

A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil

*Catherine Anne Lonngren Sampaio**

Resumo: *Este artigo descreve a investigação de alternância de código (codeswitching) em um corpus eletrônico de linguagem bilíngüe infantil (português e inglês) construído seguindo o sistema de transcrição do projeto CHILDES (Child Language Data Exchange System), um sistema de troca de dados de linguagem infantil originalmente desenvolvido na Universidade Carnegie Mellon nos Estados Unidos. A pesquisa enquadra-se na área da Lingüística de Corpus, que pode ser caracterizada pela investigação de fenômenos lingüísticos baseada em exemplos reais de linguagem, armazenados em corpora eletrônicos, compilados de forma sistemática e segundo determinados critérios. O corpus analisado no presente trabalho é composto de transcrições da linguagem falada de duas crianças bilíngües que estão aprendendo português e inglês simultaneamente. Os dados foram coletados através de gravações durante um período de aproximadamente dois anos e transcritos seguindo as convenções estabelecidas pelo CHILDES. Fora as convenções mínimas, foram inseridos códigos específicos nas transcrições que permitissem a investigação do uso que as duas crianças fazem da alternância de código, tanto de uma perspectiva estrutural quanto de uma perspectiva pragmática. No caso do estudo específico relatado aqui, foram analisadas nove conversas telefônicas gravadas durante o período de dois meses que as crianças passaram na Inglaterra. As conversas têm como principais interlocutores as duas crianças e o pai brasileiro.*

* Mestre em Lingüística Aplicada da Universidade Estadual do Ceará.

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil.*

Através das ferramentas do CHILDES investigou-se tanto a alternância de código para inglês quanto para o português.

Palavras-chave: *Linguística de Corpus; CHILDES; linguagem bilíngüe; alternância de código.*

Abstract: *This article describes the investigation of codeswitching in a computerised corpus of child bilingual language (English/Portuguese) built following the transcription system of the CHILDES (Child Language Data Exchange System) project, a data exchange system of child language originally developed at Carnegie Mellon University in the United States. The research belongs to the area of Corpus Linguistics, which is characterized by the investigation of linguistic phenomena based on examples of real-life language stored in computerised corpora compiled in a systematic manner and following certain criteria. The corpus analysed in the present article is composed of transcriptions of the spoken language of two bilingual children who are learning Portuguese and English simultaneously. The data were collected through recordings carried out over a period of approximately two years and transcribed according to the conventions established by CHILDES. Apart from the minimum conventions, specific codes were inserted in the corpus which permitted the investigation of the use which both children made of codeswitching from both structural and pragmatic perspectives. In the case of the specific study reported here, nine telephone conversations, recorded while both children were in England for a period of two months, were analysed. With the children and their father, a Brazilian, as the main interlocutors, codeswitching to both English and Portuguese was investigated through the tools of analysis which integrate CHILDES.*

Keywords: *Corpus Linguistics; CHILDES, bilingual language; codeswitching.*

1. Introdução

A investigação relatada no presente trabalho faz parte de uma pesquisa de mestrado realizada pela Universidade Estadual do Ceará na qual foi descrito

¹ Seguindo Sardinha (1999a; 1999b), os termos 'corpus' e 'corpora' não serão italizados no presente trabalho.

o processo da construção e exploração de um corpus¹ eletrônico de linguagem bilíngüe infantil seguindo os padrões do CHILDES (Child Language Data Exchange System)². Este projeto multilíngüe foi originalmente concebido em 1981, por Brian MacWhinney, Dan Slobin, Willem Levelt e Susan Ervin-Tripp, nos Estados Unidos. O objetivo original do projeto foi estabelecer um sistema de transcrição padronizado e sistematizado que permitisse a prática de troca de dados, armazenados em corpora eletrônica, entre pesquisadores do mundo todo, possibilitando progresso no estudo científico da aquisição da linguagem. O sistema CHILDES, disponível de forma gratuita pela Internet, constitui-se dos seguintes três componentes:

1. O banco de dados do CHILDES, que conta com corpora de diversas fontes, contribuições de mais de 100 pesquisadores que trabalham na área da linguagem infantil. Há dados de várias línguas³, e, sendo um banco de dados aberto, qualquer pesquisador pode contribuir com seus dados, contanto que estejam transcritos em formato CHAT (veja abaixo).
2. CHAT (Codes for the Human Analysis of Transcripts) que consiste de um sistema de transcrição cujo objetivo é uniformizar a codificação de corpora para permitir a análise através do uso de ferramentas eletrônicas.
3. CLAN (Computerized Language Analysis) que é um programa especificamente desenvolvido para analisar dados transcritos no formato CHAT.

O CHAT é um sistema de transcrição muito claro e completo, fácil de aplicar nos dados e fácil de ler. Além da transcrição propriamente dita, que focaliza “a produção de um registro escrito que nos permite compreender, embora de maneira vaga, o fluxo da interação original”⁴ (MacWhinney, 2003:13), o sistema CHAT inclui uma série de convenções para codificação, definida por MacWhinney como “o processo de reconhecer, analisar e tomar nota de fenômenos no discurso falado transcrito”⁵ (p. 13). Desde que o corpus seja anotado

² A dissertação completa está disponível no seguinte endereço: www.lonngren.com.br/cathy.

³ A natureza de cada corpus é descrita no volume II do manual mencionado na nota acima.

⁴ Tradução minha do original: “...the production of a written record that can lead us to understand, albeit only vaguely, the flow of the original interaction”.

⁵ Tradução minha do original: “...the process of recognizing, analyzing, and taking note of phenomena in transcribed speech”.

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

segundo o padrão sugerido no manual do CHILDES, o programa de análise, CLAN, pode ser utilizado na sua exploração. As possibilidades de análises aumentam com a inserção de códigos, a escolha dos quais vai variar de acordo com o objetivo do pesquisador. Em relação ao corpus da pesquisa, foram inseridos códigos criados pela própria pesquisadora que permite a análise das línguas separadamente, possibilitando, assim, a investigação de alternância de código. No presente artigo, pretende-se demonstrar a metodologia utilizada na investigação desse fenômeno e discutir os resultados obtidos das análises eletrônicas realizadas através do programa que integra ao CHILDES. Antes, entretanto, será descrita a seguir a natureza do corpus.

2. O LOBILL Corpus

Os dados usados na pesquisa fazem parte de um corpus aberto, denominado LOBILL⁶ Corpus, composto da linguagem falada de duas crianças bilíngües nas suas interações com interlocutores mono e bilíngües, em diversas situações de família. Os informantes principais são MEG⁷, uma menina de 8;2 anos⁸, e JAM, seu irmão de 5;10 anos⁹ (tendo respectivamente 5;10 e 3;5 anos no início da coleta de dados). Ambos nasceram em Fortaleza e desde 1;6 anos freqüentam uma escola brasileira. A mãe, MOT, que é a própria pesquisadora, é inglesa, casada com um brasileiro que fala inglês fluentemente. Com mais de 10 anos no Brasil, MOT fala português fluentemente. Embora desde o nascimento, a mãe interaja com os dois filhos, MEG e JAM, exclusivamente em inglês, o pai, PAI, se dirige a eles em português a maior parte do tempo. A interação entre os irmãos é predominantemente em português. Fora a interação com a mãe, o contato diário com o inglês se restringe a programas de televisão ('Cartoon Network' e 'Discovery Channel') e livros de histórias. Visitas ocasionais de parentes ingleses proporcionam uma outra fonte importante de contato com inglês e ambas as crianças têm passado temporadas curtas na Inglaterra; a última

⁶ LOBILL significa Lonngren BILingual Language, Lonngren sendo o sobrenome dos sujeitos.

⁷ Nesse artigo, os informantes são indentificados pelos códigos de identidade que foram utilizados nas transcrições: seguindo as normas do CHAT, o código é composto de uma combinação de três letras maiúsculas, que pode ser baseada no nome próprio do falante, por exemplo, MEG (Meggie), ou no seu papel, por exemplo, MOT (Mother).

⁸ Em dezembro 2003.

⁹ Em dezembro 2003.

sendo durante os meses de junho, julho e agosto de 2003. Os quatro membros da família são considerados 'bilíngües balanceados' (*balanced bilinguals*) (Baker, 2000:5). Embora o conceito 'bilíngüe balanceado' não seja bem definido (p. 5-7), aqui se refere a uma pessoa que tem 'competência aproximadamente igual em duas línguas' (p. 167)¹⁰.

Atualmente, contamos com, aproximadamente 18 horas de gravações. A coleta dos dados teve início em agosto de 2001 e continuou ao longo da pesquisa. Os dados foram coletados principalmente por meio de gravações em áudio, com um microcassete para permitir a gravação das conversas sem as crianças perceberem a sua presença. Entretanto, as gravações também incluem 'entrevistas' com os sujeitos, que, sendo pré-planejadas, ocorreram com o conhecimento pleno das crianças. Anotações de campo complementam as gravações.

No início da coleta de dados, não foi seguido nenhum sistema de transcrição devido ao fato de existirem tantos sistemas diferentes que não pareciam se adequar ao objetivo da pesquisa. As transcrições eram muito básicas e não faziam uso de convenções específicas. Ao conhecer o CHILDES, ficou evidente que o CHAT oferecia o que estava sendo procurado: uma maneira sistemática de transcrever os dados do corpus, que, ao mesmo tempo, abrisse a possibilidade de realizar análises alternativas. Fora as convenções de transcrição, foram inseridos códigos para permitir a análise das línguas separadamente: em interações em que a língua principal foi identificada como inglês (indicado pelo 'en' no nome do arquivo, por exemplo 27enJ&MNOV02.cha), somente o material em português foi marcado (<*material em português*>[@pt]); em interações em que a língua principal foi identificada como português (indicado pelo 'pt' no nome do arquivo, por exemplo 52ptJ&MJUN03.cha), somente material em inglês foi marcado (<*material em inglês*>[@en]). É fora do âmbito desse artigo descrever em detalhes todos os códigos que foram inseridos nas transcrições durante o processo da construção do LOBILL Corpus seguindo CHAT (veja dissertação para mais detalhes). Entretanto, a inserção dos códigos mencionados acima foi fundamental para a realização das análises com CLAN na investigação do uso de alternância de código pelas duas crianças, assunto da próxima seção.

3. A investigação de alternância de código no LOBILL Corpus

Atualmente o termo alternância de código (*codeswitching* em inglês) é utilizado para descrever "qualquer alternância durante uma única conversa,

¹⁰ Tradução minha do original: "...approximately equal competence in two languages".

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilingüe Infantil*.

podendo ser no nível da palavra ou sentença ou no nível de blocos de discurso”¹¹ (Baker, 2001:101). Não é um fenômeno puramente lingüístico, mas envolve muitas variáveis cognitivas, psicológicas, sociais e políticas, que determinam a sua natureza em um contexto específico. No estudo sobre alternância de código, muitos lingüistas investigam as possíveis restrições lingüísticas que acreditam governar a alternância de código (veja Poplack, 1980; MacSwan, 1999; Myers-Scotton, 2002). Embora não caiba aqui discutir os modelos de alternância de código propostos por esses teóricos, considerou-se útil para a discussão dos resultados o uso dos termos ‘Língua Matriz’ (*Matrix Language*) e ‘Língua Encaixada’ (*Embedded Language*) (Myers-Scotton, 2002). Baseado num construto gramatical, a Língua Matriz é aquela que mais contribui com as estruturas morfossintáticas para sentenças bilíngües (freqüentemente considerada a língua ‘dominante’), enquanto que a Língua Encaixada é caracterizada pela sua contribuição com morfemas de conteúdo (como substantivos, verbos, adjetivos e preposições) (p. 62). Essa assimetria entre as línguas participantes na alternância de código e as possíveis restrições estruturais que resultam podem ser investigadas no LOBILL Corpus através do CLAN, pelo uso dos comandos ‘**FREQ**’ (de ‘**FREQ**uency’ – ‘freqüência’) e ‘**KWAL**’ (‘**Key**Word **And** **Line**’ – palavra e linha chave), conforme descrito a seguir.

Na sua forma mais básica, FREQ produz uma lista de freqüência das palavras de um ou mais arquivos. Ao especificar o falante e a língua, os resultados produzidos podem ser analisados para averiguar o uso de cada língua em relação à categoria gramatical das palavras. A repetição dessas análises em outros arquivos possibilita a investigação de tendências no uso das duas línguas para certas categorias gramaticais e, portanto, a investigação da oposição entre Língua Matriz e Língua Encaixada. Uma análise detalhada das listas produzidas por FREQ pode revelar várias categorias gramaticais, como verbos, artigos, preposições, conjunções e marcadores discursivos.

O comando KWAL pode ser utilizado para investigar o uso dessas palavras em contexto, e outras questões relacionadas à estrutura morfossintática de frases bilíngües, como os possíveis pontos de alternância. Através do uso dos dois comandos (FREQ e KWAL), portanto, a relação entre o uso das duas línguas e a estrutura gramatical de frases bilíngües pode ser investigada.

Além do estudo das restrições lingüísticas de alternância de código, outras pesquisas têm como foco a investigação de aspectos pragmáticos, isto é, como os falantes utilizam a alternância de código de acordo com: ...a situação (domínios sociais, graus de formalidade do discurso etc.), os participantes da interação (idade,

¹¹ Tradução minha do original: “...any switch within the course of a single conversation, whether at word or sentence level or at the level of blocks of speech”.

sexo, *status*, origem étnica, grau de relação interpessoal etc.), o tópico (afetivo, escolar etc.) e a função da interação (um pedido, uma ordem, um argumento etc.) (Brito de Mello, 1999:122)

Para investigar no LOBILL Corpus a alternância de código em relação a essas quatro variáveis (situação, participantes, tópico e função), como veremos, novamente os comandos *FREQ* e *KWAL* oferecem uma maneira eficiente de fazê-lo. Entretanto, análises específicas somente se tornaram possíveis devido à inserção dos códigos e cabeçalhos que marcam a língua da interação, o interlocutor e seu interlocutor, e a situação.

Com respeito à variável 'tópico', ambos os comandos *FREQ* e *KWAL* podem revelar evidências sobre a relação do uso do português ou inglês e o tópico da interação. Através de outras análises similares no corpus, talvez seja possível determinar os 'domínios' de cada língua e investigar a estabilidade desses domínios nos dados longitudinais.

A investigação das várias funções da alternância de código é outra área que pode ser explorada no LOBILL Corpus através dos comandos do *CLAN*. Por exemplo, foi observado nas interações que a alternância de código freqüentemente ocorre quando um dos falantes utiliza discurso direto. No LOBILL Corpus trechos de discurso direto foram transcritos da seguinte forma:

***JAM:** é assim ó +”/.

***JAM:** +” there's too much butter on those trays +”.

Utiliza-se +”/. para indicar que o material que segue na linha seguinte é discurso direto e na linha seguinte insere-se +” para indicar o início do discurso direto e +”. para marcar seu término. Dessa forma, ao pedir uma análise de *KWAL* para o falante *JAM* em vários arquivos, seria possível averiguar outras ocorrências desse tipo de uso de alternância de código (indicadas pelos símbolos mostrados acima). Uma outra função fácil de analisar no corpus é o uso de alternância de código para fazer referências metalingüísticas do tipo ***JAM: tu sabe que é um buffer[“]?**. Na construção da linha do comando do *KWAL*, a inserção de +s”[“]” selecionaria todas as ocorrências desse tipo nos arquivos especificados do corpus. Embora, nas duas análises citadas acima, os símbolos específicos inseridos nas linhas principais facilitem a especificação dos dados para a análise, não é necessário (ou viável) codificar todas as possíveis funções de alternância de código encontradas dentro do corpus. O importante é poder selecionar nos arquivos desejados os enunciados em que ocorre alternância de código, isto é, em que se encontra a Língua Matriz e a Língua Encaixada no mesmo enunciado. Essa seleção é possível devido à inserção dos símbolos **<material em portugues>[@pt]** ou **<material em ingles>[@en]** em todas as transcri-

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilingüe Infantil*.

ções. Após a seleção através do uso de KWAL, os resultados apresentados podem ser analisados com o objetivo de identificar e classificar as várias funções de alternância de código (por falante).

Fica evidente pela discussão acima que, no LOBILL Corpus, tanto os aspectos pragmáticos quanto os aspectos estruturais da alternância de código são passíveis de investigação. Com o objetivo de ilustrar uma análise específica de forma mais detalhada, na próxima seção, serão mostrados e discutidos os resultados da investigação de alternância de código em 9 arquivos que contêm interações da mesma natureza.

4. A análise de alternância de código em nove conversas telefônicas

4.1. Os arquivos analisados

Os arquivos escolhidos para análise compartilham certas características: são gravações de conversas telefônicas que os dois sujeitos tiveram com parentes brasileiros, durante o período de dois meses que passaram na Inglaterra de férias. Durante todas as ligações, a mãe dos sujeitos esteve presente, sentada ao lado, e também conversou ao telefone. As conversas da mãe com os parentes no Brasil não foram gravadas, mas a interação entre ela e os dois sujeitos durante as conversas faz parte das gravações. A tabela a seguir traz informações sobre data e duração (em minutos) das gravações, os participantes das conversas e o nome dos arquivos:

Tabela 1: Dados dos arquivos do LOBILL Corpus utilizados nas análises

No.	Data	Duração	Participantes	Idade dos sujeitos	Nome do arquivo
1	4.07.03	4.29	JAM, VIN, MOT	JAM: 5;4.3	59ptJUL03.cha
2	4.07.03	7.40	JAM, MEG, PAI, MOT	JAM: 5;4.3 MEG: 7;8.27	60ptJ&MJUL03.cha
3	12.07.03	16.15	JAM, MEG, PAI, MOT	JAM: 5;4.11 MEG: 7;9.5	62ptJ&MJUL03.cha
4	17.07.03	11.0	JAM, MEG, PAI, MOT	JAM: 5;4.16 MEG: 7;9.10	64ptJ&MJUL03.cha
5	24.07.03	12.44	JAM, MEG, PAI, MOT	JAM: 5;4.23 MEG: 7;9.17	65ptJ&MJUL03.cha
6	1.08.03	13.53	JAM, MEG, PAI, MOT	JAM: 5;5.0 MOT: 7;9.24	69ptJ&MAUG03.cha
7	8.08.03	19.03	MEG, JAM, PAI, MOT	JAM: 5;5.7 MEG: 7;10.1	71ptJ&MAUG03.cha
8	9.08.03	2.20	JAM, MEG, VOV, SAR, MOT	JAM: 5;5.8 MEG: 7;10.2	72ptJ&MAUG03.cha
9	15.08.03	19.42	MEG, JAM, PAI, MOT	JAM: 5;5.14 MEG: 7;10.8	74ptJ&MAUG03.cha

Das 9 conversas telefônicas, totalizando 1 hora e 47 minutos, apenas duas não tinham como interlocutor o pai (PAI): conversa número 1 ocorreu entre JAM e um amigo dele, VIN (4.29 minutos) e a conversa número 8 teve como interlocutores o avô brasileiro (VOV) e a prima brasileira (SAR) (2.20 minutos). Os nomes dos arquivos indicam que o português (pt) é a língua principal das conversas, e apenas o material em inglês foi marcado (<material em inglês>[@en]). Os turnos dos interlocutores localizados no Brasil não foram transcritos, no entanto, para indicar o turno destes falantes, o símbolo **www** foi inserido nas linhas principais. No cabeçalho **@Warning** de cada arquivo, consta que a inserção dos turnos pode ser incorreta, mas para fins desta análise específica, isso não influencia os resultados.

4.2. Primeira análise dos arquivos

4.2.1. Construção da linha de comandos

Com o objetivo de investigar e comparar o uso que os dois sujeitos fizeram das duas línguas com a variável 'participante', foram realizadas as seguintes análises: através do KVAL e FREQ foram pedidas listas de frequência das palavras dos dois sujeitos por interlocutor e por língua. Os exemplos das linhas de comando reproduzidas abaixo mostram como foram construídos os comandos:

Exemplo 1:

```
kwal @ +t%add +t*JAM +s"PAI" +d +u | freq +s"<@en>" -s"*pn"+o
```

Dividida em duas partes, essa linha de comando é composta de vários itens que especificam a seleção dos dados a serem analisados e a produção dos resultados. Na primeira parte, KVAL seleciona os enunciados que JAM (+t*JAM) dirigiu ao seu pai (+t%add e +s"PAI") e faz uma análise de todos os nove arquivos selecionados (+u), isto é, não faz a análise por arquivo. Os resultados são despojados de informações extras (+d) e mandados automaticamente para uma segunda análise, realizada por FREQ (| freq). Nessa segunda parte, FREQ realiza uma análise apenas no material em inglês (+s"<@en>") encontrado nos dados resultantes da primeira análise e exclui os nomes próprios marcados (-s"*pn"). As palavras que compõem a lista final aparecem em ordem de frequência (+o).

Exemplo 2:

```
kwal @ +t%add +t*JAM +s"PAI" +d +u | freq -s"<@en>" -s"*pn" +o
```

A única diferença entre essa linha de comando e àquela do primeiro exemplo é o uso de - (sinal de menos) no lugar de + (sinal de mais) imediatamente

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

antes do item `s"<@en>`" na segunda parte da linha. Isso significa que, desta vez, o material em inglês será excluído e a lista será composta somente das palavras em português.

Exemplo 3:

`kwal @ +t%add +t*JAM +s"MOT" +d +u | freq +s"<@en>" -s"*pn" +o`

Essa linha diferencia-se da do exemplo 1, porque desta vez são selecionados os enunciados direcionados à mãe (`+s"MOT"`) em vez de ao pai (`+s"PAI"`).

Exemplo 4.

`kwal @ +t%add +t*MEG +s"PAI" +d +u | freq +s"<@en>" -s" *pn" +o`

Neste exemplo, a análise é realizada com os enunciados da MEG (`+t*MEG`) dirigidos ao pai, contendo material em inglês.

Como ficou claro nos exemplos acima, apenas três itens da linha de comando são trocados: o falante (`+t*xxx`), o interlocutor (`+s"xxx"`) e a língua (`+/-s"<@en>"`). Mantendo a mesma estrutura básica da linha do comando e utilizando a cada análise uma combinação diferente dos três itens mencionados acima, foi realizado um total de 24 análises, cujos resultados produzidos serão discutidos na próxima seção.

4.2.2. Resultados gerais

Para ter uma visão geral do papel que o português e o inglês representam para cada falante em relação aos seus interlocutores, foi feito um levantamento do número total de *tokens* (palavras) de cada lista produzida. Os resultados estão sintetizados nas Tabelas 2, 3 e 4 abaixo.

Na Tabela 2, que mostra os resultados do sujeito JAM, é possível constatar que, em termos gerais, cada língua tem um papel definido, estritamente ligado à condição monolíngüe ou bilíngüe do interlocutor.

Tabela 2: Número total e porcentagem de palavras por língua e por interlocutor do falante JAM, em nove arquivos do LOBILL Corpus.

Língua	Interlocutores									
	PAI		VOV		VIN		MOT		MEG	
Português	1531	91%	18	100%	258	100%	20	4%	23	92%
Inglês	155	9%	0	0%	0	0%	431	96%	2	8%
Total	1686		18		258		451		25	

Embora a variação entre os números totais de palavras dirigidas a cada interlocutor não permita uma comparação totalmente confiável, os resultados mostram que, ao interagir com o avô brasileiro (VOV) e o amigo VIN (nascido na Suécia e bilíngüe em sueco e português), JAM só fez uso do português. Enquanto que nenhuma palavra em inglês foi encontrada nas interações de JAM com esses interlocutores, nas conversas com os demais (pai, mãe e irmã) percebe-se o uso do inglês. Com o pai (PAI), o uso do português predomina, mas quase 10% do total de 1686 palavras foram em inglês, indicando a presença do uso de alternância de código. Com a mãe (MOT), o uso das línguas inverteu-se: de um total de 451 palavras, 96% são palavras em inglês e 4%, em português, mostrando que ao interagir com a mãe o inglês faz o papel da Língua Matriz. Com a irmã, apesar do número reduzido de palavras, os resultados indicam que a língua dominante é o português, mas que existe também a presença de alternância de código. Tudo indica que JAM considera o avô brasileiro e o amigo (bilíngüe em sueco e português) falantes com os quais a estratégia de alternância de código não pode ser utilizada, por eles não dominarem o inglês. O fato de a alternância de código ser empregada na interação com os falantes bilíngües (pai, mãe e irmã) demonstra que JAM é consciente (de forma inconsciente!) da sua escolha em usar uma ou duas línguas com um determinado falante. A natureza estrutural e pragmática do uso da alternância de código por JAM será o assunto da seção 4.3.2.

Ao examinar a Tabela 3, que contém os dados de MEG, vemos que os resultados assemelham-se aos da Tabela 2 (de JAM).

Tabela 3: Número total e porcentagem de palavras por língua e por interlocutor do falante MEG em nove arquivos do LOBILL Corpus.

Língua	Interlocutores									
	PAI		VOV		SAR		MOT		JAM	
Português	3053	96%	23	92%	12	100%	4	3%	4	80%
Inglês	127	4%	2	8%	0	0%	117	97%	1	20%
Total	3180		25		12		121		5	

Com o pai e o irmão, o português predomina (96% e 80%, respectivamente) e com a mãe, o inglês predomina (97%), mas em todos os casos há presença de material em inglês, indicando o uso de alternância de código. Em relação à prima SAR, não constam palavras em inglês, mas o número de palavras é muito reduzido para se chegar a uma conclusão: outras análises de interações entre MEG e SAR precisariam ser realizadas. Com o avô brasileiro, 2 do total de 25 palavras são palavras em inglês, o que indica o uso, embora mínimo, de alternância de código. Na seção 4.2.3., examinaremos essas palavras em mais detalhes.

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

Considerou-se interessante incluir os resultados das análises realizadas em relação ao falante MOT, que estava sempre presente nas conversas telefônicas e interagindo com os dois sujeitos. Como podemos ver na Tabela 4, a mãe utilizou inglês na maior parte do tempo, mas usou o português, especialmente ao interagir com JAM (13%).

Tabela 4: Número total e porcentagem de palavras por língua e por interlocutor do falante MOT em nove arquivos do LOBILL Corpus

Língua	Interlocutores			
	JAM		MEG	
Português	31	13%	2	2%
Inglês	200	87%	86	98%
Total	231		88	

Nos dois casos, há evidência de alternância de código, cuja natureza será investigada e discutida na seção 4.2.3.

Os números totais de palavras na Tabela 4 indicam que a mãe interagiu mais com JAM que com MEG. Ao examinar nas Tabelas 2 e 3 os totais de palavras que os dois sujeitos direcionaram à mãe, também encontramos indicação de que JAM interagiu mais com a mãe (451 palavras) se comparado à irmã (121 palavras). Ao calcular a porcentagem que estes totais representam do total geral de palavras de cada falante (2438 no caso do JAM e 3343 no caso da MEG), constatou-se que o número de palavras dirigidas à mãe por JAM representou 18,4% do total, muito mais do que a porcentagem de 3,6% de MEG. Isto é, durante as conversas telefônicas com o pai (e os outros interlocutores), em comparação a MEG, JAM interagiu muito mais com a mãe. Esta diferença entre os dois sujeitos foi investigada e os dados revelaram uma relação entre os sujeitos e o uso de alternância de código, o que será discutido nas seções a seguir.

4.2.3. Resultados mais específicos

Na discussão dos resultados mostrados acima, ficou evidente que, para os dois sujeitos, o português é a língua utilizada para interagir com falantes considerados 'monolíngües' (em português) e que a alternância de código apenas ocorre na interação com falantes bilíngües (mãe, pai e irmã). Nesta seção, o objetivo é examinar as listas de palavras produzidas pelo FREQ, buscando o que elas podem revelar sobre a natureza da alternância de código, fenômeno detectado pela análise das palavras em termos gerais.

As tabelas apresentadas nessa seção contêm as listas de palavras fornecidas pelas análises realizadas com KWAL e FREQ (veja 4.2.1). Por razões práticas, apenas as primeiras 20 palavras das listas foram selecionadas para análise e discussão. Os interlocutores aos quais não foi dirigida nenhuma palavra em inglês foram excluídos das tabelas.

Tabela 5: Listas de palavras em inglês em ordem de frequência produzidas por JAM e MEG ao interagir com os interlocutores (primeiras 20 palavras).

JAM			MEG			
PAI	MOT	MEG	PAI	VOV	MOT	JAM
10 train	19 i	1 bell	6 e+ mail	1 hey	9 he	1 yes
8 tram	17 to	1 hole	4 library	1 yeah	4 and	
6 **	15 how		4 no		4 mummy	
6 and	15 we		4 triceratops		4 the	
4 bath	14 in		3 and		4 what	
4 but	13 Portuguese		3 dictionary		3 after	
4 live	13 say		3 he		3 her	
4 train+ track	13 want		3 rock		3 him	
3 because	12 do		3 the		3 in	
3 digger	11 "		3 tram		3 says	
3 I	9 is		2 bell		3 to	
3 I'm	9 mummy		2 french		2 does	
3 lightening	9 the		2 oink		2 else	
3	8 a		2 p@l		2 get	
swimming+ po	8 you		2 river		2 just	
ol	6 bath		2 station		2 mother	
2 's	6 don't		2 then		2 said	
2 all	6 it		2 they		2 she	
2 boyfriend	6 which		2		2 something	
2 buffer	5 have		wildlife+ park		2 what's	
2 don't			1 "			
2 England						

* Veja seguinte página para explicações sobre o código " (aspas).

Essas palavras serão examinadas mais adiante nos seus contextos, através do uso do comando KWAL. Entretanto, as listas mostradas nas tabelas acima já indicam vários fatos em relação ao uso de alternância de código pelos dois sujeitos.

Na seção 3, foram apresentados os conceitos de Língua Matriz (a língua que contribui com mais estrutura morfossintática para as sentenças bilíngües) e Língua Encaixada (caracterizada pela contribuição de morfemas de conteúdo como substantivos, verbos, adjetivos e preposições). Ao examinar as listas de palavras dos sujeitos que foram dirigidas ao pai, encontramos uma dominância de substantivos e muito poucos morfemas gramaticais: das primeiras vinte palavras mais freqüentes de cada sujeito, 11 são substantivos. Fica evidente que o inglês exerceu o papel de Língua Encaixada nas interações com o pai. Ao exami-

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

nar as listas de palavras dirigidas à mãe, encontramos uma situação oposta: uma dominância de morfemas gramaticais e apenas 3 substantivos (**Portuguese**, **mummy** e **bath**) no caso de JAM e dois (**mummy** e **mother**) no caso de MEG. Isso confirma que o inglês exerceu o papel de Língua Matriz nas interações entre os sujeitos e sua mãe. A interação entre os irmãos foi pouca, mas as palavras em inglês ficaram reduzidas a substantivos (**bell** e **hole** no caso de JAM) e um marcador discursivo (**yes** no caso da MEG). As duas palavras em inglês dirigidas por MEG ao avô também ficaram na categoria de marcadores discursivos (**hey** e **yeah**), que é freqüentemente citada como um ponto de alternância de código. Conjunções formam outra categoria de palavras que, freqüentemente, marcam o ponto de alternância entre as línguas participantes. Entre as palavras dirigidas ao pai, foram encontrados **and** nas listas dos dois sujeitos e **but** e **because** na lista de JAM. A análise dessas conjunções nos seus contextos maiores (com KWAL) confirmará, ou não, a presença de alternância de código intrassentencial nessas fronteiras.

Outro aspecto que foi revelado pela lista de palavras de JAM em relação à mãe, diz respeito a uma das razões principais da necessidade de sua interação com a mãe durante as conversas com pai. Entre as primeiras dez palavras mais freqüentes, encontram-se **I**, **to**, **how**, **we**, **in**, **Portuguese**, **say**, do e o símbolo “ (aspas) que marca uma referência metalingüística. Ao reorganizar essas palavras, podemos formar a pergunta **How do I/we say “ in Portuguese?**, que, pela freqüência das palavras, deve ter sido utilizada mais de dez vezes por JAM. Isso pode ser confirmado pela análise das palavras em contexto através do KWAL.

A Tabela 6, abaixo, mostra as primeiras vinte palavras mais freqüentes em português que JAM e MEG produziram ao interagir com os pais. Ao comparar as listas dirigidas ao pai, é interessante notar que a maioria é de morfemas gramaticais, sendo eles os mesmos. Isso indica que os dois sujeitos fizeram uso muito semelhante do português ao conversar com o pai ao telefone. É claro que seria necessário comparar as listas inteiras para chegar a alguma conclusão, mas essa amostra indica que, além de atuar como a Língua Matriz, a natureza morfossintática do português empregado pelos dois sujeitos é muito similar.

Tabela 6: Listas de palavras em português em ordem de freqüência produzidas por JAM e MEG ao interagir com os pais (primeiras 20 palavras).

JAM		MEG	
PAI	MOT	PAI	MOT
110 é	2 de	155 e	2 um
84 um	2 o	123 a	1 cachorro
70 não	2 quando	115 que	1 depois
68 que	2 terminou	104 o	
63 eu	1 assistir	81 um	
53 a	1 caiu	81 é	
45 e	1 estava	75 para	
32 o	1 falar	73 eu	
32 tem	1 ligou	57 de	
31 gente	1 noite	57 está	
28 só	1 nome	50 não	
27 mas	1 qual	43 tem	
26 de	1 quê	37 aí	
23 para	1 ratinhos	36 ele	
20 estava	1 scooby+doo	36 tu	
17 no	1 trilho	34 com	
16 também		34 na	
15 aí		33 gente	
15 pai		31 ela	
14 ele		30 estava	

As listas de palavras dirigidas à mãe são muito reduzidas, indicando o uso muito limitado de português ao interagir com esse interlocutor. A natureza desse uso só pode ser examinada pelo estudo das palavras nos seus contextos (veja seção seguinte).

Embora as listas mostradas nas duas tabelas acima contenham palavras fora dos seus contextos, pela discussão nessa seção ficou evidente que uma análise da freqüência das palavras pode ser muito útil para uma investigação preliminar dos papéis que cada língua exerce na interação entre duas crianças bilíngües e seus interlocutores. Pelo estudo da natureza das palavras e suas freqüências foi possível identificar, nas interações, a presença de uma Língua Matriz e uma Língua Encaixada, variando de acordo com o interlocutor. A presença de material de ambas as línguas nas listas de JAM e MEG indica o uso de alternância de código com certos interlocutores, a saber, o pai, a mãe e a irmã. Em relação ao pai, a freqüência de palavras de conteúdo em inglês indica que a alternância de código empregada pelas duas crianças é do tipo intrassentencial, com a inserção de substantivos em inglês na estrutura morfossintática do português (a Língua Matriz). Em relação à mãe, o número reduzido de palavras para estudo

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

dificulta a identificação do tipo de alternância de código utilizado pelos dois sujeitos, o que torna necessário o estudo das palavras em contexto, assunto da próxima seção.

4.3. Segunda análise dos arquivos

4.3.1. Construção das linhas de comandos

Com o objetivo de analisar a natureza da alternância de código utilizada pelos dois sujeitos, foram realizadas seis análises através de KWAL. Como foi identificada a presença significativa de alternância de código nas interações de JAM ou MEG com seus pais, foram construídas as linhas de comandos mostradas a seguir¹²:

1. kwal @ +t%add +t*JAM +s"PAI" +d +u | kwal +s"<@en>" +d
2. kwal @ +t%add +t*JAM +s"MOT" +d +u | kwal -s"<@en>" +d
3. kwal @ +t%add +t*MEG +s"PAI" +d +u | kwal +s"<@en>" +d
4. kwal @ +t%add +t*MEG +s"MOT" +d +u | kwal -s"<@en>" +d
5. kwal @ +t%add +t*MOT +s"JAM" +d +u | kwal -s"<@en>" +d
6. kwal @ +t%add +t*MOT +s"MEG" +d +u | kwal -s"<@en>" +d

A primeira parte de cada linha de comando é basicamente igual àquela explicada em 4.2.1. Apenas trocam-se os códigos dos falantes (**+t*JAM**, **+t*MEG** e **+t*MOT**) e os interlocutores (**+s"MOT"**, **+s"PAI"**, **+s"JAM"**, **+s"MEG"**). Na segunda parte da linha, **FREQ** é substituído por **KWAL** e **+o** por **+d**. Isso significa que no lugar de listas de palavras, os resultados são apresentados em forma de concordâncias, e os enunciados são despojados de informações extras (**+d**). Foi eliminado o item **-s"*pn"** da linha do comando por uma razão importante: ao fazer testes de análises descobriu-se que, ao pedir KWAL para eliminar nomes próprios, qualquer enunciado que continha um nome próprio marcado com **@pn**, era excluído. Isso significa que além de excluir os nomes próprios, KWAL exclui outro material importante para ser analisado. Portanto, esse item não foi inserido na linha de comando.

Nas análises 1 e 3, que tratam de interações entre os dois sujeitos e o pai, foram selecionados os enunciados que contêm material em inglês (**+s"<@en>"**), isto é, que mostram alternância de código para inglês (a Língua Encaixada nesse

¹² As linhas de comando foram numeradas para facilitar a referência textual, os números não fazendo parte das linhas em si.

caso). Nas análises 2, 4, 5, e 6, que tratam de enunciados entre as crianças e a mãe, apenas enunciados que contêm material em português (a Língua Encaixada nesse caso) foram selecionados (-s"<@en>") uma vez que a Língua Matriz é o inglês. Por questões de espaço, foi necessário fazer uma seleção de alguns dos resultados produzidos pelas análises, mas os exemplos apresentados na próxima seção foram considerados representativos do todo.

4.3.2. Resultados das seis análises

Os resultados obtidos das análises através de KWAL serão discutidos de acordo com cada sujeito: primeiro, as três análises relacionadas a JAM (1, 2 e 5) e, depois, em relação a MEG (3, 4 e 6). Para facilitar a discussão dos resultados, as concordâncias foram numeradas (esse número não faz parte da concordância em si).

4.3.2.1. JAM e a alternância de código com a mãe

Foi visto em 4.2.2. que, ao interagir com a mãe, JAM utiliza inglês na maior parte do tempo (431 palavras de um total de 451), mas que existe a presença de material em português, embora pouco (20 palavras). Ao pedir KWAL para localizar esse material, os resultados apresentados foram os seguintes:

- (1) *JAM: <Vincent@pn said which[*]>[@en] Sara@pn caiu .
- (2) *JAM: <he's going to>[@en] assistir Scooby+Doo .
- (3) *JAM: <<but if we>[/] <if we>[/] if Meggie@pn>[@en] terminou, <then you call me when Meggie@pn>[@en] terminou de falar, <okay>[@en]?
- (4) *JAM: qual nome <which[*] we went>[@en]?
- (5) *JAM: <quando o>[/] <daddy he's talking[*] if[/] if Meggie@pn want[*] to talk for>[@en] +/.
- (6) *JAM: <(be)cause yesterday>[@en] quando estava de noite <I didn't have a bath>[@en] .
- (7) *JAM: o quê?
- (8) *JAM: <was saturday which we[/] Vincent@pn, he>[@en] ligou <to here>[@en]?
- (9) *JAM: <how do you say>[@en] ratinhos?
- (10) *JAM: trilho["].

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

A primeira observação a ser feita é que o tipo de alternância de código utilizado por JAM ao interagir com a mãe é intrassentencial: dentro do mesmo enunciado (que aqui equivale à unidade ‘sentença’), encontramos material das duas línguas. Em quatro dos exemplos, a alternância de código trata da simples inserção de verbos em português (**caiu**, **assistir**, **terminou** e **ligou**) na estrutura do inglês, uma contribuição típica da Língua Encaixada em sentenças bilíngües (Myers-Scotton, 2002:62). Em dois casos, a alternância para o português envolve a conjunção **quando** (**quando** o e **quando estava de noite**); novamente um ponto de alternância freqüentemente encontrado em discurso bilíngüe. O uso pragmático de alternância de código se revela nos últimos dois enunciados: no primeiro, JAM pede a tradução da palavra **ratinhos** e, no segundo, o símbolo [“] indica que ao dizer trilha o sujeito está fazendo uma referência metalingüística à palavra. Embora cada exemplo de alternância de código mostrado nos resultados mereça uma discussão detalhada à luz dos modelos desenvolvidos por teóricos como Myers-Scotton (2002), aqui fica limitado a uma observação geral em relação à alternância de código utilizada por JAM ao interagir com a mãe: pelas concordâncias produzidas por KWAL, é possível constatar que o inglês exerce o papel da Língua Matriz e que a natureza das alternâncias para português é típica da contribuição da Língua Encaixada.

Ao examinar as concordâncias que mostram o material em português produzido pela mãe ao interagir com JAM (veja análise 5), percebemos logo que a alternância de código empregada é de um tipo bem específico. Como pode ser visto nos resultados apresentados abaixo, quase todos os casos tratam de referências metalingüísticas, indicadas pelo símbolo [“].

- (11) *MOT: **bonde**[“] .
- (12) *MOT: **bonde**[“] .
- (13) *MOT: **bonde**[“] .
- (14) *MOT: **no** +...
- (15) *MOT: **sshh** .
- (16) *MOT: **<buraco grande>**[“] .
- (17) *MOT: **<bucket[“]>**[@en], **balde**, **balde** .
- (18) *MOT: **madeira**[“] .
- (19) *MOT: **proteger**[“] .
- (20) *MOT: **<assistindo um filme>**[“] .
- (21) *MOT: **<I think it's>**[@en] **trem+a+vapor**[“] .
- (22) *MOT: **<eu te amo>**[“] .
- (23) *MOT: **rato**[“] .
- (24) *MOT: **ahh**.

- (25) *MOT: sair["].
- (26) *MOT: frio["].
- (27) *MOT: trem+a+vapor["] vapor["].
- (28) *MOT: trilho["].
- (29) *MOT: trilho["].
- (30) *MOT: banheira["].
- (31) *MOT: banheira["].
- (32) *MOT: <tomar um banho>["].

O uso desse símbolo nas transcrições para marcar referências metalingüísticas foi fundamental. Se não houvesse esse tipo de marcação, ao ser apresentado com essa lista de concordâncias, o pesquisador poderia chegar a conclusões errôneas sobre o uso do português pelo falante. Fica evidente que as palavras e frases em português produzidas pela mãe são respostas a vários pedidos de JAM do tipo **How do I say XXX in Portuguese?**, que foi constatado em 4.2.3. Como será visto na próxima seção, esse uso pragmático do português não foi encontrado nos resultados da análise 6 (enunciados da mãe dirigidos à filha), o que indica um contraste na necessidade de cada criança de fazer esse tipo de alternância de código.

4.3.2.2. JAM e a alternância de código com o pai

Ao interagir com o pai ao telefone, constatou-se que 91% de um total de 1686 palavras foram de material em português. Após examinar a natureza das palavras encontradas nas listas produzidas por FREQ, conclui-se que, enquanto com a mãe inglês era a Língua Matriz, com o pai português exercia esse papel. E inglês foi claramente identificado como a Língua Encaixada. Ao pedir a seleção de todos os enunciados contendo material em inglês dirigidos ao pai (análise 1), uma lista de 87 concordâncias foi apresentada. Embora não seja possível reproduzir e analisar cada concordância, foram selecionadas 22 que mostram a natureza variada das inserções de inglês encontrada nos dados. Os exemplos foram agrupados para melhor exemplificar as categorias do material inserido e o primeiro grupo representa os casos mais comuns: a inserção de palavras de conteúdo, especialmente substantivos:

- (33) *JAM: mas só eu sou muito <lucky>[@en] que eu peguei um[/] um <train+track>[@en] e a gente comp(rou) +/.
- (34) *JAM: e um <cutting+grass+machine>[@en] também .
- (35) *JAM: eu tenho um <digger>[@en], um <digger>[@en] também e um <train>[@en] também .

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

- (36) *JAM: aí a gente ficou # # <five>[@en] dias lá .
(37) *JAM: não tinha <lightening>[@en] não, só trovão .
(38) *JAM: e eu[/] eu <found>[@en] um buraco grande, muito grande.
(39) *JAM: era <shiny, yeah>[@en] .
(40) *JAM: não, <a gente>[///] eles foi[*] pra[: para a] <Grandad@pn
's house>[@en] .
(41) *JAM: e tem um <swimming+pool>[@en] lá que a gente +/.
(42) *JAM: Mas o <North@pn>[@en] é o <England>[@en] também.

É importante lembrar que, ao conversar com o pai pelo telefone, JAM encontra-se na Inglaterra de férias. Exposto a novas experiências e novos contextos, muitas vezes as palavras em inglês inseridas no português coincidem com objetos ou conceitos que não faziam parte da sua realidade no contexto brasileiro. É o caso de palavras como **train+track**, **digger** e **cutting+grass+machine**. E, embora haja evidência de que JAM pedia à mãe a tradução para português de várias palavras em inglês, enquanto falava com o pai, fica evidente pelos exemplos que o sujeito não fazia questão de eliminar totalmente da interação o uso de inglês, especialmente em relação às palavras de conteúdo.

Em alguns casos, JAM inseriu mais do que palavras isoladas do inglês, como mostram os dois exemplos seguintes:

- (43) *JAM: <Uncle+William@pn>[@en] (es)tá <sleeping in another bed>[@en] .
(44) *JAM: não, <ele não <live>[@en]>[///] <she don't live with he(r)[/] her boyfriend>[@en] .

No primeiro exemplo, JAM está explicando ao pai que o tio veio passar alguns dias e que não vai dormir na sala (como o pai imaginava). Tendo já conversado sobre esse assunto com a mãe e a avó em inglês¹³, evidentemente ficou mais fácil mudar para inglês. No segundo exemplo, JAM está explicando sobre o namorado da tia, que não mora com ela (e a avó inglesa). Embora iniciando em português, após inserir **live**, JAM faz uma reformulação que envolve uma mudança total para inglês. Novamente parece que, nesse caso, JAM teve mais facilidade de se expressar em inglês e considerou (consciente ou inconscientemente) que seu pai não ia ter dificuldades em entendê-lo.

¹³ As anotações de campo contêm essa informação.

Em relação à inserção de conjunções em inglês, outro grupo que se destacou nos enunciados bilíngües, as concordâncias revelaram um fato interessante. Pela transcrição dos exemplos mostrados abaixo, é possível notar, que logo após a inserção de uma conjunção em inglês, JAM procura reformular seu discurso (indicado pelo símbolo [///]) e insere o equivalente em português:

- (45) *JAM: <but>[@en][///] mas, só que <o deles>[//] eles tem um debaixo e um de cima .
- (46) *JAM: não, <bec(ause)>[@en][///] porque <não tem>[/] não tem bateria[*] não.
- (47) *JAM: <and>[@en][///] e cavou muito que o[///] peguei o[///] a gente trouxe o +...

Isso também ocorre com frequência quando JAM confirma algo dito pelo pai, como mostra o seguinte exemplo, no qual o sujeito diz *yes* e imediatamente depois é:

- (48) *JAM: <yes>[@en][///] é, mas quando eu fui pro[: para o][/] pro[: para o] outro[///] pra[: para a] areia <I can very>[@en] muito muito.

Embora seja evidente que nesses exemplos a palavra em inglês é ativada primeiro (está na ‘ponta da língua’), JAM a substitui pelo equivalente em português, demonstrando uma ‘preocupação’ em falar a língua do pai. Até certo ponto essa atitude parece contradizer as ocasiões em que JAM demonstrou uma ‘falta’ de preocupação em falar apenas português (veja exemplos 33 a 42 acima). Entretanto, os casos das conjunções citadas acima parecem se enquadrar na categoria de ‘lapsos’, em que os equivalentes são fácil e rapidamente acessados, enquanto a inserção de, por exemplo, um substantivo em inglês, pode ocorrer por falta de um equivalente adequado em português ou por a palavra ser, momentaneamente, de mais difícil acesso que a equivalente em inglês. De fato, por JAM estar submerso num contexto inglês, as palavras inglesas estão mais ativadas que as portuguesas e um atraso no acesso a estas, portanto, pode ser natural.

As concordâncias mostradas a seguir revelam um uso bem específico de alternância de código que envolve referências metalingüísticas:

- (49) *JAM: tu sabe que é um <tram[“]>[@en]?
- (50) *JAM: é um banheiro mas não é um chuveiro, é um <bath, bath>[@en].

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

(51) *JAM: tu sabe que é um <buffer[""]>[@en]?

Ao fazer a alternância de código para inglês de algumas palavras, JAM demonstra certa consciência de que o pai talvez não conhecesse essas palavras e procura verificar e clarificar seus significados. Ao conferir a lista das palavras em português produzidas pela mãe (veja exemplos 11 a 32), fica evidente que, nesses casos, JAM acabou pedindo a tradução à mãe. Nos primeiros dois exemplos, a mãe forneceu **bonde** e **banheira**, mas não sabia a tradução de **buffer** ('pára-choque'), o que JAM tentou explicar para o pai, recorrendo ao inglês novamente:

(52) *JAM: é um negócio que[/] que <keep on rails>[@en].

Precisando pedir ajuda à mãe com a tradução da palavra **rail** (**trilho** aparece duas vezes, marcado como referência metalingüística na lista de concordâncias da mãe), JAM consegue fazer com que o pai entenda o que está querendo dizer.

Pelos exemplos apresentados, representativos da lista de concordâncias produzida por KWAL, em quase todos os casos, a alternância de código empregada por JAM foi do tipo intrassentencial, com a inserção de morfemas de conteúdo. Entretanto, houve uma ocorrência de uma alternância de código intersentencial mostrada na seguinte seqüência:

(53) *JAM: eu disse para mamãe no com(putador)[///] <library>[@en] (es)tá fazendo assim ó disse +”/.

JAM: +” <mummy mummy would you write on the computer <which[] saying>[/] write which[*] I'm missing Daddy and I want to see him and I'm[/] I'm[*] love him too>[@en] +”.

Aqui JAM está explicando ao pai o que falou para a mãe quando ela estava na biblioteca escrevendo em e-mail para o pai. Os símbolos +” e +”. deixam claro que se trata de discurso direto e, nesse caso, JAM escolheu manter a língua original do pedido (inglês).

Embora fossem apresentadas apenas 22 das 87 concordâncias, os exemplos escolhidos revelaram bastante sobre a natureza estrutural e pragmática da alternância de código empregada pelo falante JAM ao interagir com a mãe e o pai. Na próxima seção, examinaremos as concordâncias produzidas pelas análises 3, 4 e 6 que tratam da interação de MEG com seus pais.

4.3.2.3. MEG e a alternância de código com a mãe

Pelas análises de *FREQ*, de um total de 121 palavras direcionadas à mãe, apenas 4 foram em português: **um**, que aparece duas vezes, **cachorro** e **depois**. Ao pedir as concordâncias desse material em português, os resultados foram as seguintes:

- (54) *MEG: <and as well <when he>[//] he said when he comes in he says>[@en] cachorro[""] <and he says she starts to mi+mi+mi[""]>[@en] .
- (55) *MEG: depois um[/] um <pound>[@en]?

Na primeira concordância, MEG está contando à mãe o que o pai disse em relação ao preá dela: quando ele chega do trabalho e abre a porta, ele chama **cachorro** (uma brincadeira de nome já estabelecida) e o preá responde com **mi+mi+mi**. Além de ser utilizado para marcar uma referência lingüística, o símbolo [""] também marca discurso direto quando se trata de uma palavra isolada. Portanto, a alternância de código empregada aqui com a inserção de **cachorro** tem a função específica de citar discurso direto. A segunda concordância contém as outras três palavras em português dirigidas à mãe. Por coincidência, exatamente a mesma frase aparece na lista de concordâncias direcionadas ao pai. Através de outro comando do CLAN, COMBO, que busca grupos de palavras, (nesse caso, **combo @ t*MEG +s"depois^um"**), foi possível localizar com facilidade o arquivo específico dessa concordância e depois examiná-la em seu contexto maior (isto é ler a parte pertinente da transcrição). Ao fazer isso, constatou-se que, na realidade, esse enunciado foi direcionado tanto ao pai quanto à mãe: MEG está contando ao pai sobre a compra de um conjunto de dicionários cujo preço tinha sido reduzido de trinta libras (pounds) para sete libras e **depois um pound**. Ao dizer esse último valor para o pai, MEG olha para a mãe procurando confirmação do valor e esta responde **yes**. Portanto, podemos considerar que a mãe foi um interlocutor indireto e que a decisão de MEG de utilizar português como Língua Matriz foi baseada no seu interlocutor principal (o pai).

Em termos de números totais de palavras, MEG interagiu muito menos com sua mãe (121) quando comparado com JAM (451), o que dificulta uma comparação confiável da extensão do uso de alternância de código pelas duas crianças. Ao examinar os resultados de *KWAL* em relação ao português utilizado pela mãe ao interagir com a filha, constatou-se uma única concordância, mostrada a seguir:

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilingüe Infantil*.

(56) *MOT: bem rapidinha Meggie@pn .

Não houve ocorrências de referências metalingüísticas em português, como foi visto no caso da interação entre a mãe e JAM. Isso indica que MEG não pediu a tradução para o português de palavras inglesas. Talvez ela considerasse desnecessário ter recurso à tradução (porque o pai compreendia as palavras em inglês utilizadas) ou talvez ela não precisasse utilizar essa estratégia (por não ter dificuldade em se expressar em português). Na próxima seção, a análise das concordâncias contendo material em inglês dirigidas ao pai, revelará mais sobre as diferenças no uso de alternância de código pelos dois sujeitos.

4.3.2.4. MEG e a alternância de código com o pai

Pelos resultados de FREQ foi visto que, ao interagir com o pai, MEG utilizou 127 palavras em inglês, de um total geral de 3053 palavras (representando 4%). Ao examinar as primeiras vinte palavras da lista produzida por FREQ, constatou-se a presença de muitos substantivos e poucas palavras gramaticais. Ao pedir a lista de concordâncias do material em inglês através de KWAL, os resultados mostram que, na grande maioria dos casos, a alternância de código trata-se da inserção de palavras inglesas isoladas na estrutura do português. De um total de 71 concordâncias, 59 contêm uma ou, às vezes, duas palavras isoladas e dessas 59 concordâncias, 56 contêm substantivos. Alguns exemplos típicos podem ser vistos a seguir:

- (57) *MEG: e também eu[/] eu fui com umas botas que eu comprei no <car+boot+sale>[@en] e fui lá no[/] no <river>[@en] num cantinho que tem água e fiquei andando .
- (58) *MEG: hoje <de man(hã)>[/] de[/] de tarde quando a mamãe foi pro[: para o] <library>[@en] para mandar um <e+mail>[@en] para tu, sabe o que que ela trouxe?
- (59) *MEG: também eu ganhei o sinal que tem o <bell>[@en] que tem no <tram>[@en] .
- (60) *MEG: um, comprou por noventa e nove <p@l[/] pence>[@en] .
- (61) *MEG: ah, e hoje na praia a gente comprou um <rock>[@en] .
- (62) *MEG: quando eu estava lá no <station, train+station>[@en] +...
- (63) *MEG: aí sabe que[/] que o tiranosauo+rex estava comendo o <triceratops>[@en] .
- (64) *MEG: tinha tudo preto, e estavam[/] quando James@pn corria, eles corriam embaixo desta[/] deste bush>[@en] .

- (65) *MEG: ei pai, quando o James@pn estava falando de cair da cama quando eu estava deitando para dormir eu e a mamãe na cama dela, o[/] o Jame(s)@pn caiu da cama e caiu de[/] de costas na[//] no <bricks>[@en] .
- (66) *MEG: era erm aquele que passa no filme <que tem aquela>[///] que parece <triceratops>[@en] só que não é afiado a[/] a[/// /] <the horn>[@en] .

A ausência de símbolos que marcam hesitações (#) e marcadores de discurso como **erm**, indica que, pelo menos nestes casos, MEG empregou esse tipo de alternância de código com a maior naturalidade. Ao examinar a natureza das palavras, encontramos algumas que são estritamente ligadas ao contexto experiencial em que MEG se encontrava: **car+boot+sale**, **pence**, **rock**, **river**, **library**, **bell**, **tram**, **train+station**, **bush**. É importante salientar que tal afirmação é possível devido ao fato de a pesquisadora encontrar-se na posição de um observador participante, compartilhando as experiências dos sujeitos. Para outro pesquisador poder fazer essa mesma afirmação seria necessário incluir esse tipo de informação extralingüística numa linha dependente.

Em relação à natureza estrutural da alternância de código utilizada por MEG, os dados nos revelam uma coisa interessante: com exceção do último exemplo, o substantivo em inglês sempre vem acompanhado pelo artigo em português e fica evidente que a preferência do falante é utilizar o artigo masculino: **no car+boot+sale**, **no library**, **o bell**, **um rock**. Nos exemplos (64) e (65), essa preferência fica ainda mais clara: apesar de inicialmente utilizar o artigo feminino, MEG imediatamente faz uma reformulação e o substitui pelo artigo masculino (**desta[//] deste bush** e **na[//] no bricks**). Parece que a maneira que MEG encontrou de lidar com a falta de gênero dos substantivos ingleses é de generalizar, inserindo o artigo masculino mesmo no caso de palavras cujas traduções em português requereriam o artigo feminino, como por exemplo, **library** – ‘biblioteca’ e **station** – ‘estação’. Ao re-examinar as concordâncias referentes a JAM, foi encontrado o mesmo fenômeno. Em relação à exceção (o último exemplo da lista acima), MEG utiliza o artigo feminino, mas, após repetir uma vez, insere o substantivo inglês junto com o artigo definido em inglês: **...afiado a[/] a[///] <the horn>[@en]**. Talvez a inserção do **the** fosse uma estratégia para evitar um encontro de sons que soasse meio desajeitado (**a** com **horn**) ou talvez MEG quisesse evitar uma possível ambigüidade de sentido: sem a inserção do **the**, **a**, seguido por **horn**, podia ser entendido tanto como ‘um’ (o artigo indefinido inglês) quanto ‘a’ (artigo definido em português). Pelo enunciado, fica claro que o sentido desejado é este último e, portanto, talvez para

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

evitar a construção de um enunciado potencialmente incoerente, MEG decide utilizar **the**.

Embora a grande maioria das palavras inglesas inseridas de forma isolada fosse de substantivos, as concordâncias abaixo mostram outros tipos de palavras isoladas inseridas (no máximo duas ocorrências de cada):

- (67) *MEG: <then>[@en] sete .
- (68) *MEG: é, dois dias, <no>[@en] duas noites .
- (69) *MEG: <yeah>[@en], uma grade .
- (70) *MEG: <fine>[@en] .
- (71) *MEG: tinha dois <little>[@en] ratinhos ali .
- (72) *MEG: ela nas escadas <already>[@en] .

Dessas poucas ocorrências, os últimos dois exemplos indicam um uso mais marcado de alternância de código: a inserção de **little** e **already** parece ter a função de enfatizar o que o falante está contando. Entretanto, como já foi dito, as ocorrências de concordâncias contendo esses tipos de palavras foram tão poucas que quaisquer conclusões seriam especulações.

As concordâncias que evidenciaram a inserção de mais que palavras isoladas podem ser divididas em dois grupos. Em um grupo, encontram-se 4 concordâncias que mostram a inserção de frases curtas em inglês no português, reproduzidas a seguir:

- (73) *MEG: mas o James@pn, ele é maluco, se pulou no mar, <to the sea>[@en].
- (74) *MEG: aí a gente vai vir essa semana que vem (por)que o <library>[@en] está fechado <friday, saturday and sunday>[@en].
- (75) *MEG: <then it was>[@en] +...
- (76) *MEG: aí a <Janet@pn>[@en] conseguiu ver <but>[@en][///] mas o <Grandad@pn didn't know>[@en] .

Embora o pesquisador possa apenas especular sobre as razões por que MEG inseriu essas frases curtas, em termos estruturais o que pode ser observado em três das concordâncias é que há coerência entre as estruturas da Língua Matriz e da Língua Encaixada. Novamente é importante salientar que não há evidência suficiente para fazer afirmações concretas em relação a essa coerência entre as línguas. Entretanto, a falta de ocorrências desse tipo de inserção (frases curtas) quando comparada ao número de ocorrências da inserção de palavras isoladas, reafirma

que, na maior parte do tempo, a alternância de código empregada por MEG foi de um tipo bem específico e restrito (de palavras isoladas).

Outro grupo pequeno de 6 concordâncias evidenciou o uso pragmático da alternância de código para citar o discurso direto. As primeiras quatro citações da lista mostrada a seguir tratam de falas de personagens de um seriado inglês *Fawlty Towers*, que as crianças assistiam em vídeo (repetidas vezes) na Inglaterra e que já tinha passado no canal EuroChannel no Brasil (o pai conhecia o seriado). Nos primeiros dois casos, são mostradas em parênteses as falas anteriores à citação:

- (77) ***MEG:** e tu já viu no <Fawly+Towers>[@en] o Manuel@pn?
***MEG:** +” <two dead pigeons in the water+tank>[@en] +”.
- (78) ***MEG:** aí e[/] e na linguagem do Manuel@pn, ele pensa que é porcos
e
[/] <e ele disse>[///] <and he said>[@en] +”/.)
***MEG:** +” <how did they get up there>[@en] +”?
- (79) ***MEG:** +” <they flew>[@en] +”, disse o Basil@pn .
- (80) ***MEG:** +” <oink oink>[@en] +”, e tudo assim .
- (81) ***MEG:** aí o[/] <Jake@pn and Max@pn>[@en] também vieram mas
tu sabe como <Jake@pn>[@en] é +”/.)
***MEG:** +” <why doesn’t he eat his tail>[@en] +”.
- (82) ***MEG:** +” <what on earth is he thinking>[@en] +”.

As últimas duas concordâncias são citações das falas do primo inglês, JAK, e da mãe. Embora esse tipo de alternância de código seja utilizado mais por MEG, quando comparado ao irmão (veja exemplo (53)), sua presença nas interações com pai é significativa: indica a consciência plena da condição bilíngüe do pai. Entretanto, as concordâncias revelam que, apesar de saber que a alternância de código para inglês não cria muito problema para o pai, ambas as crianças mostram o uso de certas estratégias para evitar o seu uso excessivo. Enquanto JAM recorre bastante à ajuda da mãe para traduzir palavras em inglês para o português (veja exemplos (11) a (32)), MEG usa a estratégia de reformulação. A seguir são reproduzidas as cinco concordâncias que revelam essa estratégia:

- (83) ***MEG:** e estava colecionado <err <ca(rds)>[@en]>[///] err tiquete
de trem .
- (84) ***MEG:** tinha o[/] o <whale>[@en][///] a baleia .
- (85) ***MEG:** <do outro <si(de)>[@en]>[///] do outro lado .

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

- (86) *MEG: estava lá no <North@pn>[@en] e <mum(my)>[@en][//]
mamãe estava no <conference>[@en] ela, eu[/] eu saí com
a[/] a <Janet@pn>[@en] e o <Grandad@pn>.
- (87) *MEG: era um banco # # que[/] que estava na[/] em frente do
<river>[@en] e o ratinho <jump(ed)>[@en][//] err pu-
lou na água, aí[/] aí eu vi aí uma coisa tinha pulada na água .

As palavras para as quais MEG forneceu a tradução em português (**cards, whale, side, mummy, jump(ed)**) não são do tipo específico que justifica a necessidade de uma tradução. Portanto, essa estratégia de reformulação indica que MEG, talvez por empatia com o pai, estava disposta a limitar o uso de inglês quando possível, isto é, quando o equivalente em português fosse facilmente recuperado. No caso de palavras mais específicas ou mais recorrentes no ambiente de convívio imediato (exemplos (57) a (66)), essa estratégia foi raramente utilizada: o desconhecimento do equivalente em português, ou o fato de ser de difícil acesso no momento da fala, levou à inserção liberada de substantivos em inglês.

5. Conclusão

A análise das listas de palavras e das concordâncias permitiu identificar a natureza da alternância de código empregada por MEG e JAM ao interagir com o pai pelo telefone. Os dados indicam que, no caso de MEG, seu uso é bastante específico, limitado na maioria dos casos à inserção de substantivos ingleses isolados e a algumas ocorrências de citações de discurso direto. Pela análise dos dados relacionados à criança JAM, constatou-se que seu uso da alternância de código é muito parecido. Entretanto, ao examinar as listas de palavras e as concordâncias que resultaram das análises da interação entre mãe e filho, ficou evidente uma diferença notável entre os dois sujeitos. Embora utilizando alternância de código com o pai, JAM também recorria muito à ajuda da mãe em busca de traduções para o português de palavras inglesas. Essa diferença pode ser indicativa de uma relação entre o uso de alternância de código e o desenvolvimento lingüístico das duas crianças. Pelas evidências encontradas nos dados, MEG (7;10 na época da gravação das conversas) mostra uma certa consistência e segurança no uso de alternância de código, utilizando quando e quanto ela julga necessário na interação com seu pai, não achando necessário recorrer à ajuda da mãe. Em relação a JAM (5;6), por outro lado, os dados mostram uma certa incoerência no uso de alternância de código: ora insere palavras em inglês ora pede a tradução para a mãe. Isso

Crop, 10, 2004

talvez indique uma imaturidade no que diz respeito à consciência de sua condição bilíngüe e da de seu pai. É evidente que JAM sabe que pode empregar alternância de código ao interagir com o pai, mas parece inseguro em relação a 'quando' e 'quanto'. Talvez, por causa do seu estágio de desenvolvimento lingüístico, falte para JAM o conhecimento necessário para estabelecer os parâmetros que regem o uso de alternância de código com este determinado falante.

6. Considerações finais

Os resultados das análises apresentados e discutidos neste artigo deixa evidente o valor de poder realizar tais análises eletronicamente. Em questão de segundos, o pesquisador pode dispor de dados confiáveis, os quais podem ser examinados tanto de forma quantitativa quanto qualitativa. O exame das listas de palavras e das listas de concordâncias dos dados de JAM e MEG permitiu detectar várias facetas do fenômeno de alternância de código e seu uso pelas duas crianças. Além dos aspectos estruturais e pragmáticos revelados pelos dados, foram levantadas algumas hipóteses que merecem ser investigadas nos demais dados do LOBILL corpus. Através do CLAN, esse tipo de investigação torna-se fácil e rápida, e também permite a comparação entre dados de corpora diferentes. É evidente que essa metodologia eletrônica pode contribuir muito para a investigação de vários fenômenos lingüísticos e não somente a da alternância do código. As opções que o CHILDES pode oferecer em termos de construção e exploração de um corpus oral ou escrito são amplas. Cabe ao pesquisador delimitar o tipo de corpus que precisa construir para a realização dos seus objetivos e transcrevê-lo com suas especificidades.

Referências Bibliográficas

- BAKER, Colin. *The care and education of young bilinguals: an introduction for professionals*. Clevedon: Multilingual Matters, 2000.
- _____. *Foundations of bilingual education and bilingualism*. 3 ed. Clevedon: Multilingual Matters, 2001.
- BERBER-SARDINHA, Tony. *Usando WordSmith Tools na investigação da linguagem*. [online]. 1999a. [cited 05.08.2002]. <http://ael.pucsp.br/DirectPapers/40.pdf>
- BRITO DE MELLO, Heloísa Augusta. *O falar bilíngüe*. Goiânia: Ed. da EFG, 1999.
- MACSWAN, Jeff. *A minimalist approach to intrasentential Code Switching: Spanish-Nahuatl Bilingualism in Central Mexico*. [online]. 1997. [cited 23.03.2003]. (dissertação de doutorado) <http://cpnet.mit.edu/MITECS/Entry/gumperz.html>.
- MACWHINNEY, Brian. *Manual de CHAT*. [online]. 2003. [cited 23.03.03]. <http://childes.psy.cmu.edu/manuals/CHAT.pdf>

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilingüe Infantil*.

_____. *Manual de CLAN*. [online]. 2003. [cited 23.03.03]. <http://childes.psy.cmu.edu/manuals/CLAN.pdf>

_____. *Out of the Baby Book and Into the Computer: Child Language Research Comes of Age* [online]. 2003. [cited 23.03.03]. <http://childes.psy.cmu.edu>

MYERS-SCOTTON, Carol. *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford: Oxford University Press, 2002.

POPLACK, Shana. 'Sometimes I'll Start a Sentence in Spanish y termino español': Toward a Typology of Code-Switching. *Linguistics* 18, 1980: 581-618.

_____. *Noções de compilação de corpus*. [online]. 1999b. [cited 05.08.2002].