

# CODIFICAÇÃO DA MORFOLOGIA DO PB E ANÁLISE DA FALA DIRIGIDA À CRIANÇA

## MORPHOLOGICAL CODING OF BRAZILIAN PORTUGUESE AND CHILD DIRECTED SPEECH ANALYSIS

Leonor Seliar-Cabral<sup>1</sup>

Universidade Federal de Santa Catarina - UFSC

### Resumo

Neste artigo, relataremos o que já foi realizado pelo *Grupo de Pesquisa Produtividade linguística emergente* no projeto **Codificação da morfologia do PB e análise da fala dirigida à criança**, a fundamentação teórica e respectiva metodologia. Acreditamos que seja de particular interesse para os estudiosos de aquisição da linguagem e dos sociolinguistas que investigam registros de fala ter à disposição a gramática automática da morfologia do PB que lhes permitirá, após homogeneização de seus *corpora*, a rápida codificação da morfologia, para subsequente análise estatística e qualitativa. Explicamos, em detalhes, os procedimentos para a depreensão da gramática, de acordo com o programa CLAN.

**Palavras-chave:** Codificação. Morfologia do PB. Fala dirigida à criança.

### Abstract

A report on what has been done by the project **Morphological coding of Brazilian Portuguese**, altogether with its theoretical referential and respective methodology is presented. This project is sponsored by the Researching Group “Emergent Linguistic Productivity”, and is relevant for language acquisition researchers and sociolinguists interested on speech registers, since an automatic grammar for morphological analysis will be available for statistic and qualitative analysis of examining *corpora*. Procedures for building the automatic grammar with the help of CLAN program are extensively described.

**Keywords:** CLAN; automatic grammar; registers; morphology

## 1 HISTÓRICO

O grupo de pesquisa, até o momento, realizou a segmentação dos enunciados da fase 1 (20 meses e 21 dias) e da fase 2 (22m e 21d), do sujeito Pá, com a respectiva colocação dos *bullets* para que o enunciado possa ser ouvido pelos pesquisadores e cotejado com a transcrição fonética dos enunciados da criança; em adendo, foi feito o levantamento automático dos *type/tokens* dos itens lexicais e a codificação da linha morfológica dos enunciados da criança (com a tradução para o inglês, da 1ª e da 2ª fases), conforme exemplo a seguir: %mor: v|bot-TV1&PFV=put, em que v é o verbo, seguido do radical, da vogal temática da 1ª conjugação, do expoente aspectual acabado e da

---

<sup>1</sup> Profa. Emeritus

tradução. O principal achado da pesquisa foi considerar o acento de intensidade como morfema verbal (suprafixo), com a função de assinalar na 1ª fase diferenças aspectuais (posteriormente, redundante e cumulativamente também assinalará tempo/modo). Propusemos, então, a implementação da fórmula de Mattoso Camara Jr. (2004:134) que passa a: **T(R+VT) + SF (SMTA+SNP +SPF)**. Também foi levantado o paradigma verbal da 1ª e 2ª fases. A seguir, passou-se à homogeneização do *corpus* da 3ª fase do sujeito Pá, no arquivo PAU003 (24;20 dias), corrigindo a notação da superposição de falas e partindo enunciados muito longos, a partir das instruções do coordenador geral do banco mundial de dados de aquisição da linguagem, MacWhinney, conforme amostra abaixo:

@New Episode

@Situation: na sala, brincando

\*MOT: vem xxx no seu quarto para a Leonor ver seus brinquedos ?  
 \*INV: vamos lá [x 2]  
 \*MOT: vamos lá no seu quarto (.) vamos ?  
 \*CHI: lâ [: lá] dent(r)o .  
 %pho: 'lã 'dētũ  
 \*MOT: ahn@i ?  
 \*CHI: vamos depoi(s) .  
 %pho: 'vãmuz de'poj  
 %act: Paulo está entrando no quarto e acendendo a luz=Paulo is coming\_in the room and switching\_on the lights  
 \*INV: não aguento o baixinho .  
 %par: rindo=laughing  
 \*CHI: vamos acendi [: acende(r)] .  
 %pho: 'vãmuzasē'dʒi  
 \*MOT: (es)tá bom xxx [?] .  
 \*INV: vem (.) vamos sentar [?] ?  
 %com: overlapping nos três últimos enunciados=the three last utterances overlapped  
 \*MOT: xxx xxx xxx .  
 \*CHI: acende a luz .  
 %pho: 'sējŋ a 'lus  
 \*INV: ah@i (.) você acendeu a luz ?  
 \*CHI: (a)cendi .  
 %pho: sɪ'dʒi  
 \*INV: m@i .  
 %com: overlapping nos dois últimos enunciados=the two last utterances overlapped

*INV:	m@i .
*CHI:	fico(u) (es)t(r)agada .
%pho:	fi'ko tɐ'gadɐ
%com:	overlapping nos dois últimos enunciados=the two last utterances overlapped

Teve início a tabulação das classes sintáticas em três arquivos das falas dos adultos entre si, com a criança e da criança, pré-requisito para a comparação entre os registros dos falantes, dos quais damos exemplos, com os números dos enunciados onde os itens ocorrem:

- (1) **scat.adch.n.alfa**  
7 boca {[scat n]} =mouth  
1552, 1556, 1796, 10519, 10523, 10524, 10534
- (2) **scat.adad.n.alfa**  
1 pele {[scat n][gen fem]} =skin  
270
- (3) **scat.ch.n.alfa**  
1 luz {[scat n][gen fem]} =light  
7515  
1 zebrinha {[scat n][gen fem]} ="zebra-DIM-SG" =little zebra=

(Códigos para interpretação dos exemplos: scat = categoria sintática; sdch = adulto quando se dirige à criança; adad = adulto quando se dirige a adulto; ch = fala da criança; n = substantivo; alfa = em ordem alfabética; dígito = número de ocorrências do item; scat n = categoria sintática substantivo; dígitos = número do enunciado onde o item ocorre para recuperação do contexto).

Já se encontram prontas as tabulações, nos arquivos CLAN, dos substantivos, das preposições, dos pronomes adjetivos (det) e substantivos (pro) possessivos, com as respectivas traduções para o inglês, para depreensão da gramática automática do PB, conforme exemplos:

#### **Determinantes = Determinants**

essa {[scat det:dem]} "esse&FEM" =this=  
um {[scat det:art]} =a=

#### **Preposições = Prepositions**

com {[scat prep]} =with=  
entre {[scat prep]} =between=

#### **Contrações e combinações com preposições**

ao {[scat prep]} "a~det|o&MASC" =to the=

#### **Pronomes indefinidos = Indefinite Pronouns**

alguém {[scat pro:indef]} =someone=

#### **Pronomes interrogativos = Interrogative Pronouns**

quem {[scat pro:int]} =who=

**Pronomes pessoais = Personal Pronouns**

% forms that are the same as subject and object (in the last case, always preceded by preposition)

ela        {[scat pro:pers]} "ele&FEM=3S " =she;her=

**Pronomes relativos = Relative pronouns**

quem      {[scat pro:rel]} =who=

**Substantivos = Nouns**

adultos    {[scat n][gen masc]} "adulto&MASC-PL" =adults=

Os dados já tabulados na sua versão final, inclusive a audição em áudio, estão à disposição da comunidade científica internacional no *site*:

<http://childes.psy.cmu.edu/data/Romance/Portuguese/florianopolis.zip>

Para que se tenha uma ideia da magnitude do projeto, leve-se em conta que o mundo contemporâneo dos computadores e da linguística computacional tornou possível a catalogação e análise de uma quantidade antes nunca conhecida de dados da comunicação verbal, em tempo muito menor. Isso possibilita comparações e generalizações a partir de uma massa de dados muitíssimo mais robusta. A base de dados da plataforma CHILDES, com a qual o presente projeto opera, contém 44 milhões de palavras faladas em 28 línguas diferentes. Trata-se do maior *corpus* de fala atualmente existente. Em segundo lugar, vem o *British National Corpus*, com 5 milhões de palavras. Todos os dados do sistema CHILDES estão codificados de forma consistente num formato de transcrição denominado CHAT, inclusive os dados da 3ª fase do sujeito Pá, dos quais será depreendida a gramática automática, contendo mais de 8.500 enunciados.

Atualmente já foram construídas gramáticas MOR de 10 línguas: cantonês, holandês, inglês, francês, alemão, hebraico, japonês, italiano e espanhol, das quais servirão de modelo para a depreensão da gramática do PB as gramáticas do italiano e espanhol.

## 2 FUNDAMENTAÇÃO

A expansão do projeto **Codificação da morfologia do PB e análise da fala dirigida à criança** se deve à necessidade de aprofundar e de adaptar à morfologia do português do Brasil, os procedimentos de montagem das regras que compõem o aparato para a análise automática da morfologia, dentro da plataforma CLAN, conforme se observará no presente artigo.

Com efeito, no decorrer da última pesquisa, comparando-se a formalização das classes sintáticas e respectivas regras do espanhol e do italiano, que servem de modelo, com as já efetuadas para o PB, chegou-se à conclusão de que elas necessitavam ser refinadas e expandidas, com o concurso de especialista em linguística computacional, com sólida formação em linguística.

Persiste o nosso propósito de colocar à disposição da comunidade científica um programa que possibilite o processamento automático das unidades morfológicas do português do Brasil (PB) pelo programa CLAN do projeto CHILDES (MacWHINNEY, 2000b), porém, com as adaptações e expansões mencionadas acima, depreendido dos

enunciados do *corpus* da 3ª fase PAU003 (24;20 dias), já transcritos canonicamente, com os respectivos *bullets* e transcrição fonética, na linha %pho e dos enunciados dos adultos, produzidos pelos pais da criança, ambos falantes da variedade paulista (cidades de Campinas e São Paulo), o pai, linguista, com doutorado, e a mãe, psicóloga; pela investigadora, falante da variedade gaúcha (cidade de Porto Alegre), na época, em fase de doutorado em linguística e, esporadicamente, sobre alguns enunciados das empregadas. Já dispomos hoje, de quase todos os arquivos com as classes sintáticas, que permitirão a análise estatística e qualitativa, comparando os registros da fala dirigida à criança (*CDS*, isto é, *Child Directed Speech*), com a fala dos adultos entre si, bem como da própria fala da criança, com vários propósitos teóricos, tais como: 1) a investigação do *intake*, ou seja, o que, a criança consegue processar para fins de produção, procurando, a nível explanatório, justificá-lo por limites linguísticos e cognitivos; 2) comparação entre o registro da fala praticada entre os adultos e o da fala dirigida à criança, demonstrando, através das diferenças morfossintáticas, lexicais, semânticas, textuais e pragmáticas, a intuição que os adultos têm dos limites linguísticos e cognitivos impostos ao processamento verbal, os quais vão gradualmente diminuindo com o desenvolvimento da criança.

Com vistas a contribuir para o debate sobre como e quando as crianças adquirem a gramática (BITTNER; DRESSLER; KILANI-SCHOCH, 2003, p. 269), deter-nos-emos no aprofundamento e expansão do programa para o processamento automático das unidades morfológicas do PB (gramática do adulto) pelo programa CLAN do projeto CHILDES e comparação do registro de fala praticada entre os adultos e o da fala dirigida à criança.

A elaboração do programa para o processamento automático das unidades morfológicas do PB pelo programa CLAN do projeto CHILDES permitirá não só verificar o quanto a intuição dos adultos permite a adequação de seu registro ao desenvolvimento cognitivo e linguístico da criança e a possível influência da gramática do adulto na construção da gramática da criança, como também colocará à disposição de qualquer pesquisador que deseje investigar o PB um instrumento que lhe permitirá a análise automática dos morfemas.

No decurso das pesquisas até agora realizadas, foram evidenciados os seguintes achados:

- 1) as primeiras categorias verbais dominadas pela criança são aspecto (imperfectivo *versus* perfectivo) e modalidade (permissão, possibilidade e volição). A precedência da categoria aspectual sobre tempo confirma os achados de Bittner *et al.* (2005). A modalidade (permissão, possibilidade e volição), além da entoação e da negação, é produzida com os auxiliares po(d)e, qué e vô (nos dois primeiros já se observa a ocorrência da metafoia verbal que vai se mostrar produtiva na segunda fase, já com verbos regulares). Observem-se a oposição entre as vogais /a/ e /o/ acentuadas, assinalando a oposição entre imperfectivo (infinitivo) na 1ª conjugação (a mais frequente) e 3ª pessoa singular (forma não marcada) do perfectivo; e a oposição entre a vogal

temática acentuada (infinitivo e 3ª pessoa singular (forma não marcada) do perfectivo *versus* a vogal acentuada do radical (3ª pessoa singular (forma não marcada)) do imperfectivo (presente e imperativo, isto é, formas rizotônicas) nas três conjugações.

- 2) Em decorrência, propusemos a implementação da fórmula de JMC que passa a: **T(R+VT) + SF (SMTA+SNP +SPF)**, com a inclusão do suprafixo e da categoria de aspecto. Essa inclusão do acento de intensidade com a função de assinalar diferenças aspectuais tem sido negligenciada na literatura sobre aquisição da linguagem, apresentando problemas para a codificação automatizada pelo CLAN, ainda não solucionados, o que pretendemos fazer durante este projeto.
- 3) A estrutura canônica nos verbos regulares é dissilábica, permitindo a distribuição do acento. Toda vez em que a criança retém apenas uma sílaba (na 1ª fase isso ocorre apenas no imperfectivo presente e no imperativo), ela mantém a vogal acentuada do radical, como em ti(ra), tã(ma) e (a)tê(nde).
- 4) Levanta-se a hipótese de que, em virtude da alta saliência perceptual da vogal acentuada, oposta à não-acentuada e à relevância funcional, elas podem desempenhar o papel de alavancar (*bootstrapping*) a apreensão do sistema vocálico do PB e a respectiva prosódia vocabular.
- 5) Na 1ª fase, a oposição entre as pessoas é marginal e a categoria de tempo está ausente, por limites cognitivos gerais. Há, contudo, limites específicos da língua que explicam as diferenças encontradas na literatura sobre como emergem as categorias verbais nas diversas línguas. Por exemplo, observe-se que, no inglês (a língua mais estudada em aquisição da linguagem), o infinitivo é homônimo com a 3ª pessoa singular do presente do indicativo (a forma não marcada sem /s/, quando a criança começa a falar), o que leva muitos pesquisadores a considerarem o *root infinitive* como a forma predominante. O mesmo foi constatado em relação ao holandês (WIJNEN; KEMPEN; GILLIS, 2001), porém os autores exploram a possibilidade de que o fato esteja relacionado aos padrões da língua do *input*. Hoekstra e Hymes (1998), ao compararem crianças que estavam adquirindo o italiano, o holandês e o inglês, constataram a frequência de infinitivos raiz respectivamente de 7%, 33% e 78%.
- 6) Também constatamos que, na 1ª fase (MLU 1.45), não há superioridade significativa dos nomes sobre os verbos

(SINIGAGLIA, 2006), ao contrário do que justificam Gentner (1982) por serem mais fáceis de recortar, já que se referem a objetos concretos, ou Kilani-Schoch (2003) ao afirmar que a explosão lexical ocorre com substantivos e não com verbos. Concordamos com Caselli *et al.* (1999, p.77) em que “o surgimento e crescimento da morfologia flexional pode variar significativamente de uma língua para outra”, ou seja, existem diferenças consideráveis dependentes da estrutura da língua adquirida.

- 7) Na presente fase do projeto, o exame dos substantivos permite afirmar que os adultos têm uma intuição de como adequar sua fala ao interlocutor ao qual eles se dirigem: o número de substantivos abstratos, o número de substantivos por cláusula e o número de substantivos derivados é muito maior quando os adultos falam entre si do que quando se dirigem à criança. Em adendo, a criança retira dos dados a que está exposta (*input*) só aquilo de que é capaz linguisticamente e cognitivamente (*intake*), resultando um léxico diferente dos adultos na criança.

As divergências entre os pesquisadores de aquisição da morfologia nos conduzem à necessidade de aprofundar a reflexão teórica e de encontrar a nível explanatório um embasamento cada vez melhor para equacionar problemas em morfologia, em particular aqueles que dizem respeito ao PB, principalmente o sistema verbal, ou seja, fenômenos como a cumulação, redundância, suprafixo, metafonía verbal, alternância vocálica (*umlaut*) nos verbos irregulares de maior frequência de uso e a interface entre morfologia e fonologia. Isso vem ao encontro do interesse crescente pela morfologia, com novas propostas teóricas, distintas das que caracterizaram o distribucionalismo. Com o intuito de sustentar teoricamente a presente proposta, nos debruçaremos em especial sobre a morfologia distributiva (HALLE; MARANTZ, 1994); o modelo organização e paradigmas (BYBEE, 1995, 1991; DRESSLER, 2001, 1999, 1997a, b) e o modelo conexionista (ELMAN *et al.*, 1996; MacWHINNEY, 2000a; PLUNKETT, K.; JUOLA, 1999).

Como nos dispomos a elaborar um programa para o processamento automático das unidades morfológicas do PB, é necessário compulsar as novas propostas para a sua descrição. Sendo assim, examinaremos as obras de Abaurre; Rodrigues (2002); Azevedo Filho (2004); Basílio (2004, 2000, 1999, 1987, 1980); Bechara (1999); Borba *et al.* (2002); Castilho (1989, 2002a e b); Cunha; Lindley-Cintra (1987); Gonçalves (2005); Ilari (2002); Ilari; Basso (2006); Kato (2002); Koch (2002); Lobato (1986); Macedo; Roncarati; Molica (1996); Mattos e Silva (2001); Melo (1986); Moura Neves (2000, 1999); Naro; Scherre (1993); Preti (1993); Roncarati; Abraçado (2003); Teixeira Durate (2001). Obviamente, continuaremos consultando os clássicos de Mattoso Câmara Jr. (1997 [1970]; 1976; 1971), bem como os dispersos (FALCÃO UCHOA, 2004).

Após programarmos o processamento automático das unidades morfológicas do PB pelo programa CLAN do projeto CHILDES, será possível realizarmos vários estudos quantitativos e qualitativos, comparando os registros entre adultos e aqueles dirigidos à criança (CDS). Tais estudos visam à discussão da influência do *input* sobre o *intake* da criança que contribuiria para a apreensão gradativa de sua gramática, levando em conta as especificidades dos registros sociolinguísticos do português utilizado pelos adultos quando a ela se dirigem.

A partir das contribuições da sociolinguística sobre os registros, vários pesquisadores têm se dedicado ao estudo dos universais (FERGUSON (1964, 1978; SNOW, 1997 [1995]) do que hoje se denomina fala dirigida à criança. Nosso intuito é demonstrar, através das diferenças morfossintáticas, lexicais, semânticas, textuais e pragmáticas entre o registro praticado pelos adultos e aquele quando fala à criança, a intuição que têm dos limites linguísticos e cognitivos impostos ao processamento verbal.

### 3 METODOLOGIA

A metodologia é bibliográfica e instrumental.

**3.1** A pesquisa bibliográfica consiste na leitura e discussão em seminários semanais da bibliografia arrolada, com os integrantes do **Grupo Integrado Produtividade Linguística Emergente**, com vistas à formação teórica de seus integrantes e ao embasamento a nível explanatório dos fenômenos que serão observados e descritos.

**3.2** A pesquisa instrumental visa à montagem do programa para processamento automático das unidades morfológicas do PB das linhas principais dos adultos, já fatiadas e acopladas ao *bullet*, nos três *corpora* colhidos. As informações a seguir são traduzidas de MacWhinney (2008), com os exemplos adaptados ao PB:

O programa MOR é usado para gerar uma linha %mor para todas as linhas principais em um arquivo CHAT. É requerido um tempo substancial para usar um programa tão complexo, a fim de explorar todas as suas capacidades analíticas.

O *design* computacional do MOR seguiu a orientação do sistema MORPH de Roland Hausser e foi implementado por Mitzi Morris. O sistema foi projetado para maximizar o seu trânsito através das várias línguas, a expansão do léxico e da gramática, e sua compatibilização com os programas CLAN. A maquinaria básica do *parser* é independente da especificidade linguística. A informação linguística específica é registrada em arquivos separados de dados. As regras da língua estão em arquivos de dados que podem ser modificadas pelo usuário. As entradas lexicais também são guardadas em arquivos ASCII e há várias técnicas para implementar o emparelhamento do léxico com um *corpus* específico. A fim de evitar arquivos lexicais muito extensos, apenas os radicais de um *corpus* são registrados e as formas flexionadas apropriadas para cada radical são compiladas por regras para cada radical.

O programa MOR gera automaticamente a linha %mor, conforme vem explicado na codificação morfossintática do manual CHAT. As palavras são rotuladas pela sua categoria sintática, conforme o exemplo a seguir:

\*CHI:     botô .  
 %mor:    v|bot-TV1&PFV=put

O programa MOR considera cada palavra da linha principal, sem levar em consideração o contexto. Para que o programa MOR possa aplicar com sucesso, é necessário configurar os arquivos da gramática e do léxico em suas posições adequadas, no catálogo da biblioteca MOR. Será necessário criar um catálogo de biblioteca específico para o MOR que seja distinto do catálogo geral da **lib do CLAN**. Serão necessários três arquivos: ar.cut (o arquivo das regras alomórficas); o cr.cut (o arquivo das regras de concatenação) e o sf.cut (o arquivo dos marcadores das formas especiais, isto é, formas marcadas). Haverá também um subcatálogo denominado lex, que contém todas as classes abertas e fechadas do léxico, tais como adj.cut, prep.cut, v.cut, ou n.cut.

A construção da gramática MOR do PB obedece aos seguintes itens:

**ar.cut:** são regras que geram variantes alomórficas a partir dos radicais ou temas.

**cr.cut:** são regras que especificam as combinações possíveis entre os morfemas da esquerda para a direita numa palavra.

O processo de construção do ar.cut e do cr.cut envolve uma lenta interação entre a construção do léxico e das regras.

**debug.cdc:** este arquivo preserva o traço completo e mais recente de uma análise para uma dada palavra pelo MOR.

**docs:** este arquivo contém instruções sobre como treinar o POST e uma lista das categorias usadas na gramática do PB.

**port.db:** a ser construída para ser usada pelo POST.

**ex.cut:** este arquivo inclui análises que serão “generalizadas” pelo MOR, as quais deverão ser filtradas ou excluídas quando ocorrerem.

**lex:** é o coração da gramática MOR e, como é o objeto essencial desse projeto, será examinado em maior detalhe: há um arquivo para listar o léxico pertencente a cada uma das partes do discurso, conforme os exemplos acima”.

O arquivo tem que dar conta dos afixos, os quais são listados no arquivo 0affix.cut.

O arquivo começa com uma lista de prefixos tais como: in-, com-, de- e inclui uma listagem para as contrações como né no, à, etc.

Elemento	Símbolo	Exemplo	Representação	Parte
prefixo	#	refazendo	re#v faze-PROG	re#
tema	R	refazendo	re#v faze-PROG	faze
sufixo	-	refazendo	re#v faze-PROG	PROG
fusão	&	refiz	re#v faze&PAST	PAST
tradução	=	gato	n gato=cat	cat
outros	o	-	-	-

**sf.cut:** este arquivo nos diz como proceder com palavras que terminam com certos marcadores como @wp (para balbucios, brincadeiras com palavras, canções etc.).

**posttags.cut:** trata-se de uma lista de tags gramaticais do PB que deverão ser incluídas quando se roda o POST.

**traintags.cut:** esta é uma lista para tags que são usados pelo POSTTRAIN quando ela cria a base de dados port.db.

Roda-se o Programa POST, inclusive o POSTMORTEM para resolver as ambiguidades, como, por exemplo, ocorrências como a, que, se.

Roda-se o programa GRASP para o estabelecimento das relações gramaticais para o *parsing*. O GRASP, também chamado de MEGRAPSP, produz estruturas de dependência rotuladas para os dados transcritos pelo CHILDES, conforme exemplos abaixo:

GR	Definição	Exemplo
SUBJ	sujeito não oracional	Paulo desenhou.
CSUBJ	oração finita sujeito	Quem não estudou dançou.
XSUBJ	oração nominal sujeito	Comer doce faz mal.

#### 4 EQUIPE

Além do coordenador, a equipe é constituída pelos membros que integram o **Grupo Integrado Produtividade Linguística Emergente**, inclusive os três bolsistas IC. O programa de Pós-Graduação em Linguística da UFSC solicitou uma bolsa de Pós-Doutorado ao PND da CAPES para obtenção de um especialista em linguística computacional que integrará o projeto. Uma assistência constante é dada pelo coordenador mundial do Projeto CHILDES, Brian MacWhinney, com quem temos apresentado trabalhos conjuntos a vários Congressos e Encontros Internacionais.

#### 5 COMENTÁRIOS FINAIS

Espera-se com este projeto dar continuidade às pesquisas que resultam na alimentação do Banco Mundial CHILDES, com dados referentes ao PB. Até o momento, já segmentamos todos os enunciados de três *corpora*, colocando os respectivos *bullets*, que possibilitam a audição simultânea, com respectivas análises no espectrograma e cotejo com a transcrição fonética dos enunciados da criança. Foi também efetuada a codificação da linha %mor dos enunciados da criança de duas fases, levantado o léxico dos determinantes, pronomes adjetivos e substantivos, preposições e substantivos. Efetuou-se, igualmente, o levantamento do léxico relativo às falas dos adultos entre si, quando se dirigem à criança e quando essa fala. A elaboração do programa para processamento automático das unidades morfológicas do PB das linhas principais dos adultos coloca à disposição dos pesquisadores que trabalham com a morfologia do português uma ferramenta poderosa para análises quantitativas e qualitativas. No plano teórico, contribui a nível explicativo para uma melhor compreensão sobre a construção das gramáticas do PB, particularmente do sistema verbal, bem como ampliar o entendimento sobre o papel do *input* na construção de tais gramáticas.

## REFERÊNCIAS

ABAURRE, M. B.M.; RODRIGUES, A. C. S. (Orgs.). **Gramática do português falado**. 4. ed. revista. Campinas: UNICAMP, 2002, v. VIII.

AZEVEDO FILHO, L. A. de. **Descrição e funcionamento da língua portuguesa**. Rio de Janeiro: Comunicação, 2004.

BASÍLIO, M. **Estruturas lexicais do português**. Petrópolis: Vozes, 1980.

\_\_\_\_\_. **Teoria lexical**. São Paulo: Ática, 1987

\_\_\_\_\_. **A delimitação de unidades lexicais**. Rio de Janeiro: Grypho, 1999.

\_\_\_\_\_. Em torno da palavra como unidade lexical: palavras e composição. **Veredas**, v. 4, n. 2, 2000.

\_\_\_\_\_. **Formação e classes de palavras no português do Brasil**. São Paulo: Contexto, 2004.

BECHARA, E. **Moderna gramática portuguesa**. 37. ed. Ver. e amp. Rio de Janeiro: Lucerna, 1999.

BITTNER, D.; DRESSLER, W.U.; KILANI-SCHOCH, M. (Orgs.). **Development of verb inflection in first language acquisition**. Berlin: Mouton de Gruyter, 2003, p. 269-295.

\_\_\_\_\_.; GAGARINA, N.; KUHNAST, M; POPOVA, P. *Aspect before Tense* in first language acquisition of Bulgarian, Russian, and German. **Proceedings of the VII<sup>th</sup> Conference on Cognitive Modelling in Linguistics, Moskau/Varna**, Varna: CML, 2005, p. 281-291.

BORBA, F. S. *et al.* **Dicionário de usos do português do Brasil**. São Paulo: Ática, 2002.

BYBEE, J. Natural morphology: The organization of paradigms and language acquisition. *In*: HUEBNER, T.; FERGUSON, C. (eds.). *Crosscurrents in second language acquisition and linguistic theories*. Amsterdam: John Benjamins, 1991, p. 67–91.

\_\_\_\_\_. Regular morphology and the lexicon. **Language and Cognitive Processes**, v. 10, p. 425–455, 1995.

CASELLI, C.; CASADIO, P.; BATES, E..A comparison of the transition from first words to grammar in English and Italian. **Child Lang**, v. 26, p. 69-111, 1999.

CASTILHO, A. T. (org.). **Português culto falado no Brasil**. Campinas: UNICAMP, 1989.

\_\_\_\_\_. (Org.). **Gramática do português falado**, 4. ed. Revista. Campinas: UNICAMP/FAPESP. 2002a, v. I., A ordem.

\_\_\_\_\_. (Org.). **Gramática do português falado**, 3<sup>a</sup> ed. Revista. Campinas: UNICAMP/FAPESP. 2002a, v. III., As abordagens.

CUNHA, C.; LINDLEY-CINTRA, L.F. **Nova gramática do português contemporâneo**. Rio de Janeiro: Nova Fronteira, 1987.

DRESSLER, W.U. (ed.). **Studies in pre- and protomorphology**. Viena: Verlag der Oester-reichischen Akademie der Wissenschaften, 1977a.

\_\_\_\_\_. Universals, typology, and modularity in natural morphology. In: HICKEY, R.; PUPPEL, S (eds.). **Language history and linguistic Modeling**, Berlin/New York: de Gruyter, 1997b, p. 1399-1421.

\_\_\_\_\_. What is natural in natural morphology? **Travaux du Cercle Linguistique de Prague/Prague Linguistic Circle Papers**, v. 3, p. 135-144, 1999.

\_\_\_\_\_. Extragrammatical vs. marginal morphology. *In*: DOLESCHAL, U.; THORNTON, A. M. (Eds.). **Extragrammatical and Marginal Morphology**. Munique: Lincom, 2001, p. 1-10.

EDELMAN, S.; SOLAN, Z.; HORN, D. & RUPPIN, E. Bridging computational, formal and psycholinguistic approaches to language. **Cognitive Science Society**, 2004.

ELMAN, J. L.; BATES, E.; JOHNSON, M.H.; KARMILOFF-SMITH, A.; PARISI, D.; PLUNKETT K. **Rethinking innateness: A connectionist perspective on development**. Cambridge, Mass./London: MIT Press, 1999.

FALCÃO UCHOA, C. E. (org.). **Dispersos de J. Mattoso Câmara Jr.** Rio de Janeiro: Lucerna, 2004.

FERGUSON, C.A. Baby tal in six languages. **American Anthropologist**, v. 66, n. 5, p. 103-14, 1964.

\_\_\_\_\_. Talking to children: A search for universals. In: GREENBERG, J.; FERGUSON, C.A.; MORAVSIC, E.A. (Eds.). **Universals of human language**, v. 1, Stanford, CA: Stanford University, 1978, p. 203-24.

GENTNER, D. Why are nouns learned before verbs: linguistic relativity versus natural partitioning. In: KUCZAJ, S. II (ed.). **Language development**, v. 2: **Language, thought and culture**. Hillsdale, NJ: Erlbaum, 1982.

GONÇALVES, C. A. **Flexão & derivação em português**. Rio de Janeiro: Faculdade de Letras/UFRJ, 2005.

HALLE, M.; MARANTZ, A. Some key features of distributed morphology. **MIT working papers in linguistics**, v. 21, p. 275-288, 1994.

HOEKSTRA, T.; HYMES, N. The syntax and interpretation of “dropped” categories in early grammar: A unified account. **Proceedings of the WCCFL 14**. Stanford: Stanford University, 1998.

ILARI, R. (org.). **Gramática do português falado**. 4. ed. revista. Campinas: UNICAMP, 2002, v. II, Níveis de análise linguística.

\_\_\_\_\_.; BASSO, R. **O português da gente** – a língua que estudamos, a língua que falamos. São Paulo: Contexto, 2006.

KATO, M. (org.). **Gramática do português falado**. 2. ed. revista. Campinas: UNICAMP, 2002, v. V, Convergências.

KILANI-SCHOCH, M. Early verb inflection in French: An investigation of two corpora. *In*: BITTNER, D.; DRESSLER, W.U.; KILANI-SCHOCH, M. (eds.). **Development of verb inflection in first language acquisition**. Berlin: Mouton de Gruyter, 2003, p. 269-295.

KOCH, I. G. (org.). **Gramática do português falado**. 2. ed. revista. Campinas: UNICAMP, 2002, v. VI.

LOBATO, L.M.P. **Sintaxe gerativa do português: da teoria padrão à teoria da regência e ligação**. Belo Horizonte: Vigília, 1986.

MACEDO, A.; RONCARATI, C.; MOLICA, M.C. (Orgs.). **Variação e discurso**. Rio de Janeiro: Tempo Brasileiro, 1996.

MacWHINNEY, B. Lexicalist connectionism. *In*: BROEDER, P.; MURRE, J. (eds.). **Models of language acquisition**. Oxford: Oxford University Press, 2000a, p. 9-31.

\_\_\_\_\_. **The CHILDES Project: Transcription on Format and Programs**. 3. ed. New Jersey: Lawrence Erlbaum, 2000b, 3. ed, v. I e II.

\_\_\_\_\_. **Enriching CHILDES for Morphosyntactic Analysis**. Plataforma CHILDES, 2008. <http://childes.psy.cmu.edu/morgrams/morphosyntax.doc>, Acessado em 22/09/2008.

MALVERN, D. D., RICHARDS, B. J., CHIPERE, N., & PURÁN, P. **Lexical diversity and language development**. New York: Palgrave Macmillan, 2004.

MARCUS, G., ULLMAN, M., PINKER, S., HOLLANDER, M., ROSEN, T., & XU, F. Overregularization in language acquisition. **Monographs of the Society for Research in Child Development**, 57(4), 1-182, 1992.

MATTOS E SILVA, R. V. (org.) **Para a história do português brasileiro**. São Paulo: Humanitas/FAPESP, 2001, v. II, Primeiros estudos.

MATTOSO CAMARA JR., J. **Problemas de linguística descritiva**. Petrópolis: Vozes, 1971.

\_\_\_\_\_. **História e estrutura da língua portuguesa**, 2. ed. Rio de Janeiro: Padrão, 1976.

\_\_\_\_\_. **Estruturas da língua portuguesa**. Petrópolis: Vozes, 1997 [1970], 26. ed.

\_\_\_\_\_. Para o estudo descritivo dos verbos irregulares. *In*: FALCÃO UCHOA, C. E. (Org.). **Dispersos de J. Mattoso Câmara Jr.** Rio de Janeiro: Lucerna, 2004, p.131-146.

MELO, G. C. de. **A lingual do Brasil**. 4. ed. Rio de Janeiro: Padrão, 1986.

MOURA NEVES, M. H. de. (org.). **Gramática do português falado**. 2. ed. revista. Campinas: UNICAMP/Humanitas, 1999, v. VII, Novos estudos.

\_\_\_\_\_. **Gramática de usos do português**. São Paulo: UNESP, 2000.

NARO, A. SCHERRE, M. M. P. Sobre as origens do português popular do Brasil. **DELTA**, v. 9, p. 437-54, 1993.

PLUNKETT, K.; JUOLA, J. A connectionist model of English past tense and plural morphology. **Cognitive Science**, v. 23, p. 463–490, 1999.

PRETI, D. (org.). **Análise de textos orais**. São Paulo: FFLCH/USP, 1993.

RONCARATI, C.; ABRAÇADO, J. (Orgs.). **Português brasileiro: contato linguístico, heterogeneidade e história**. Rio de Janeiro: Letras/FAPERJ, 2003.

SINIGAGLIA, J. Emergência dos verbos no português do Brasil. Pôster apresentado à **58ª Reunião Anual da SBPC**, Florianópolis: UFSC, 2006.

SNOW, C. E. Questões no estudo do *input*: sintonia, universalidade, diferenças individuais evolutivas e causas necessárias. *In*: FLETCHER, P.; MacWHINNEY, B. (Orgs.) **Compêndio da linguagem da criança**. Trad. de Marcos A.G. DOMINGUES; Regina Ritter LAMPRECHT. Porto Alegre: Artes Médicas, 1997, p. 153-163.

TEIXEIRA DUARTE, P. M. **Elementos para uma morfologia do português**: em torno da noção de radical. Fortaleza: UFC, 2001.

WIJNEN, F; KEMPEN, M.; GILIS, S. Root infinitives in Dutch early child language: An effect of input. *J. Child Lang.*, v. 28, p. 629-660, 2001.