicon (with direct effects on the syntax) what is systematic in a language – conceptual/intentional distinctions, logical relations, means of satisfying discourse demands are systematically expressed in word order patterns and in inflectional morphology” (Corrêa, 2009, p. 39).

Functional elements are characterized for expressing grammatical properties (for example, number, gender and person), showing morphological dependency and lack of descriptive content. Determiner (D) is one of functional categories, a class of function words that determine referential and quantitative properties of a noun (Radford, 1997) and do not appear recursively in the same phrase, such as articles (the), demonstratives (this), possessives (my) and quantifiers (every) (Adger, 2002; Radford, 2007). This point leaves open the fact that in some languages, such as Portuguese, Italian and Hungarian for example, demonstratives, possessives and/or quantifiers can co-occur with articles. This led some authors to propose more functional projections under DP (Carsten, 2000; Adger, 2002; Longobardi, 2004; Alexiadou et al., 2007). Nonetheless, for purpose of our experiments, we take BP demonstratives as determiners.

As regard to subject pronouns, Radford (1997) also considers them as determiners, due to the fact that they encode grammatical properties as person, number, gender and case. In addition, they refer to entities and can appear along with a noun (e.g. we psychologists don’t trust you linguists). However, this classification is controversial: it has been argued that subject pronouns’ phi-features are interpretable as the nouns’ ones, unlike the determiners’ ones (Cardinaletti, 1993; Cardinaletti & Starke, 1994). Furthermore, agreement between determiner and noun is established as a feature checking/sharing operation in a syntactic relation. In contrast, pronoun-noun agreement is established in a semantic relation of co-reference between these elements. Finally, as we will see below, determiners and subject pronouns seem to guide noun and verb categorization, respectively. This suggests that grammatical
features of these elements can be coded differently in the phonological interface, which would make infants identify them as distinct classes. For these reasons, we will consider subject pronouns as a separate class from determiners.

From a perceptual point of view, function words are different from lexical items because of their phonetic, phonological and distributional features in several languages. They tend to be monosyllabic and unstressed. Also, as closed-class elements, they are very frequent in sentences and can be predictable by the syntactic context, that is, they occur in specific syntactic positions and appear at the borders of prosodic units, showing a distributional pattern across languages (Morgan, Shi & Allopenna, 1996; Gout & Christophe, 2006). Because of these properties, these items become salient for children. Thus, we would expect to find an early sensitivity to closed-class elements.

Indeed, experimental evidence suggests that 1-to-3-day-old American newborns are sensitive to a form-based distinction between functional and lexical items when these elements are presented in separated lists (Shi, Werker & Morgan, 1999). At 11 months of age, American infants identify function words in continuous speech (Shady, 1996; Shafer et al., 1998); around 16 month-olds, they are sensitive to their syntactic positions in sentences.

With regard to determiners, studies testing infants acquiring French, English or German show that they become sensitive to higher frequent determiners of their native language from the age of 6-8-month-olds; this sensitivity increases around 11 months and low frequent determiners are also recognized. At 13 month-olds, infants are able to distinguish real determiners from slightly phonetically modified pseudo-determiners. They also use determiners to segment the DP (Höhle & Weissenborn, 2000, 2003; Shi, Werker & Cutler, 2003, 2006; Shi, Marquis & Gauthier, 2006; Shi et al., 2006; Hallé et al., 2008; Shi & Lepage, 2008). At 14 months, German and French learning children categorize novel words as nouns when these words are preceded by determiners, but fail to categorize subject pronoun preceded words as verbs (Höhle et al., 2004; Shi & Melançon, 2010).

Taken together, these experimental results show enhancement of infants’ abilities to perceive function words in continuous speech during their first year of age. Particularly, in relation to determiners, infants use them to segment DPs and categorize novel words as nouns. Although these elements show different features in tested languages (for example, function words in English are more reduced than in French and prosodically weaker than in German; some are bi-syllabic in German), the stimuli presented in the experiments used monosyllabic, unstressed function words. One could suppose that the task infants had to do (identify function words, segment DP or categorize novel words) was facilitated because of the contrast between the function word phonological features and the lexical word ones.

Brazilian Portuguese is a good case for testing the weight of contrastive phonological properties between function and lexical words on language acquisition. As we pointed above, function words have up to 3 syllables and may be stressed. Concerning to determiners, indefinite feminine article and demonstratives are trochaic and have two or three syllables. As regard to subject pronouns, except for the first person singular pronoun, the other ones are all trochaic and bi-syllabic. The majority of lexical items in BP show a trochaic stress pattern as well. Therefore, function word phonological form is similar to lexical word ones, which could difficult infants’ task to identify these items in continuous
speech and use them as cues to word segmentation and categorization. Nonetheless, as determiners are concerned, experimental evidence suggests 14-month-old Brazilian children’s recognition of these items, since they listened significantly longer to stories containing real BP determiners compared to the same stories with phonetically modified pseudo-words \(^3\) (Name & Corrêa, 2003).

In the present study we ask if 13-month-old Brazilian infants are sensitive to BP determiners and are able to use them as cues to segment continuous speech. Determiners were masculine articles and demonstratives formed by one, two or three syllables. Monosyllabic determiners were unstressed; 2- and 3-syllable determiners were trochaic. If phonological features play a crucial role on the identification of elements as functional or lexical items, infants acquiring Brazilian Portuguese would have trouble to represent determiners as closed-class items and to use them in word segmentation. On the other hand, if lexical item-like phonological features are less important than other cues, as frequency and syntactic position, infants would succeed in both tasks.

The second goal of this study was to investigate whether BP function words (determiners and subject pronouns) can allow 13-month-olds Brazilian infants to categorize a novel word as noun or verb according to which function word precedes it (a determiner or a pronoun, respectively). This presupposes that Brazilian infants already represent determiners and subject pronouns as distinct functional classes, despite their phonologically strong forms. Previous infant categorization studies showed that 14-month-old French or German learning infants succeeded only in noun categorization task using weak function words (Höhle et al., 2004; Shi & Melançon, 2010).

3. EXPERIMENTS

3.1 Experiment 1: Sensitivity to determiners and word segmentation

3.1.1 Methods

Participants
Sixteen 13-month-old Brazilian infants completed the experiment. These infants were divided into two familiarization groups. Another 8 infants were tested but their data were excluded from the analysis due to fussiness (4) and looking time too short (4).

Stimuli
Stimuli were four real BP determiners and five pseudo-determiners (see figure 1, below). The pseudo-determiners were segmentally but not prosodically different from the real determiners. Phonotactics and typical word shape were preserved. In addition, two nonsense, pseudo-nouns, bape and tofe, were created, which also followed phonotactics and typical word shape of BP. Determiner Phrases (DP) were created using a real determiner or a pseudo-determiner plus a pseudo-noun.

In addition, the same adjective, lindo (beautiful), followed all DPs. We opted for using a real adjective so ungrammatical condition trials would not be consisted of only pseudo-items.

\(^3\) This result does not mean that Brazilian younger babies were not sensitive to determiners. It remains to be tested experimentally. It is worth mentioning that 9 to 12-month-old Brazilian infants perceived alterations on morphophonological pattern of verbal affixes (Bagetti & Corrêa, 2011).
A native Brazilian female speaker recorded the stimuli in an infant-directed register. Determiners and pseudo-determiners were recorded in isolation as well as the pseudo-nouns presented during the habituation phase. The pseudo-nouns are also recorded together with the adjective. Using the software PRAAT (Boersma & Weenick, 2008), we created strings of determiner or pseudo-determiner + pseudo-noun + adjective to be presented in the test phase. A 50msec pause was inserted between determiner or pseudo-determiner and noun.

Procedure and design
Infants were tested individually in a visual fixation procedure. The infant was seated on the parent’s lap in front of a TV screen and a loudspeaker. Auditory linguistic stimuli were presented together with a visual display of blue-and-green waves. The parent wore headphones delivering masking music.

The visual fixation procedure consists of four phases. The pre-test phase acquainted the infant with the procedure. The posttest phase allowed us to determine whether the infant was on task throughout testing. A 16-second trial consisting of four unrelated utterances were presented in these phases. In familiarization phase infants listened to the pseudo-nouns bape and tofe in different melodic contours and in isolation during 2 minutes. After that, in the test phase, they were presented to 12 trials divided in two conditions (6 trials per condition): Grammatical condition, formed by determiners, pseudo-nouns and the adjective; and the Ungrammatical condition, in which pseudo-determiners were combined with pseudo-nouns and the same adjective. Thus, pseudo-nouns were in the middle of the phrase, between the (pseudo-)determiner and the adjective. So, it would be necessary to infants to segment the phrase in order to track the familiarized pseudo-nouns. Each trial consisted of four DP + adjective phrases separated by 1sec pause and lasted about 16sec. Each child participated of the two conditions.

One group of infants (Group 1) heard sequences of DPs consisting of a real determiner followed by bape and the adjective (o/um/esse/aquele + bape + lindo) (Grammatical condition) and “DP”s consisting of a pseudo-determiner followed by tofe and the adjective (one/or/ugue/ofupe + tofe + lindo) (Ungrammatical condition). Group 2 heard the reverse pattern: DPs consisting of a real determiner followed by tofe and the adjective (o/um/esse/aquele + tofe + lindo) (Grammatical condition) and “DP”s consisting of a pseudo-determiner followed by bape and the adjective (one/or/ugue/ofupe + bape + lindo) (Ungrammatical condition). The time of looking fixation was measured and, at the end, it was calculated the average looking time for each condition.

We expected that if the real determiners facilitate infants’ recognition of the following previously habituated pseudo-nouns, infants will listening longer to grammatical condition trials. In this case,
the word recognition presupposes that infants would be able to segment the phrase to track the familiarized pseudo-nouns between the determiner and the adjective.

**Results and discussion**

Each infant’s average looking times for the grammatical trials and for the ungrammatical trials during the test phase were calculated. The data were analyzed in a 2 x 2 analysis of variance (ANOVA), with Type of Determiner (real versus pseudo) as within-subject factor, and Group (Group 1 versus Group 2) as between-subject factor. The results showed a main effect of Type of Determiner, \( F(1, 14) = 17.8, p = .0008 \) (Grammatical condition: \( M = 9.18s, SE = 0.718s \); Ungrammatical condition: \( M = 7.88s, SE = 0.657s \)), and no effect of Group, \( F(1, 14) = 0.33, p = .57 \). These results are shown in figure 2. Follow-up paired t-tests assessed the performance of each group. For both groups, looking time was significantly longer for the Grammatical trials than for Ungrammatical trials (Group 1: \( t(1, 7) = 3.42, p = .011, M = 9.46s, SE = 1.2s \) versus \( M = 8.39s, SE = 1.08s \); Group 2: \( t(1, 7) = 2.88, p = .024, M = 8.9s, SE = 0.87 \) versus \( M = 7.37s, SE = 0.78s \)).

![Figure 2: Mean looking times and standard errors during Grammatical Condition trials and Ungrammatical Condition trials](image)

Our results show that infants were able to identify the familiarized pseudo-nouns easier when they were preceded by real BP determiners than when they were preceded by pseudo-determiners. Since both pseudo-nouns were familiar to infants and the only difference between the two conditions trials was the kind of determiner (real versus pseudo), we may conclude that infants recognized the real determiners of their native language. Besides, it is reasonable to interpret these results as a consequence of infants’ ability to segment the phrases containing a real determiner in order to track the familiarized pseudo-noun. We could think that infants listened longer to grammatical trials simply because they had recognized the real determiners in these streams. However, trials in both conditions also contained a frequent adjective which is already familiar to infants at this age. So, these results could not be explained by a familiarity effect. In the next experiment we went one step further and investigated infants’ capacity to use phonologically strong function words (determiners and subject pronouns) as cues to categorize a novel word as noun or verb.
3.2 Experiment 2: noun and verb categorization

3.2.1 Method

Participants
Sixteen 13-month-old Brazilian infants completed the experiment. These infants were divided into two familiarization groups (Determiner or Pronoun). Another 3 infants were tested but their data were excluded from the analysis due to fussiness (2) and technical problems (1).

Stimuli
Stimuli were three feminine determiners – a (the), uma (a/an) and essa (this), three personal pronouns – ele (he), ela (she) and você (you) and two nonsense words – piva and dema. This nonsense, pseudo-words were created preserving phonotactic and typical word shape patterns of BP nouns and verbs. A-ending nouns are typically feminine gender nouns; -a verbal suffix marks 2\textsuperscript{nd} and 3\textsuperscript{rd} persons of present tense in first verb class (infinitive ar-ending verbs, the most productive verbal class). A native Brazilian female speaker recorded the stimuli in an infant-directed register. Determiners and pseudo-words were recorded in isolation. Using the software PRAAT (Boersma & Weenick, 2008), we created strings of determiner + nonsense word and subject pronoun + nonsense word presented in familiarization and test phases. A 50msec pause was inserted between determiner or pronoun and noun. Two of the determiners (a, uma) and two of the pronouns (ele, ela) were presented to infants during familiarization phase, the remainder determiner (essa) and pronoun (você) were used in the test phase.

Procedure and design
Infants were tested individually in a visual fixation procedure. In the familiarization phase, each group of children listened to stimuli during 2 min. Infants of Group 1 (Determiner) listened to sequences of DPs consisting of a determiner followed by a pseudo-word (a/uma piva; a/uma dema). Infants of Group 2 (Pronoun) listened to sequences of phrases consisting of a subject pronoun preceding a pseudo-word (ele/ela piva; ele/ela dema). During the test phase, both groups were presented to 12 trials divided in two conditions (6 trials per condition): Determiner condition, formed by a new, non familiarized determiner (essa, this) + pseudo-word (dema/piva); Pronoun condition, formed by a new, non familiarized pronoun (você, you) + pseudo-word (dema/piva). Trials of Determiner condition were grammatical (Same Category) for infants of Group 1 and ungrammatical (Different Category) for infants of Group 2. Conversely, trials of Pronoun condition were grammatical (Same Category) for infants of Group 2 and ungrammatical (Different Category) for infants of Group 1. Figure 3 summarizes stimuli presentation in familiarization and test phases.
The underlying hypotheses are that (i) 13-month-old infants identify determiners and pronouns as distinct classes within the category of functional items, and (ii) this information helps them to categorize new words as nouns or verbs. Therefore, we predict that after being familiarized to DPs, infants of Group 1 would show differential looking time responses to the test trials containing the new determiner *essa* versus the test trials presenting the pronoun *você*. On the other hand, infants of Group 2 should show differential looking time responses to the test trials in the reverse direction after being familiarized to pronouns preceding the pseudo-words.

**Results and discussion**

Each infant’s average looking times for the Same Category trials and for the Different Category trials during the test phase were calculated. The data were analyzed in a 2 x 2 analysis of variance (ANOVA), with Test Category (same versus different) as within-subject factor, and Familiarization (D + Noun versus Pron + V) as between-subject factor.

The results showed an effect of Test Category, $F(1, 14) = 7.63, p = .015$, and no effect of Familiarization, $F(1, 14) = 0.004, p = .947$. Follow-up paired t-tests assessed the performance of each familiarization group. For both groups, looking time was significantly longer for the Same Category trials than for Different Category trials (Group 1, Determiner: $t(1, 7) = 2.469, p = .042, M = 9.36s, SE = 1.35s$ versus $M = 7.27s, SE = 1.4s$; Group 2, Pronoun: $t(1, 7) = 2.26, p = .058, M = 9.09s, SE = 1.92s$ versus $M = 7.26s, SE = 1.6s$). These results in Figure 4 agree with our predictions.
Our results show that infants’ reaction to the test trials was related to the kind of function word they were exposed to during familiarization phase. Infants of the Determiner Group treated the new determiner in the Same Category test trials as the same kind of items as the determiners in the familiarization phase. Similarly, infants of the Pronoun Group considered the new pronoun in the Same Category test trials as the same kind of items as the pronouns heard in the familiarization phase. In other words, infants generalized determiners to one same class and pronouns to another, different class. Moreover, our results suggest that infants were able to categorize novel words to nouns or verbs when preceded by a determiner or a subject pronoun, respectively. This study goes beyond previous research in word categorization in that we showed noun and also verb categorization by infants as young as 13-month-olds.

Although determiners and subject pronouns are phonologically strong in Brazilian Portuguese, infants at 13-month-olds recognize them as different from lexical items. More than that, they represent them in different classes of functional elements. It allows infants to assign novel words to noun and verb categories.

4. CONCLUSION

Our study showed experimental evidence that phonologically strong forms of functional items do not prevent 13 month-old Brazilian infants to consider them different from lexical ones. They succeeded to represent determiners and subject pronouns as distinct closed-class items as well as to categorize novel words as noun or verb according to their preceding function word. Contrastive phonological properties between function and lexical words do not seem to play a crucial role on Brazilian Portuguese; frequency and syntactic position may be more robust cues to infants in order to track function words.
To conclude, we would like to emphasize that our findings are consistent with the hypothesis that functional elements play an important role on language development, as proposed by Christophe et al. (1997; 2008) model. Regardless of their lexical-like word forms, function words are very frequent in BP, have a predictable syntactic context and tend to appear at the edges of prosodic units. Infants may pay attention to these properties to identify these items. Then, they might rely on function words and prosodic boundary clues to a preliminary processing of syntactic structure. Indeed, using determiners and morpheme-like word markers, Laguardia et al. (2015) showed that 11-month-old infants learned non-adjacent dependency patterns only when these elements were both aligned with phonological phrase boundaries.

We also would accentuate that our findings are compatible with language acquisition proposals based on minimalist assumptions that morphophonological patterns of functional elements express the possible values that formal features may assume in a language. In that case, function word recognition might bootstrap the computational system operation and allow infants to fix the specific values that formal features assume in the language they are acquiring (Corrêa, 2009; 2011).

REFERENCES


