

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE LETRAS
CURSO DE PÓS-GRADUAÇÃO EM LETRAS

PROJETO:

**AQUISIÇÃO DA ANÁFORA PRONOMINAL NO PORTUGUÊS DO BRASIL:
TRANSCRIÇÃO E INVESTIGAÇÃO DE *CORPORA* DO CEAAL/PUCRS**

por

Prof. Dr. Sergio Menuzzi
Curso de PG em Letras da PUCRS

Porto Alegre, outubro de 1999.

1. Título do Projeto:

Aquisição da Anáfora Pronominal no Português do Brasil:
Transcrição e Investigação de *Corpora* do CEAAL/PUCRS

2. Dados de Identificação

Responsável: Prof. Dr. Sergio de Moura Menuzzi
Instituição: Curso de Pós-Graduação em Letras, PUCRS
Modalidade: Bolsa de Iniciação Científica
Número de Bolsas Solicitadas: duas (02)
Período: março de 2000 a março de 2001

3. Situando o Projeto

Em pesquisa já em andamento, o responsável pelo presente projeto estuda a aquisição dos elementos pronominais *se*, *ele* e *ele mesmo* em português do Brasil, com o objetivo de avaliar algumas teorias recentes sobre a natureza e a aquisição de elementos anafóricos (ver anexo 2). Entretanto, uma etapa importante da pesquisa, a investigação do aparecimento das formas referidas na produção normal de fala da criança, não pode progredir a contento devido à ausência de uma transcrição apropriada de parte fundamental dos *corpora* disponíveis.

Especificamente, os *corpora* de fala atualmente disponíveis de crianças nascidas em Porto Alegre são aqueles encontrados no CEAAL (“Centro de Estudos de Aquisição e Aprendizagem da Linguagem”, PUCRS) e os do projeto DELICRI (“Desenvolvimento da Linguagem da Criança na Fase de Letramento”, UFRGS). Estes últimos estão transcritos e computadorizados de acordo com o sistema CHAT do programa internacional CHILDES (cf. Guimarães e Lamprecht 1995, MacWhinney 1995); estão, portanto, em formato apropriado para o trabalho de levantamento dos dados. Entretanto, a amostra do projeto DELICRI cobre apenas o período de desenvolvimento infantil que vai dos 5 aos 9 anos de idade; isso deixa em aberto um período crucial no desenvolvimento da anáfora, que são os anos iniciais da criança, especialmente entre os 3 e os 5 anos (ver dados das demais línguas já investigadas, cf. anexo).

Quanto aos *corpora* disponíveis no CEAAL/PUCRS, apenas os coletados por Denise Issler e por Luciene Simões encontram-se em condições de terem os dados levantados (estão transcritos e computadorizados de acordo com uma adaptação do CHAT, cf. Simões 1997 e Issler 1998).¹ Estes dois *corpora*, entretanto, cobrem o período de desenvolvimento que vai de 1 ano e seis meses a 3 anos; a lacuna entre 3 e 5 anos continua em aberto, portanto.

¹ Os dados destes já foram levantados, e os resultados são discutidos em Menuzzi (1999).

Na realidade, o CEAAL dispõe de gravações que cobrem este período, aquelas coletadas para a constituição do banco de dados AQUIFONO (“Aquisição da Fonologia do Português do Brasil”, PUCRS). São fitas contendo entrevistas de cerca de 30 minutos com 310 sujeitos (divididos por sexo e por faixas etárias de 2 em 2 meses); no diz respeito ao período que nos interessa, são 120 sujeitos (60 meninos e 60 meninas) gravados num total de cerca de 60 horas de entrevistas.

As entrevistas utilizam como método de elicitación de dados a “homeação espontânea” e a “narração e fala espontâneas”, em geral motivadas pela apresentação, por um experimentador, de objetos, figuras e bonecos à criança (cf. Yavas, Hernandorena e Lamprecht 1991). As entrevistas se revestem, em geral, do caráter de brincadeira e, na maioria dos casos, resultam em situação informal de interação. Por isso, oferecem uma amostra representativa da fala natural da criança e podem servir para a investigação não apenas de aspectos fonológicos da fala infantil, mas também de outros aspectos, como o desenvolvimento da sintaxe, de capacidades de interação discursiva, etc.

Entretanto, por se encontrar em “estado bruto”, este material até agora não pôde ser investigado por pesquisadores interessados em aspectos *não-fonológicos* do desenvolvimento da fala infantil: como foram originalmente coletadas para o estudo da aquisição da fonologia do português do Brasil, as entrevistas só receberam transcrição e computadorização apropriada para a investigação de aspectos fonológicos; em particular, não há uma transcrição global das entrevistas, mas apenas das palavras isoladas que atestam os diferentes padrões fonológicos originalmente investigados. O que o presente projeto propõe é que se execute, portanto, uma nova transcrição de parte destas entrevistas: uma transcrição que seja, entretanto, apropriada à investigação de aspectos não-fonológicos do desenvolvimento da fala infantil.

A transcrição que se pretende fazer seguirá parâmetros do sistema CHAT de transcrição do projeto CHILDES (ver Metodologia abaixo). Com isso, o projeto visa não apenas atender a necessidades da pesquisa específica à qual está vinculado (cf. anexo 2), mas também cooperar para a consolidação e ampliação dos bancos de dados de fala infantil do CEAAL, e de sua integração ao sistema internacional do projeto CHILDES. Nisso, o presente projeto se junta ao projeto “Preparo e Codificação de *Corpora* do Banco de Dados INIFONO no Sistema CHAT do Projeto CHILDES”, que também está sendo submetido à FAPERGS pela Profa. Dra. Regina Lamprecht: ambos fazem parte de um plano comum de longo prazo do CEAAL/PUCRS — gradualmente transcrever e codificar *todos* os *corpora* do CEAAL em formato CHAT e integrá-los ao banco internacional de dados do projeto CHILDES.

4. Objetivos

Em resumo: os *objetivos específicos* do presente projeto são os seguintes:

- (1) treinar bolsistas de iniciação científica para a executar a transcrição e a codificação de dados de acordo com sistema CHAT;
- (2) transcrever um subconjunto específico das entrevistas acima descritas e codificar certos aspectos morfosintáticos e semântico-pragmáticos delas de acordo com o sistema CHAT (ver Metodologia abaixo);
- (3) levantar, de acordo com os critérios discutidos na seção 5.1 da pesquisa em anexo, a distribuição das formas *se*, *ele* e *ele mesmo* no subconjunto de entrevistas transcritas.

E os *objetivos gerais* do projeto são os seguintes:

- (4) contribuir para a compreensão do processo de aquisição da anáfora pronominal do português do Brasil executando tarefas cruciais para o andamento da pesquisa em anexo;
- (5) contribuir para a ampliação e a instrumentalização dos bancos de dados do CEAAL/PUCRS, digitalizando *corpora* ainda não-transcritos em formato CHAT, apropriado para o uso dos recursos computacionais do sistema CLAN (também desenvolvido pelo projeto CHILDES, cf. MacWhinney 1995);
- (6) contribuir para a eventual integração dos bancos de dados do CEAAL/PUCRS no banco internacional do projeto CHILDES.

5. Justificativa

O presente projeto tem como justificativa científica imediata o fato de permitir o andamento de uma pesquisa que procura responder a perguntas centrais da literatura recente sobre a aquisição e o desenvolvimento do conhecimento sintático relacionado à anáfora pronominal (ver anexo 2). Além disso, o projeto também se justifica por seu papel pequeno, mas importante, no desenvolvimento da pesquisa em aquisição da linguagem em nosso estado e no Brasil: contribuirá não apenas para o esforço de consolidação do sistema de bancos de dados do CEAAL/PUCRS, mas também, para formação de pessoal qualificado a gerenciar adequadamente estes bancos no futuro. Finalmente, o presente projeto também se justifica porque colaborará com processo de adequação dos bancos de dados do CEAAL/PUCRS aos requisitos de um sistema internacional integrado de bancos de fala infantil, o do projeto CHILDES: isso permitirá que pesquisadores de outros países interessados nos dados de aquisição do português do Brasil tenham fácil acesso a estes dados, o que evidentemente serve de estímulo à cooperação de nossos pesquisadores com a comunidade científica internacional.

6. Metodologia

(1) Treinamento dos Bolsistas:

Os bolsistas passarão por um período inicial (cerca de um mês) de preparação teórica básica, em que estudarão algumas noções, problemas e objetivos da pesquisa sobre a

aquisição da linguagem (cf. Elliot 1981, Ingram 1989). Este estudo será supervisionado pelo pesquisador responsável por meio de encontros semanais para discussão das leituras. Após a preparação teórica, os bolsistas serão treinados na transcrição e na codificação dados de fala de acordo com o sistema CHAT, bem como na operação das ferramentas básicas do sistema CLAN de *softwares*. Esta treinamento se dará por um período de aproximadamente dois meses, em dois encontros semanais: um encontro servirá para supervisão da leitura dos manuais dos sistemas CHAT e CLAN (em MacWhinney 1995), e o outro, para exercícios supervisionados de prática de transcrição, codificação, e de operação dos *softwares* do CLAN no computador. Finalmente, após a transcrição e a codificação dos *corpora*, os bolsistas passarão por mais um período de treinamento para o levantamento de dados previsto pelo projeto anexado. Este treinamento se dará em duas semanas: a primeira será dedicada ao estudo supervisionado do projeto em anexo (dois encontros), e a segunda será dedicada a exercícios práticos supervisionados de levantamento de dados (dois encontros).

(2) *Corpora a serem transcritos e codificados:*

Em princípio, pretende-se transcrever e codificar os *corpora* dos 120 sujeitos (60 meninos e 60 meninas) que pertencem à faixa etária de interesse deste projeto (entre os 3;0 e os 5;0 anos de idade), o que totaliza cerca de 60 horas de entrevistas. Se a extensão da tarefa se revelar impraticável no período de duração do projeto, far-se-á uma seleção dos *corpora* (p. ex., reduzindo-os para 80 sujeitos, 20 na faixa entre 3;0 e 3;6, 20 na faixa entre 3;6 e 4;0, 20 na faixa entre 4;0 e 4;6, e 20 na faixa entre 4;6 e 5;0).

(3) *Transcrição e Codificação dos Dados*

Como se disse antes, os dados serão transcritos e codificados de acordo com o sistema CHAT. Este sistema oferece um conjunto amplo de opções de transcrição e de codificação de diferentes aspectos da fala (ver MacWhinney 1995). Como o presente projeto tem interesse imediato em certas estruturas sintáticas da fala da criança e em alguns aspectos da sua interpretação (em particular, qual a referência dos pronomes e outras formas anafóricas da fala da criança), far-se-á a transcrição de acordo com as opções mínimas de satisfação dos requisitos do sistema CHAT e, adicionalmente, codificar-se-á aqueles aspectos que atendem aos interesses imediatos do projeto. Isso significa que os enunciados terão de receber não apenas uma transcrição na chamada “linha principal”, mas também codificação ocasional em algumas “linhas dependentes”:

Linha principal (*Main Tier*): transcreve exatamente o que o falante disse de acordo com uma adaptação convencional da ortografia oficial: p. ex., partes omitidas de uma palavra aparecem entre parênteses (“(es)tá”); palavras ou trechos não reconhecíveis são transcritos como “xxx”; pausas, como “#”, “##” ou “###”, dependendo da duração; etc.

Linha “normalizada” (*Flow Tier*): uma versão da transcrição tão livre quanto possível de convenções de transcrição, e o mais próxima possível da ortografia oficial (p. ex., só

registra as pausas que receberiam representação na ortografia oficial); tem a função de facilitar da leitura da transcrição;

Linha de explicação (*Explanation Tier*): serve para especificar a referência (dêitica ou não), de expressões que aparecem no enunciados transcrito no tier principal: p. ex., se “ele” é usado pelo falante para se referir deiticamente a João, codifica-se esta informação no tier (“ele=João”); serve também para codificar aspectos da situação relevantes para interpretação de expressões do enunciado: p. ex., se a criança diz “Toma.” e entrega uma boneca ao ouvinte (“F entrega um boneca a O”);

Linha morfossintática (*Morphosyntactic Tier*): codifica as palavras do enunciado de acordo com sua categorização morfossintática e sua segmentação morfológica: p. ex., “Quero bolacha” é codificado como “V/querer-PRES&1S N/bolacha”.

Para ilustração do resultado final de uma transcrição em formato CHAT, ver anexo 1 a este projeto (em que “*PAT” é o código do tier principal, “%flo” do tier normalizado, e “%mor” do tier morfossintático).

(4) *Levantamento de Dados*

O levantamento da distribuição das formas *se*, *ele* e *ele mesmo* nos *corpora* transcritos se dará de acordo com os critérios discutidos na seção 5.1 da pesquisa em anexo.

7. Plano de Trabalho dos Bolsistas

(Para maiores detalhes acerca das atividades do bolsista, ver Metodologia.)

- (1) Março, Abril e Maio de 2000:
treinamento dos bolsistas, primeira fase;
- (2) Junho, Julho, Agosto, Setembro e Outubro de 2000:
transcrição dos *corpora* selecionados;
- (3) Novembro de 2000:
treinamento dos bolsistas, segunda fase;
- (4) Novembro e Dezembro de 2000, Janeiro e Fevereiro de 2001:
levantamento dos dados.

8. Referências Bibliográficas

Elliot, Alison J. *Child language*. Cambridge: Cambridge University Press, 1981.

- Guimarães, Ana Maria de Mattos & Lamprecht, Regina Ritter. "The Use of the CHILDES Database for Brazilian Portuguese." Em I. H. Faria & M. J. Freitas, eds., *Studies on the Acquisition of Portuguese*, págs. 207-214. Lisboa: Colibri/Associação Portuguesa de Lingüística, 1995.
- Ingram, David. *First language acquisition: method, description, and explanation*. Cambridge: Cambridge University Press, 1989.
- Issler, Denise Silveira. "A aquisição de 'eu' e 'tu' : intersecções entre a lingüística e a psicologia." Porto Alegre: Tese de Doutorado, PUCRS, Inst. de Letras e Artes, 1998.
- MacWhinney, Brian. *The CHILDES Project: Computational Tools for Analyzing Talk*. Hillsdale (EUA): Lawrence Erlbaum, 1995.
- Menuzzi, Sergio. "Topics in Binding Theory: Constraint Interaction, Chains, Indexing and Reflexivity, with Particular Reference to Brazilian Portuguese." Porto Alegre: Tese de Doutorado, PUCRS, Inst. de Letras e Artes, 1998.
- Menuzzi, Sergio. "Multidisciplinaridade na Sintaxe do Século XXI: Teoria Sintática e Aquisição da Linguagem." Porto Alegre: 3o. Encontro do CELSUL, 1999.
- Simões, Luciene Juliano. "Sujeito nulo na aquisição do português brasileiro: um estudo de caso." Porto Alegre: Tese de Doutorado, PUCRS, Inst. de Letras e Artes, 1997.
- Yavas, Mehmet; Hernandorena, Carmen L. M. e Lamprecht, Regina R. *Avaliação fonológica da criança: reeducação e terapia*. Porto Alegre: Artes Médicas, 1992.