

O português na
Plataforma Childes

Leonor Scliar-Cabra
(Organizadora)

EDITORA  INSULAR

O Corpus CHILDES *MCF*: Primeiras Produções de Três Crianças Multilíngues

Madalena Cruz-Ferreira

Resumo

O Corpus CHILDES *MCF* documenta as produções faladas de três irmãos durante os primeiros 18 meses de vida. As crianças, duas meninas e um rapaz, são bilingues simultâneos em português (dialecto de Lisboa) e sueco (*rikssvenka*, dialecto central) através da mãe e do pai, respetivamente. Antes do aparecimento das primeiras palavras, o balbucio das crianças caracterizou-se pela habilidade em isolar componentes prosódicos da fala, nomeadamente frequência, amplitude e ritmo, confirmando que a aquisição da prosódia linguística precede e assiste a aquisição de componentes segmentais da fala. O balbucio prosódico das crianças também favoreceu características típicas da entoação do português e do sueco, dependendo dos interlocutores e de outras variáveis situacionais, manifestando a sua diferenciação das duas línguas. Este capítulo apresenta o enquadramento e os métodos de recolha e análise de dados que guiaram a constituição deste corpus.

Palavras-chave

balbucio; bilinguismo/multilinguismo; diferenciação de línguas; entoação; prosódia.

Abstract

The CHILDES *MCF* Corpus gathers the speech productions of three siblings during the first 18 months of life. The children, two girls and one boy, are simultaneous bilinguals in Portuguese (Lisbon dialect) and Swedish (Central Standard dialect) through their mother and father, respectively. Before the appearance of the first words, the children's babble evinced dexterity in isolating prosodic components of speech, namely, pitch, amplitude and rhythm, confirming that the acquisition of linguistic prosody precedes and assists the acquisition of segmental components of speech. The children's prosodic babble also favoured salient intonational features of either Portuguese or Swedish, depending on interlocutor and other situational variables, pointing to the children's differentiation of their two languages. This chapter presents the background, the data collection methods and the analytical choices which guided the constitution of this corpus.

Keywords

babbling; bilingualism/multilingualism; intonation; language differentiation; prosody.

O corpus CHILDES *MCF* (Cruz-Ferreira, 2009) contém o registo da fala de três irmãos, duas meninas e um rapaz, durante o primeiro ano e meio de vida. As crianças são bilingues simultâneos em português e sueco, principalmente através da mãe e do pai, falantes do dialecto de Lisboa e do dialecto padrão sueco (*rikssvenska*, em sueco, *Central Standard Swedish*, em inglês), respetivamente.

Este capítulo resume o enquadramento e os métodos de recolha e análise de dados que guiaram a constituição deste corpus. A publicação que deve ser citada sempre que o corpus CHILDES *MCF* for referido ou utilizado é Cruz-Ferreira (2006), onde também se descreve o ulterior desenvolvimento linguístico, social e cultural destas crianças, incluindo a sua aquisição trilingue do inglês a partir da idade escolar e o seu contacto com outras línguas.

1. Enquadramento da investigação

À data do início da recolha de dados para este corpus, em 1986, não existia informação sobre a aquisição e desenvolvimento da prosódia infantil nos primeiros meses de vida, nomeadamente em contextos multilingues. Um trabalho de doutoramento, completado três anos antes (Cruz-Ferreira, 1983, ver também Cruz-Ferreira, 1987), tinha investigado o tópico da aquisição e uso da prosódia multilingue em adultos, de onde o interesse em complementar esse tópico com dados da aquisição infantil. O estudo levado a cabo em Cruz-Ferreira (1983) incluiu também a primeira descrição do sistema prosódico do português europeu, com base num levantamento de dados de 30 informantes nativos. Essa descrição foi mais tarde alvo de publicação específica (Cruz-Ferreira, 1998) e informou a análise das produções infantis do corpus *MCF*.

O interesse pela produção prosódica de crianças multilingues prende-se com dois factos adicionais. Primeiro, o facto de a fonologia suprasegmental constituir um sistema tão específico de cada língua como a fonologia segmental (Cruz-Ferreira, 1985). Convém esclarecer que o termo *suprasegmental* não designa características da fala que de algum modo pairam ‘acima’ de sequências segmentais, ou que ocorrem independentemente destas, mas sim a concomitância necessária de características segmentais e prosódicas, que são indissociáveis em qualquer produção falada. Segundo, o facto de a produção prosódica preceder a produção segmental, na aquisição infantil. Esta observação ganhou reconhecimento na literatura especialmente a partir do trabalho de Crystal (1979), embora tivesse já sido notada no final do século XIX por Stevenson (1893) e Lukens (1894). Trabalhos mais recentes salientam não só a precedência da prosódia na aquisição da linguagem, como o seu papel fundamental no próprio desenvolvimento da produção segmental: o fraseamento prosódico enquadra e assiste a estruturação gramatical de cada língua, assinalando contrastes e significados pertinentes a cada uma. Duas publicações, Cruz-Ferreira (2006; 2010), contêm resenhas críticas da literatura sobre este tópico, e Cruz-Ferreira (1999b) relata os usos que as três crianças do corpus *MCF* mais tarde fizeram da prosódia, durante a sua aquisição do inglês entre os 5 e os 10 anos.

A literatura acessível no período de recolha de dados era também escassa quanto à produção e desenvolvimento fonético segmental nos primeiros meses de vida, incluindo em crianças monolingues. Ainda hoje, em algumas publicações, a fase do balbucio continua a ser designada como “pré-linguística”, como se apenas a produção segmental fosse digna de caracterização *linguística*. A fase do balbucio é, pelo contrário, uma fonte valiosa de provas de desenvolvimento linguístico, pela riqueza da gama e variedade prosódica que revela. Os balbucios prosódicos evidenciam aquisição de destreza fonatória suprasegmental quanto a frequência, amplitude, compasso temporal e duração, sendo estas as dimensões que formam a

substância de acentos e ritmos típicos de cada língua bem como, mais geralmente, a substância da demarcação de saliências linguísticas. No caso de crianças multilíngues como as que informam este corpus, interessava verificar se, como e quando as crianças produziam prosódia diferencial, específica de cada língua, antes das primeiras “palavras”. Os resultados preliminares desta investigação encontram-se em Cruz-Ferreira (2003) e, mais detalhadamente, em Cruz-Ferreira (2006).

Subsidiariamente, o corpus *MCF* integra dados sobre fala dirigida a crianças em português e sueco, e sobre os modelos adultos reais a que as crianças são expostas. Estes modelos não podem ser assumidos implicitamente, por exemplo como reproduções de pronúncias palavra por palavra, nem identificados com formas padrão de cada língua, por dois motivos: nem adultos nem crianças falam por “palavras”, sejam quais forem as línguas que usem no dia-a-dia; e falas coloquiais, em família, estão longe de corresponder a formas padrão de cada língua.

2. As crianças

Karin e Sofia nasceram na Suécia em 1986 e 1988, respetivamente, Mikael nasceu em Portugal em 1990. Devido a compromissos profissionais do pai, a família mudou-se da Suécia para Portugal em 1988 (Karin 2;0, Sofia 7 semanas) e depois para a Áustria, em 1990 (Karin 4;2, Sofia 2;4, Mikael 3 semanas), onde permaneceu até 1992 (Karin 5;9, Sofia 4;11, Mikael 1;8).

Cada uma das crianças viveu assim num país diferente durante o primeiro ano de vida. O corpus *MCF* dedica-se especificamente à fala de cada criança durante esse primeiro ano, mas contém também produções das crianças mais velhas presentes em sessões de gravação das mais novas, nomeadamente de Karin e Sofia até às idades de 5;7 e 3;8, respetivamente. Durante a totalidade dos anos que o corpus abrange, Karin frequentou um infantário local na Suécia entre as idades de 1;1 e 1;9, cerca de 15 horas por semana, e um infantário sueco em Portugal entre as idades de 3;0 e 3;9, diariamente. Na Áustria, frequentou um infantário local diariamente dos 4;2 aos 5;9. Também na Áustria e também diariamente, Sofia frequentou um infantário local dos 3;1 aos 3;11. Tanto Karin como Sofia tiveram assim alguma escolaridade e contacto com crianças em alemão, embora a língua alemã não faça parte deste corpus. (Mikael começou a sua escolaridade em Hong Kong aos 3;1, em inglês, numa idade não abrangida pelo corpus.)

As crianças sempre viveram com os pais e sempre participaram ativamente na vida familiar. O seu desenvolvimento é típico, com duas exceções que podem refletir-se na produção de duas das crianças durante o período de tempo abrangido pelo corpus. Aos 4 anos, Sofia, cuja fala se desenvolveu em geral mais tarde que a dos irmãos, foi diagnosticada com 40% de surdez causada, como então se verificou, por otites médias sucessivas que a tinham afetado praticamente desde o nascimento. O problema foi resolvido por duas intervenções cirúrgicas, aos 4 e 6 anos. Uma consequência interessante desta circunstância é que a fala de Sofia se caracterizou por alguma indistinção articulatória até cerca dos 10 anos de idade, tanto em português como em sueco, enquanto que as suas produções em inglês, cuja aquisição decorreu justamente a partir dos 4 anos e com as suas faculdades auditivas em recuperação ou já

recuperadas, foram desde o início mais claras e mais fluentes. Por seu lado, a fala de Mikael destacou-se por um assinalável ceceo até à idade de 5;9 nas três línguas que então falava, data em que ele espontaneamente corrigiu a sua articulação.

3. As línguas

Durante os anos a que o corpus se refere (1986 a 1991), a exposição das crianças ao português e ao sueco efetuou-se principalmente através dos pais, devido a mudanças de país como acima se referiu. Os pais tinham por essa altura adquirido a língua um do outro, embora a mãe usasse predominantemente português e o pai sueco com as crianças. As crianças regularmente usaram cada uma das suas línguas com cada um dos pais e com outras pessoas que identificavam como falantes de uma ou de outra. Também diferenciaram o uso de cada língua, por exemplo, para conversar com brinquedos e outros objetos que lhes tinham sido oferecidos por parentes e amigos identificados da mesma forma.

No primeiro ano de vida de cada criança, o contacto com o português e o sueco foi secundariamente mantido através de visitas feitas a Portugal e à Suécia com duração de algumas semanas, ou de visitas com a mesma duração recebidas de parentes e amigos portugueses e suecos. Estas visitas proporcionaram reforço da familiarização das crianças com a cultura, os hábitos e as tradições de cada país. As crianças foram também expostas a variedades locais das suas duas línguas, bem como a outras línguas e a variedades do inglês, por necessidades de comunicação nos países de residência da família, ou entre parentes e amigos portugueses e suecos.

A mãe teve a seu cargo o cuidado dos filhos, a tempo inteiro, durante os seus primeiros anos de vida, resultando em que a exposição das crianças ao português suplantou a sua exposição ao sueco durante este período. Para compensar esta assimetria, acentuada por ausências regulares do pai em viagens de serviço, os pais optaram por comunicar um com o outro em sueco na presença dos filhos. As crianças começaram por preferir usar português entre si, exceto para conversar acerca de acontecimentos ou comemorações especificamente associados com a Suécia, como férias de esqui ou celebrações do solstício de verão, quando usavam sueco. Em 1993, a família mudou-se para Hong Kong e depois para Singapura, em 1994, países onde o inglês é uma das línguas oficiais. As crianças residiram em Singapura até aos 16 anos, onde foram escolarizadas nessa língua. Desde o início da sua escolarização em inglês, as crianças começaram gradualmente a usar esta língua cada vez mais entre si. Cerca de dois a três anos depois da mudança para Singapura, o inglês passou a ser a língua preferencial de comunicação entre elas, situação que se mantém até hoje.

4. Dados do corpus

O corpus *MCF* contém produções referentes às seguintes idades de cada criança:

Karin - de 0;1 a 2;4

Sofia - de 0;1 a 2;6

Mikael - de 0;1 a 1;5

4.1 Colheita dos dados

A recolha de dados foi feita pela mãe das crianças, que é também a investigadora, através de gravações de áudio, de vídeo, e de diários pormenorizados dedicados a cada criança.

Os dados foram colhidos desde o nascimento de cada criança, com regularidade mensal durante o primeiro ano de vida. Mais tarde, especialmente a partir do aparecimento das primeiras palavras, a colheita não obedeceu a regularidade cronológica, incidindo principalmente sobre características novas e/ou diferentes da fala das crianças, como por exemplo um uso de entoação inédito ou uma nova sobregeneralização lexical ou gramatical, da mesma forma que novidades e/ou diferenças de comportamento em outras áreas de desenvolvimento se registaram nos diários.

Comportamentos linguísticos novos seguiram-se, em regra, a períodos mais ou menos extensos de repetição e variações em torno de uma estrutura segmental ou suprasegmental básica já adquirida e documentada em recolhas de dados anteriores. Repetições deste tipo foram especialmente frequentes durante as primeiras tentativas das crianças em comunicar, antes do aparecimento das primeiras palavras. Nesta fase, as crianças pareceram contentar-se com produções estereotipadas que replicavam, com variações mínimas, produções previamente acolhidas favoravelmente pelos seus interlocutores, ou que de outro modo satisfaziam as suas necessidades de comunicação.

Registos de áudio foram feitos em cassetes, com gravadores Sony portáteis dispendo de um microfone plano com sensibilidade ajustável. Para registos de vídeo, usaram-se Handycams Sony Fe/Cr Hi8, também portáteis. Dados em áudio e vídeo foram revistos tão prontamente quanto possível dentro das 24 horas após cada recolha, e complementados sempre que necessário por anotações nos diários.

As gravações e os diários de cada uma das crianças cobrem períodos de tempo aproximadamente iguais e situações interativas semelhantes. Os dados contidos no corpus *MCF* correspondem, no total, a cerca de duas horas de gravação para cada criança, incluindo sessões em que duas crianças ou as três participaram. O nascimento da segunda criança deu origem a participações múltiplas deste tipo, dado que por vezes se tornou impossível excluir um irmão mais velho do local e hora de recolha de dados planeados para um irmão mais novo, ou vice-versa. As intromissões de irmãos mais velhos e, mais tarde, também de mais novos, que foram recolhidas nessas condições fazem naturalmente também parte do corpus.

Os dados retratam assim uma gama diversa de situações. Além de sessões especificamente dedicadas ao registo de progresso, linguístico ou outro, e que por regra decorreram em casa, as gravações incidem sobre rotinas diárias, brincadeira a sós ou com os irmãos, dentro ou fora de casa, e interação com pais e parentes. Tentou-se captar uma gama de aptidões de linguagem e de socialização nos primeiros meses de vida, incluindo modulações de voz e estratégias de chamada de atenção, ou de superação de intrusões de outras crianças ou de ruído ambiente.

Os diários foram manuscritos e relatam extensivamente o progresso geral de cada criança, tanto linguístico como noutras áreas de desenvolvimento, nomeadamente motor, cognitivo, social e cultural. Registam comentários sobre traços de personalidade de cada criança, bem como preferências, hábitos, atividades bem recebidas ou evitadas, e episódios do dia a dia que revelam a individualidade de cada criança. Os diários incluem também notas sobre data, local e circunstâncias das gravações, interlocutores, estado de espírito destes e das crianças, rotinas e/ou acontecimentos até à hora da gravação e, sempre que necessário, sobre os motivos que levaram a um registo de dados em especial. Dos diários constam ainda notas complementares gravadas num Dictaphone, que se revelou um dispositivo ideal para registo rápido de observações sempre que o acesso a papel e lápis era inconveniente, por exemplo durante a condução de um automóvel. Notas assim ditadas foram passadas a papel logo que possível, e as cassetes reusadas. Anotações em diários foram especialmente valiosas quanto às sessões de gravação da fala de Karin até à idade de 1;6, data em que a família pôde dispor do seu primeiro gravador de vídeo.

A maioria dos dados reflete fala espontânea, exceto nos casos em que a criança interagia com um dos pais ou parentes, ou foi intencionalmente convidada a falar, ou cantar, ou 'ler' um livro ilustrado, por exemplo para recitar rimas e lengalengas, ou dizer os nomes de cores ou de objetos e animais. Cuidou-se, no entanto, em não perturbar a naturalidade das atividades, linguísticas ou outras, durante as sessões de gravação, e em evitar que as crianças se sentissem de alguma maneira forçadas a participar nelas. O consentimento dos adultos que participaram nas sessões de gravação para figurar no corpus foi devidamente pedido e obtido.

Durante os primeiros meses de vida das crianças, sessões típicas de gravação decorreram com a criança confortavelmente recostada num ambiente familiar, brincando sozinha ou interagindo com um dos pais ou um parente. Mais tarde, o gravador de áudio deixou-se ligado num local discreto onde as crianças se entretinham ou estavam a ser cuidadas. Em sessões de recolha de vídeo, as crianças naturalmente começaram por manifestar a sua consciência da presença do gravador, mas a regularidade do seu uso depressa o tornou num objeto rotineiro e, portanto, desinteressante.

Uma crítica possível ao método de levantamento de dados para este corpus é a de que a mãe das crianças esteve regularmente presente durante as sessões de gravação, como participante ou como operadora técnica, excetuando os casos em que outros adultos operaram os gravadores ou em que se ativou o gravador áudio quando as crianças brincavam sozinhas. Outros problemas inerentes a recolhas de dados de fala infantil espontânea, à parte os que são comuns a recolhas efetuadas em ambientes naturais, são do conhecimento geral de investigadores nesta área, incluindo falhas de equipamento devidas ao seu uso inesperado como brinquedos, relutância ou recusa absoluta das crianças em participar, ou necessidade urgente de atendimento a uma das crianças. Em casos destes, a informação registada nos diários constitui uma base preciosa de dados suplementares.

4.2 Transcrição, análise e codificação dos dados

O tratamento dos dados, quanto a transcrição, análise e codificação, esteve a cargo da investigadora, que é especializada em fonética e fala infantil, e fluente em português e sueco.

As transcrições são auditivas e foram feitas logo que possível depois de cada sessão de gravação, recorrendo-se a uma segunda transcrição independente de cada sessão entre duas a quatro semanas mais tarde. Todas as transcrições foram mais uma vez reverificadas no ano 2000, data da primeira publicação do corpus *MCF*, durante o processo de codificação dos dados em formato CHAT (MacWhinney, 2000). Dada a impossibilidade de recorrer a outros foneticistas com experiência de fala infantil em português e sueco, que pudessem transcrever os dados e/ou verificar as transcrições, optou-se assim por um autocontrole múltiplo das transcrições, de modo a salvaguardar a sua fiabilidade tanto quanto possível.

A transcrição dos dados progrediu a uma média de cerca de 30 minutos por cada minuto de fala gravada, tendo as gravações sido mais tarde digitalizadas para efeitos de uso eletrónico. Em 2009, o corpus *MCF* foi sincronizado com os ficheiros áudio que atualmente fazem parte da base de dados. O acesso a estes ficheiros facilita uma avaliação independente adicional das transcrições propostas.

Os dados foram transcritos por inteiro, incluindo falas indistintas e sons não linguísticos, como respirações fundas ou soluços. A codificação dos dados inclui também notas com informação adicional sobre, por exemplo, ações e comentários das crianças ou dos seus interlocutores, antes e depois de uma sessão de gravação. Em casos de discrepância flagrante entre as várias transcrições, por exemplo se uma produção infantil se atribuiu a línguas diferentes em cada transcrição, os dados foram descartados da análise e codificados como fala ininteligível na transcrição. Variações de menor monta, por exemplo a notação de uma fricativa como palatal versus pós-alveolar em duas transcrições, foram mantidas e devidamente assinaladas na codificação.

Todos os ficheiros do corpus incluem um renque %pho, dedicado a transcrição fonética, e um renque %int, dedicado a transcrição prosódica. Os dois renques servem também a transcrição de produções de adultos que ou não correspondem a formas padrão das sua línguas respetivas, ou apresentam características típicas de fala dirigida a crianças. A descrição que segue refere-se às convenções especificamente criadas para efeitos de codificação do corpus *MCF*. Outras convenções seguem as do Manual CHAT.

4.2.1 O renque %pho

Alfabeto fonético. Aquando da codificação dos dados, no ano 2000, o alfabeto fonético disponível em CHAT era o IPAPhon. O corpus *MCF* está agora transcrito no Alfabeto Fonético Internacional (AFI). A transcrição é tão estreita quanto possível, embora tendo em consideração a sua legibilidade. Falas balbuciadas transcrevem-se por inteiro, com indicação e discussão de sons problemáticos no renque %com (comentários). O renque %mod (modelo) regista a transcrição de formas coloquiais adultas e infantis, tal como eram usadas na família.

Símbolos fonéticos. A transcrição de fala adulta e de falas infantis que podem ser facilmente reconhecidas como reproduções de um modelo adulto segue as análises propostas em Cruz-Ferreira (1999a) para o português e em Engstrand (1999) para o sueco. Os símbolos usados em transcrições de balbucios ou de falas não identificáveis como reproduções de uma língua representam os valores padrão que lhes são atribuídos no AFI. Por exemplo, o símbolo [ə] do

AFI representa uma vogal com um timbre próximo do timbre de uma vogal central média tanto do português como do sueco. Em produções infantis próximas de modelos adultos, esta vogal transcreve-se [ɐ] em português e [ə] em sueco. Em balbucios, transcreve-se exclusivamente [ə].

Obstruentes. Símbolos fonéticos representando consoantes sonoras e que se transcrevem com o diacrítico de desvozeamento, por exemplo [d̥], indicam articulações surdas frouxas.

Ditongos. Sequências de vogais em que a segunda vogal continua o tom melódico iniciado na primeira são interpretadas como ditongos. O segmento glide do ditongo transcreve-se com os símbolos [j] ou [w], que assim representam vocoides. Por exemplo, a transcrição [aw] representa uma sílaba, enquanto que a transcrição [au] representa duas.

Sílabas. Sequências de consoantes intervocálicas são silabificadas como ataque de sílaba, para efeitos de atribuição de acento. Esta análise reflete uma escolha teórica entre outras possíveis, sem presumir a sua legitimidade preferencial ou definitiva na descrição silábica de fala infantil. A análise segue as regras fonotáticas de cada uma das línguas, tanto para produções de adultos como para produções infantis reconhecíveis como réplicas de modelos adultos. A transcrição de dois símbolos vocálicos idênticos adjacentes, por exemplo [aa], indica que a criança pronunciou a vogal como duas sílabas consecutivas.

Acento. Dado que a percepção do acento é relativa, a atribuição de acento a uma sílaba depende de uma comparação de sílabas adjacentes. Ressaltos cumulativos de frequência, duração e intensidade destacam a sílaba que está a ser acentuada como diferente das sílabas que a rodeiam. Ressaltos de frequência assinalam a sílaba acentuada como mais alta ou mais baixa, enquanto que os ressaltos de duração e de intensidade que resultam na percepção de acentos são geralmente de grau positivo, assinalando sílabas mais longas e mais fortes. Nas transcrições, o símbolo ['] indica que a sílaba seguinte é acentuada.

Palavras / frases. Na transcrição de fala infantil ou de fala dirigida a crianças, um espaço delimita o que se pode interpretar como uma “palavra” (ou “frase”) dentro de um grupo melódico, mais uma vez sem pretender atribuir foros de legitimidade analítica a uma notação de cariz essencialmente prático. Por exemplo, no ficheiro ks890105.cha do corpus, a primeira linha de cada um dos pares reproduzidos abaixo assinala as produções das três falantes como não correspondendo a uma língua específica, e a segunda linha fornece a transcrição fonética dessas falas:

29	*MUM:	yyy yyy yyy .
30	%pho:	gɔ'dʲɔ gɔ'dʲɔ gɔ'dʲɔ
69	*SOF:	yyy yyy yyy .
70	%pho:	gwa'wa jɛ'jɛ jɛjajɛja'jɛ:
96	*KAR:	yyy yyy ?
97	%pho:	gɔ'dʲɔgɔ'dʲɔ 'tãw̃tãw̃'tãw̃

4.2.2 O renque %int

Este renque transcreve os contornos melódicos da fala tanto infantil como adulta, e inclui indicação de qualidade de voz e de características paralinguísticas como, por exemplo, laringealização e compasso temporal.

A transcrição da entoação adapta a análise nuclear originalmente concebida para descrição da língua inglesa, já usada nos trabalhos anteriores sobre entoação do português europeu, referidos acima. Esta análise assume a divisão de enunciados em grupos melódicos, delimitados por pausas silenciosas, e que por norma correspondem a ciclos de respiração cómoda durante a fala (ver Cruttenden, 1997, para discussão desta e de outras abordagens à análise da entoação). Em produções adultas, cada grupo melódico contém um tom nuclear, por exemplo ascendente ou descendente, que afeta a última sílaba acentuada desse grupo. Clarifica-se, mais abaixo, a análise de grupos melódicos em fala infantil e de constituintes opcionais de um grupo melódico.

Pares de símbolos abreviados descrevem a entoação dos enunciados. No caso de tons simples (unidirecionais), descendentes, ascendentes ou fixos, o primeiro símbolo indica a frequência em que o tom se inicia, alta (H, do inglês *high*), média (M, *mid*) ou baixa (L, *low*), e o segundo símbolo indica o tipo de movimento de altura da voz, descendente (F, *fall*), ascendente (R, *rise*) ou fixo (L, *level*). O tom descendente-baixo grave do português constitui a única exceção a esta notação, como se esclarece mais abaixo. Os termos ‘alto’, ‘médio’ e ‘baixo’ denotam frequências relativas: uma altura de voz ‘média’ refere-se à gama de frequências habitualmente usada por um falante, tal como se pode intuir auditivamente por contacto regular com esse falante. Uma altura de voz ‘alta’ ou ‘baixa’ define-se, por conseguinte, em relação à ‘média’ de cada falante.

No caso de tons complexos (bidirecionais), a sucessão de símbolos indica o tipo de movimento de altura da voz do tom nuclear. Esquemáticamente, as convenções usadas são como segue.

Tons descendentes simples.

LF (do inglês *low-fall*): descendente-baixo

MF (*mid-fall*): descendente-médio

HF (*high-fall*): descendente-alto

eLF (*extra-low fall*): descendente-baixo grave, com início baixo e queda abaixo da gama baixa habitual do falante.

Tons ascendentes simples.

LR (*low-rise*): ascendente-baixo

MR (*mid-rise*): ascendente-médio

HR (*high-rise*): ascendente-alto

Tons fixos.

LL (*low-level*): fixo-baixo

ML (*mid-level*): fixo-médio

HL (*high-level*): fixo-alto

Tons complexos.

RF (*rise-fall*): ascendente-descendente

FR (*fall-rise*): descendente-ascendente

RFR (*rise-fall-rise*): ascendente-descendente-ascendente

FRF (*fall-rise-fall*): descendente-ascendente-descendente

Sempre que necessário, dão-se indicações complementares quanto à altura inicial ou final dos tons, por exemplo, ‘descendente-alto até médio’.

Entoação pré-nuclear.

preH (do inglês *prehead*): pré-cabeça, se a houver, incorporando as sílabas átonas que precedem a primeira sílaba acentuada do enunciado.

H (*head*): cabeça, se a houver, com início na primeira sílaba acentuada do enunciado e estendendo-se até ao tom nuclear.

Estes dois símbolos seguem sempre os símbolos que indicam o início ou o tipo de movimento de altura de voz, pelo que não se cria confusão entre o H que indica ‘alto’ e o H que indica ‘cabeça’. Por exemplo:

LpreH (do inglês *low prehead*): pré-cabeça baixa

MpreH (*mid prehead*): pré-cabeça média

HpreH (*high prehead*): pré-cabeça alta

LH (*low head*): cabeça baixa

MH (*mid head*): cabeça média

HH (*high head*): cabeça alta

FH (*falling head*): cabeça descendente

RFH (*rising-falling head*): cabeça ascendente-descendente

A transcrição de cada grupo melódico (abreviado tg, do inglês *tone group*) é feita sucessivamente no renque %int, sendo cada grupo separado por ponto e vírgula, e numerado por ordem. Sinais de + indicam os limites prosódicos da sucessão de pré-cabeça, cabeça e tom nuclear. Por exemplo, no ficheiro k880500.cha, a linha 691 transcreve ortograficamente uma fala, neste caso em sueco, e a linha 693 transcreve a entoação dessa fala:

691 *DAD: vad heter det [!] för nåt (.) vad är det [!] för nåt (.) vad heter [!] det .

693 %int: 1tg , MH+ML; 2tg , MH+LR; 3tg , LH+MF .

A transcrição de balbucios não pressupõe a existência de um tom nuclear, registando apenas altura e movimento de voz em cada sílaba balbuciada, de acordo com as mesmas convenções. A única diferença é que nas transcrições de balbucios os sinais de + indicam limites silábicos. Por exemplo, no ficheiro m901215.cha, a linha 64 transcreve ortograficamente um balbucio, a linha 65 transcreve a sua composição segmental, e a linha 66 transcreve a sua entoação:

64 *MIK: yyy.

65 %pho: qʔk:ee:: hmʔ ə'mw̃ã

66 %int: 1tg , ML+long MF; 2tg , LL; 3tg , HL+short LF .

4.2.3 Outras convenções e símbolos

Convenções e símbolos adicionais são:

tg(s) (do inglês *tone group*): um ou mais grupos melódicos

syll(s) (*syllable*): uma ou mais sílabas

dipt(s) (*diphthong*): um ou mais ditongos

min(s) (*minute*): um ou mais minutos

sec(s) (*second*): um ou mais segundos

Ortografia. Falas adultas e falas infantis reconhecíveis como reproduções de formas adultas transcrevem-se de acordo com as convenções ortográficas de cada uma das línguas. Tornou-se no entanto necessário criar uma ‘ortografia infantil’ ad hoc para representar fala infantil fluente que, embora replicando modelos adultos, apresenta distorções tanto segmentais como prosódicas que tornam inaplicável qualquer uso legível das convenções CHAT que guiam a transcrição de truncamentos em fala infantil. Esta ‘ortografia infantil’ é uma ortografia fonética, por assim dizer, codificando falas distorcidas tal e qual como se produziram. Nos casos em que esta ortografia figura na transcrição, transcreve-se a ortografia padrão no renque %gls (glosa). Por exemplo, no ficheiro sm910100.cha, as linhas 15 e 16 transcrevem uma fala em português em ‘ortografia infantil’ e em ortografia padrão, respetivamente:

15 *SOF: amã (.) klhi klhkó ?

16 %gls: mamã , a Karin está na escola ?

Línguas. Os ficheiros do corpus *MCF* contêm produções monolíngues e/ou bilingues de uma ou mais das três crianças. Sempre que a indicação da língua em questão se julgou relevante, as abreviaturas *Ptg* e *Sw* representam citação de falas em português e sueco, respetivamente, nos renques %lan (língua), %com (comentários) ou %exp (explicação). As notações *PtgSw* e *SwPtg* referem-se às mesmas línguas quando usadas em fala multilíngue, onde o primeiro símbolo indica a língua de acolhimento (a língua que aceita elementos de outra língua) e o segundo a língua visitante (a língua que oferece os seus elementos a outra língua).

5. Ficheiros de dados

O quadro abaixo lista os ficheiros do corpus *MCF*, por ordem alfabética, e com indicação da idade de cada criança à data da recolha de dados, para cada ficheiro. Nas colunas relativas a cada criança, parênteses indicam que o ficheiro correspondente contém muito pouca fala dessa criança.

Os nomes dos ficheiros incluem um prefixo com as iniciais da(s) criança(s) que participaram na gravação (k, s, m), e a data da recolha de dados em formato aammdd (ano, mês, dia). Por exemplo, o ficheiro ks890510.cha contém dados de Karin e Sofia recolhidos no dia 10 de

maio de 1989. A notação 00 para um dia significa que não foi possível anotar o dia exato da recolha de dados.

Nome do Ficheiro	Karin – Idade	Sofia – Idade	Mikael – Idade
k861020.cha	K 0;1.18		
k861113.cha	K 0;2.11		
k870117.cha	K 0;4.15		
k870203.cha	K 0;5.1		
k870319.cha	K 0;6.17		
k870500.cha	K 0;8		
k870600.cha	K 0;9		
k870800.cha	K 0;11		
k880500.cha	K 1;8		
ks881200.cha	K 2;3	S 0;5	
ks890105.cha	K 2;4.3	S 0;5.24	
ks890510.cha	(K 2;8.8)	S 0;9.	
ksm901205.cha	K 4;3.3	S 2;4.24	M 0;1.25
ksm910700.cha	(K 4;10)	(S 3;0)	M 0;9
ksm920408.cha	(K 5;7.6)	(S 3;8.27)	M 1;5.28
m901100.cha			M 0;1
m901215.cha			M 0;2.5
m910108.cha			M 0;2.28
m910500.cha			M 0;7
m910528.cha			M 0;7.18
m910600.cha			M 0;8
m910800.cha			M 0;10
m910900.cha			M 0;11
m911125.cha			M 1;0.15
s880815.cha		S 0;1.4	
s880900.cha		S 0;2	
s881000.cha		S 0;3	
s881004.cha		S 0;2.23	
s890100.cha		S 0;6.	

s890300.cha		S 0;8	
sm910100.cha		S 2;6	M 0;3

Referências

- Cruttenden, A. (1997). *Intonation* (2ª ed.). Cambridge: Cambridge University Press.
- Cruz-Ferreira, M. (1983). *Non-native comprehension of intonation patterns in Portuguese and in English*. Dissertação de Doutorado, Universidade de Manchester, Inglaterra.
- Cruz-Ferreira, M. (1985). Elementos para um estudo comparativo dos sistemas prosódicos do português e do inglês. Em *Miscelânea de Estudos Dedicados a Fernando de Mello Moser* (pp. 373-388). Lisboa: Faculdade de Letras de Lisboa.
- Cruz-Ferreira, M. (1987). Non-native interpretive strategies for intonational meaning: an experimental study. Em A. James & J. Leather (Eds.), *Sound Patterns in Second Language Acquisition* (pp. 103-120). Berlin: de Gruyter.
- Cruz-Ferreira, M. (1998). Intonation in European Portuguese. Em D. Hirst & A. Di Cristo (Eds.), *Intonation Systems: A Survey of Twenty Languages* (pp. 167-178). Cambridge: Cambridge University Press.
- Cruz-Ferreira, M. (1999a). Illustrations of the IPA. Portuguese (European). Em *Handbook of the International Phonetic Association. A Guide to the use of the International Phonetic Alphabet* (pp. 126-130). Cambridge: Cambridge University Press.
- Cruz-Ferreira, M. (1999b). Prosodic mixes: Strategies in multilingual language acquisition. *International Journal of Bilingualism*, 3(1), 1-21.
- Cruz-Ferreira, M. (2003). Two prosodies, two languages: Infant bilingual strategies in Portuguese and Swedish. *Journal of Portuguese Linguistics*, 2(1), 45-60.
- Cruz-Ferreira, M. (2006). *Three is a Crowd? Acquiring Portuguese in a Trilingual Environment*. Clevedon: Multilingual Matters.
- Cruz-Ferreira, M. (2009). CHILDES MCF Multilingual Corpus. Acessível em <http://childes.psy.cmu.edu/browser/index.php?url=Biling/MCF/> [acedido em 21.08.2014]
- Cruz-Ferreira, M. (Ed.). (2010). *Multilingual Norms*. Frankfurt: Peter Lang.
- Crystal, D. (1979). Prosodic development. Em P. Fletcher & M. Garman (Eds.), *Language Acquisition: Studies in First Language Development* (pp. 33-48). Cambridge: Cambridge University Press.

Engstrand, O. (1999). Illustrations of the IPA. Swedish. *Em Handbook of the International Phonetic Association. A Guide to the use of the International Phonetic Alphabet* (pp. 140-142). Cambridge: Cambridge University Press.

Lukens, H. T. (1894). Preliminary report on the learning of language. *Pedagogical Seminary*, 3, 424-460.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Vol. 1: *Transcription Format and Programs* (3^a ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Stevenson, A. (1893). The speech of children. *Science*, 21, 118-120.
