



Ana Rita Campos **A Extensão Média de Enunciado
como indicador de
desenvolvimento linguístico no
português europeu**

Dissertação de Mestrado em Desenvolvimento e
Perturbações da Linguagem na Criança

dezembro 2014

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Desenvolvimento e Perturbações da Linguagem na Criança – Área de Especialização em Terapia da Fala e Perturbações da Linguagem, realizada sob a orientação científica da Professora Doutora Maria Lobo e coorientação da Professora Eileen Sua Kay.

Declaro que esta Dissertação é o resultado da minha investigação pessoal e independente. O seu conteúdo é original e todas as fontes consultadas estão devidamente mencionadas no texto, nas notas e na bibliografia.

O candidato,

Setúbal, 20 de dezembro de 2014

Declaro que esta Dissertação se encontra em condições de ser apreciada pelo júri a designar.

A orientadora,

A coorientadora,

Setúbal, 20 de dezembro de 2014

*A todas as pessoas que acreditam, nem que seja por breves momentos,
que um impossível é um possível longínquo.*

AGRADECIMENTOS

Gostaria de expressar os meus sinceros agradecimentos a todas as pessoas que, de alguma maneira, deram o seu contributo para a finalização desta etapa, mas em especial às que se revelaram essenciais.

À orientadora científica, Professora Doutora Maria Lobo, pela incansável e rigorosa ajuda nos momentos mais difíceis e determinantes.

À coorientadora científica, Professora Eileen Sua Kay, pelo seu valioso contributo ao longo deste trabalho.

Ao Professor Doutor Hélder Alves, pela constante colaboração e incentivo, e por ter despertado o meu interesse pela estatística.

À Professora Doutora Ana Castro, um agradecimento pela disponibilidade e interesse, demonstrados no momento certo.

À diretora do Agrupamento de Escolas de Santa Comba Dão, Professora Madalena Dinis, pela confiança depositada em todo o trabalho desenvolvido.

Aos Encarregados de Educação dos alunos que integraram o presente estudo e aos docentes que colaboraram no procedimento de recolha, pela manifesta valorização do projeto, agora concretizado.

Às crianças, participantes, e a todas as outras com quem já tive o privilégio de interagir, por manterem viva a paixão pela minha profissão.

À minha amiga, e companheira nesta investigação, Raquel João Campos, um agradecimento especial pela infindável partilha, científica, profissional e pessoal, em todos os desafios que trilhamos juntas.

Aos meus familiares, particularmente pais, irmã, padrinho e avós, por fazerem, sempre, parte da realização dos meus sonhos.

Aos meus amigos, por todo o apoio demonstrado e pela força que me transmitiram em momentos de dúvida ou angústia.

Finalmente, à Petra, porque a simplicidade da sua companhia também iluminou este percurso.

RESUMO

A Extensão Média de Enunciado como indicador de desenvolvimento linguístico no português europeu

Ana Rita Campos

PALAVRAS-CHAVE: Extensão Média de Enunciado - palavras (EME-p), discurso narrativo, discurso espontâneo, desenvolvimento da linguagem, nível de escolaridade

A Extensão Média de Enunciado (EME) é um índice que resulta do rácio entre o número de morfemas (EME-m) ou número de palavras (EME-p) e o número de enunciados produzidos pela criança. Este índice tem vindo a ser estudado e utilizado em diferentes países e línguas, e, apesar de muitos estudos o apontarem como um bom indicador de desenvolvimento gramatical e linguístico, outros são controversos, principalmente quando aplicados a línguas com morfologia flexional mais rica, como o caso do português.

Dada a escassez de dados para o português europeu, confinados a duas faixas etárias, e atendendo às evidências de que a EME-m e a EME-p apresentam uma boa correlação, neste estudo procurou-se analisar a EME-p, dos três aos sete anos. Os principais objetivos da presente dissertação são verificar se a EME-p será um bom indicador de desenvolvimento gramatical e linguístico, através da sua correlação com a idade e analisando diferenças entre faixas etárias, entre contextos de discurso (narrativo e espontâneo) e o nível de escolaridade (pré-escolar e escolar).

Participaram no estudo cinquenta crianças, monolíngues do português europeu e com desenvolvimento típico de linguagem, a frequentar Centros Educativos do Concelho de Santa Comba Dão. Foram gravadas, em dois momentos, amostras de discurso, espontâneo e narrativo, perfazendo, pelo menos, cem enunciados, que, posteriormente, foram transcritos através do Software Computerized Language Analysis (CLAN), que também realizou, de forma automática, o cálculo da EME-p.

Os resultados indicam uma associação forte entre a idade e a EME-p, apesar de não existirem diferenças entre todas as faixas etárias, e uma correlação positiva entre a EME-p e o desenvolvimento da linguagem de crianças monolíngues do português europeu. O valor da EME-p em discurso narrativo é estatisticamente superior ao seu valor em discurso espontâneo e o grupo em idade pré-escolar apresentou menor desempenho do que o grupo escolar.

Em suma, considera-se que a EME-p pode fornecer informação pertinente sobre o desenvolvimento expressivo morfossintático da criança, contudo não poderão ser considerados valores padrão nem se poderá utilizar este índice como critério de diagnóstico.

ABSTRACT

The Mean Length of Utterance as indicator of language development in European Portuguese

Ana Rita Campos

KEYWORDS: Mean Length of Utterance (MLUw), narrative discourse, spontaneous speech, language development, education level.

The Mean Length of Utterance (MLU) is an index that results from the ratio between the number of morphemes (MLUm) or number of words (MLUw) and the number of utterances produced by the child. This index has been studied and used in different countries and languages and although many studies tell us that it is a good grammatical and linguistic development indicator, others are controversial, especially when applied to languages with richer inflectional morphology, such as Portuguese.

Given the lack of data for European Portuguese confined to two age groups, and given the evidence that MLUm and MLUw have a good correlation, in this study we tried to analyze the MLUw in a group from three to seven years old. The main objectives of this dissertation are to check if the MLUw is a good indicator of grammatical and linguistic development, verifying its correlation with age and analyzing differences between age groups, discourse contexts (narrative and spontaneous) and the level of education (pre -school and school-age).

In this study participated fifty children. They all were monolingual speakers of European Portuguese with typical language development, attending educational centers of the municipality of Santa Comba Dão. Speech samples were recorded in two different moments, in spontaneous speech and narrative context, making at least an hundred utterances which were later transcribed through Computerized Language Analysis Program (CLAN) which also calculated automatically the MLUw.

The results indicate a strong association between age and the MLUw, although there are no significant differences between all age groups, and a positive correlation between MLUw and the development of European Portuguese children. The value of the MLUw in narrative discourse is statistically higher than in spontaneous speech and the group of pre-school age showed lower performance than the school-age group.

In short, it is considered that the MLUw can provide relevant information on child development of morphosyntactic expressive competence, but cannot be considered as normative data or for diagnostic criterion.

ÍNDICE

INTRODUÇÃO.....	1
CAPÍTULO I: ENQUADRAMENTO TEÓRICO.....	3
I. 1. Revisão da literatura.....	3
<i>Aquisição e desenvolvimento morfossintático.....</i>	3
<i>Avaliação da morfossintaxe: a recolha de produções orais</i>	5
<i>Extensão Média de Enunciado (EME).....</i>	8
<i>Aplicabilidade e limitações da EME.....</i>	13
I. 2. Objetivos do estudo.....	15
I. 3. Hipóteses	15
CAPÍTULO II: METODOLOGIA	17
II. 1. Participantes e considerações éticas	17
II. 2. Procedimento de recolha e materiais.....	19
II. 3. Constituição dos corpora	21
CAPÍTULO III: RESULTADOS.....	26
III.1. Análise dos corpora.....	26
III.2. Descrição dos resultados	28
CAPÍTULO IV: DISCUSSÃO	35
IV.1. Discussão dos resultados	35
IV.2. Limitações do estudo e trabalho futuro	40
CONCLUSÃO.....	42
REFERÊNCIAS BIBLIOGRÁFICAS	44
LISTA DE TABELAS	50
LISTA DE GRÁFICOS.....	51
LISTA DE FIGURAS	51
APÊNDICE A.....	i
APÊNDICE B.....	iii
APÊNDICE C.....	vii
APÊNDICE D.....	ix
APÊNDICE E.....	xiv
ANEXO I.....	xvi
ANEXO II.....	xxi

LISTA DE ABREVIATURAS

EME – Extensão Média de Enunciado

MLU – Mean Length of Utterance

EME-m – Extensão Média de Enunciado, medida em morfemas

EME-p – Extensão Média de Enunciado, medida em palavras.

EME-p-DN – Extensão Média de Enunciado, medida em palavras, no contexto de discurso narrativo.

EME-p-DE - Extensão média de enunciado, medida em palavras, no contexto de discurso espontâneo.

MLR - Mean Length of Response

PE – Nível pré-escolar

E – Nível escolar

DL – Desenvolvimento Linguístico

GOL_E – Grelha de Observação da Linguagem – Nível Escolar

TALC – Teste de Avaliação da Linguagem da Criança

CHILDES - Child Language Data Exchange System

CLAN - Software Computerized Language Analysis

SPSS - Software Statistical Package for Social Sciences

H - Hipótese

M – Média

A - anos

M – meses

DP – Desvio padrão

p – Probabilidade de cometer um erro estatístico tipo I “falso positivo” (nível de significância)

n – número de itens

SW – Teste Shapiro-Wilk

TLC – Teorema do Limite Central

H - Teste Kruskal-Wallis

r – correlação linear *r* de Pearson

rs – correlação de Spearman

F – Análise de Variância – ANOVA

CD – CD-ROM contendo os *corpora* em formato CHAT

INTRODUÇÃO

A aquisição da língua materna constitui, indubitavelmente, um dos marcos mais significativos da primeira infância e, possivelmente, de toda a vida de um indivíduo, enquanto chave de acesso ao passado, desenvolvimento do presente e planificação do futuro (Albuquerque, 2000). É neste domínio da linguagem oral que se foca o campo de ação de muitos terapeutas da fala, em Portugal, assegurando a prevenção, avaliação, diagnóstico, intervenção e estudo científico.

O presente estudo nasce da crescente necessidade de implementar uma prática profissional baseada na evidência científica, para a qual contribui o estudo e utilização de indicadores de desenvolvimento linguístico. Estes constituem uma mais-valia no processo terapêutico, não só para avaliar a presença de perturbação linguística como também para verificar a eficácia da intervenção, ambos elementares num bom plano de avaliação, adaptado às necessidades da criança (Acosta, Moreno, Ramos, Quintana, & Espino, 2003).

Um dos indicadores que tem vindo a ser, há várias décadas, estudado e utilizado como índice de desenvolvimento gramatical é a Extensão Média de Enunciado (EME), que corresponde ao comprimento, em palavras ou morfemas, designadamente EME-p ou EME-m, dos enunciados produzidos pela criança. Trata-se de uma medida de produtividade linguística objetiva que, segundo Shipley e McAfee (2009), fornece informação importante sobre o desenvolvimento linguístico da criança. A EME é calculada a partir de amostras de produções orais, que devem ser representativas das reais competências da criança e, por isso, obedecer a um método cuidadosamente planeado (Demuth, 1998; Lund & Duchan, 1983; Miller, 1981).

Apesar de existirem referências prévias a este índice, contabilizado em palavras pelo investigador Nice (1925), citado por MacWhinney (2000), foi Roger Brown, em 1973, que ficou conhecido como o proponente do estudo da EME, designada por “Mean Length of Utterance”, em morfemas, para a língua inglesa. A partir daqui, outros investigadores procuraram validar estes dados na década de 80 (Chapman & Miller, 1981) e, também, mais recentemente (Rice, Redmond, & Hoffman, 2006; Rice *et al.*, 2010). Na mesma linha, diversos estudos têm vindo a confirmá-lo como indicador de

desenvolvimento sintático e linguístico, inclusive no português brasileiro (Araujo, 2007; Fensterseifer & Ramos, 2003). A EME tem apresentado um papel de destaque, não só em estudos de aquisição e desenvolvimento de linguagem, como também com a sua aplicabilidade noutros domínios, como por exemplo enquanto índice valioso na investigação de crianças com perturbações de linguagem (Rice *et al.*, 2010). Contudo, a utilização da EME como indicador de desenvolvimento gramatical e linguístico não é consensual, pois alguns autores apresentaram resultados controversos, assumindo que a EME não reflete a complexidade e qualidade dos enunciados produzidos.

Estes factos espoletaram o interesse de investigar este índice em crianças falantes do português europeu, em faixas etárias abrangentes, dos três aos sete anos, uma vez que são recentes e escassos os estudos em Portugal no âmbito da EME-p, confinados às faixas etárias dos quatro e cinco anos de idade (Cacela, 2013; Faria, 2013). Nestes foram pesquisados valores médios da EME, bem como a influência do género e do nível de escolaridade dos pais.

O principal objetivo deste estudo é verificar se a EME-p é um indicador de desenvolvimento linguístico em crianças monolingues do português europeu, através da análise da associação desta medida com o desenvolvimento linguístico e com a idade da criança. Pretende-se, ainda, caracterizar a EME-p, de acordo com a faixa etária, com o tipo de discurso (narrativo e espontâneo) e com o nível de escolaridade (pré-escolar e escolar).

A presente dissertação encontra-se dividida em quatro partes. No primeiro capítulo far-se-á um levantamento da literatura mais relevante, enquadrando a temática em estudo, de forma a integrar os objetivos e hipóteses. Na segunda parte, apresentam-se os procedimentos metodológicos, quanto aos participantes, materiais, métodos de recolha de dados e constituição dos *corpora*. Seguidamente apresentam-se os resultados obtidos, em termos de análise dos *corpora* e estatística, seguindo-se a discussão, menção de algumas limitações do estudo e seu impacto futuro. Finalmente, a conclusão fará uma abordagem integrada dos aspetos mais importantes do presente trabalho, em jeito conclusivo.

CAPÍTULO I: ENQUADRAMENTO TEÓRICO

I. 1. Revisão da literatura

Aquisição e desenvolvimento morfossintático

O estudo da organização formal do sistema linguístico de uma língua denomina-se tradicionalmente por gramática (Acosta *et al.*, 2003). Neste âmbito, a morfologia e a sintaxe são duas das componentes que têm sido consideradas nos estudos inerentes ao desenvolvimento linguístico. Segundo Halle (1973), citado por Azuaga (1996), a primeira tem como objeto de estudo a palavra e as suas alterações sistemáticas na forma, que se relacionam com mudanças do seu sentido; e, ainda, o modo como a estrutura da palavra reflete a relação com outras palavras (Aram, Morris, & Hall, 1993). Denomina-se morfema à unidade de análise dessas mesmas palavras, que, durante muito tempo, foi tratada como sendo uma “unidade mínima com significado que integra a estrutura da palavra”, tal como refere Sim-Sim (1998), (p.147). Segundo Villalva (2008), numa redefinição do conceito, e uma vez que nem todas as unidades de análise são portadoras de significado, morfema é qualquer elemento que participa da construção de uma palavra. A sintaxe, por sua vez, é a componente da gramática que se ocupa da representação de todos os aspetos do conhecimento intuitivo dos falantes (Prieto, Estrella, Thorson, & Vanrell, 2011), que Campos e Xavier (1991) sumariam como “o estudo da organização das sequências linguísticas maiores que a palavra” (p.22).

A morfologia e sintaxe são abordadas, na linguística moderna, cada vez mais, conjuntamente, o que se reflete, também, na construção e validação de testes de avaliação da linguagem na criança, utilizados pelos terapeutas da fala em Portugal (Santos *et al.*, 2006), nos quais estes domínios, surgem, muitas vezes, de forma emparelhada com a designação de morfossintaxe. O conhecimento da morfossintaxe, associado ao dos outros domínios da linguagem (fonologia, semântica e pragmática), converge para a produção e reconhecimento de expressões bem formadas, ou seja, frases gramaticais (Campos & Xavier, 1991). A quebra dessas regras de formação linguística provoca agramaticalidade, por vezes, presente nas produções orais da criança, com ou sem patologia linguística.

De acordo com Gonçalves, Guerreiro, e Freitas (2011) o desenvolvimento do discurso infantil emerge a partir dos dois anos e só estabiliza no final da adolescência. Para outros autores, o período holofrástico (utilização de palavras isoladas que encerram todo o conteúdo frásico que a criança pretende transmitir), antes dos dois anos, corresponde às primeiras manifestações de desenvolvimento sintático (Sim-Sim, 1998). Aos dois anos a criança utiliza sobretudo dois tipos de palavras, nomes e verbos (Albuquerque, 2000), aumentando a sua competência morfossintática por volta dos três anos, altura em que passa a produzir frases com categorias puramente gramaticais (e.g., conjunções, artigos, preposições, pronomes átonos), aumentando a sua extensão (Costa & Santos, 2003).

Os cinco anos têm sido apontados na literatura como um marco importante no desenvolvimento morfossintático, nomeadamente com surgimento de frases mais longas, novas estruturas sintáticas produzidas e compreendidas e expansão de constituintes (Gonçalves *et al.*, 2011).

Parece consensual (Gonçalves *et al.*, 2011; Sim-Sim, 1998) que a aquisição morfossintática é pautada por aquisições mais precoces e aquisições tardias, surgindo algumas estruturas depois da entrada no primeiro ciclo, seis anos, que contribui para a estabilização destas aquisições (e.g., compreensão das estruturas passivas ou estruturas coordenadas tardias, como as disjuntivas, e algumas subordinadas). O período que se estende até aos sete anos é extraordinariamente rico na explosão e consolidação de conhecimento morfológico, simultaneamente com o domínio da estruturação frásica (Sim-Sim, 1998).

Em conversação, a criança vai experimentando diferentes tipos de texto, dentro dos quais se insere, como sendo um dos formatos mais precoces, o narrativo (Gonçalves *et al.*, 2011). A narração de uma história é, provavelmente, a primeira situação de monólogo com a qual a criança se depara, sem suporte de um interlocutor. Contar uma história requer que a criança produza sequências de frases que sigam uma estrutura específica de coesão e de coerência (Sim-Sim, 1998). Será dos três até aos quatro anos que se começa a desenvolver a capacidade de elaborar uma narrativa, que mantém uma forma rudimentar até à entrada para a escola (Gonçalves *et al.*, 2011), apesar de já se encontrar adquirida a sua estrutura básica (Sim-Sim, 1998). Depois dos cinco anos a criança apropria-se, progressivamente, de uma maior variedade de funções para as

formas que já utiliza e concretiza novas formas morfossintáticas para velhas funções linguísticas que já utilizava (Gonçalves *et al.*, 2011).

Preece (1987) pesquisou, longitudinalmente, a ocorrência de narrativas no discurso conversacional, entre pares, dos cinco aos sete anos, através de 90 horas de gravações. Concluiu que produziram 14 tipos de narrativas, entre as quais os tipos anedóticos e de fantasia, e que este tipo discursivo surgiu no discurso com elevada regularidade em crianças pequenas. Um estudo levado a cabo por Bento e Befi-Lopes (2010), com 60 crianças dos sete aos dez anos, permitiu concluir que com o aumento da idade a criança tende a realizar narrações mais complexas. Verificou, ainda, que o tipo de sequência de imagem fornecida facilita a produção de discursos mais ou menos complexos.

Avaliação da morfossintaxe: a recolha de produções orais

Quando se avalia a morfossintaxe na criança, os procedimentos e estratégias podem ser variados: testes padronizados, escalas de desenvolvimento, observação do comportamento e testes não padronizados. Dentro destes quatro blocos apontados por Acosta *et al.* (2003), também outros autores (Lund & Duchan, 1983; Miller, 1981; Retherford, 1993) consideram que a recolha de amostras de discurso será o procedimento que confere uma análise mais fidedigna da produção linguística real da criança, produzindo um tipo de informação mais relevante e mais abrangente.

Triadó e Forns (1989), citados por Acosta *et al.* (2003) afirmam que são produzidas até 30% mais estruturas sintáticas do que as geradas em situação de teste, no entanto Crystal (2002) considera que são os testes que permitem avaliar determinadas estruturas que podem não surgir em amostras de discurso espontâneo.

Segundo Diessel (2009), as recolhas de discurso espontâneo são uma abordagem metodológica que tem vindo a ser utilizada desde muito cedo nos estudos de linguagem. O mesmo autor refere que os primeiros registos, baseados em observações diárias por parte dos próprios pais, remontam há mais de cem anos; contudo, a partir dos anos 40 e 50 é que as recolhas ganharam alguma sistematização, em estudos, essencialmente longitudinais, como os de Brain (1963), Brown e Miller (1964) e Ervin (1966), citados por Bloom (1970), que variavam essencialmente na idade dos informantes, tamanho da amostra e procedimentos de recolha.

Desde a década de 70 que se regista a existência de programas computacionais de análise do discurso, mas, apesar do avanço que podem representar, o clínico ou investigador necessita de ser muito cuidadoso para recolher uma amostra representativa das capacidades da criança, através de uma transcrição e codificação apropriadas (Retherford, 1993).

Diferentes variáveis podem afetar a qualidade da amostra, tais como: o local, natureza da interação, dimensão da amostra, os materiais utilizados, o método de gravação e especificação do contexto (Demuth, 1998; Miller, 1981).

As melhores amostras de linguagem são recolhidas em contexto familiar, quando a criança está manifestamente mais relaxada.. Longhurst e Grubb, 1974, citados por Lund e Duchan (1983), concluíram que as situações menos estruturadas são aquelas em que é possível elicitare maior quantidade e complexidade de linguagem. Por vezes, o caráter artificial de alguns contextos comunicativos pode comprometer a representatividade da amostra e não revelar a sua real habilidade linguística. Enquanto não se percebe qual será o melhor local para interação com a criança, Retherford (1993) indica que será preferencial uma recolha variada, em diferentes locais.

Escolher a situação de interação é sobretudo importante quando se pesquisam estruturas linguísticas específicas (Acosta *et al.*, 2003). Em termos de interlocutores, se por um lado Miller (1981) indica a necessidade de realizar diferentes recolhas, com vários interlocutores, incluindo um dos pais, o irmão ou par, e clínico; por outro lado, um estudo levado a cabo por Olswang e Carpenter (1978), que analisou recolhas realizadas pela mãe e pelo clínico, ambos em contexto clínico, permitiu concluir, de forma contraditória, que não havia diferenças nas análises linguísticas, mas apenas na quantidade de enunciados produzidos. Durante a interação com a criança é importante atender a aspetos linguísticos, adequando-os ao seu nível de funcionamento (Acosta *et al.*, 2003). Neste contexto, Retherford (1993) apresenta um conjunto de estratégias a ter em linha de conta.

Ao nível da dimensão, a recolha pode ser realizada tendo como base o tempo ou o número de enunciados produzidos. É recomendado pelo menos meia hora, numa ou em várias situações, como referem os autores Bloom e Lahey, 1978; Crystal, Fletcher, e Garman, 1978, citados por Lund e Duchan (1983); que, segundo Acosta *et al.* (2003), normalmente equivale a 100 enunciados. Este tamanho é, também, aconselhado por outros autores, nomeadamente Lee, 1974; Tyack e Gottsleben, 1974; citados por Lund e

Duchan (1983), havendo referências à utilização de um mínimo de 50, para o cálculo da EME (Chapman, 1981; Retherford, 1993).

A gravação deve contemplar gravação áudio, mas também vídeo ou outra forma de evidenciar aspetos pragmáticos, que podem tornar a análise mais precisa. O equipamento referido deve ser profissional, sendo determinante a qualidade do microfone, e testado antes de proceder às recolhas (Demuth, 1998; Retherford, 1993).

Finalmente, os materiais utilizados devem, segundo Miller (1981), ser selecionados de acordo com os interesses individuais e nível de desenvolvimento (motor, cognitivo e sensorial). Pretende-se que os materiais utilizados na elicitación sejam um aliado para focar a conversação e não um substituto da mesma.

O passo seguinte à recolha da amostra de discurso será a sua transcrição, de forma a atingir os objetivos da análise. A forma como se transcreve variará, naturalmente, consoante os objetivos em causa (Demuth, 1998; Lund & Duchan, 1983). Demuth (1998) indica a importância da codificação do *corpus* linguístico na facilitação da análise para quem já conhece a amostra e, especialmente, para permitir a partilha com outros pesquisadores.

Se por um lado, como já se referiu, diversos autores consideram a recolha de produções orais como um método de análise de elevada riqueza linguística, por outro lado são apontadas algumas limitações. Por vezes, é difícil obter as 100 produções exigidas, sobretudo em crianças mais pequenas; trata-se de um procedimento trabalhoso principalmente ao nível da transcrição; há falta de critérios consensuais no momento da transcrição e segmentação do discurso e, ainda, ausência de índices evolutivos que sirvam de referência na interpretação dos resultados (Acosta *et al.*, 2003). De facto, um estudo muito recente dos autores Finestack, Payesteh, Disher, e Julien (2014), que fizeram uma análise de 726 artigos, nos quais se estudava a EME, através de recolhas de discurso, permitiu concluir que existia elevada inconsistência na metodologia reportada pelos diferentes investigadores.

Extensão Média de Enunciado (EME)

A produção de enunciados, em qualquer língua, implica sempre o conhecimento da sintaxe dessa língua, resultando de uma aquisição progressiva de estruturas gramaticais de complexidade crescente (Sim-Sim, 1998).

O conceito de enunciado é complexo e surge na literatura em diferentes perspectivas, dependendo do nível de análise linguística. Para Campos e Xavier (1991), para além de, tal como a frase, poder ser entendido como uma sequência linguística que corresponde a uma proposição, necessita de evidenciar categorias gramaticais (número, género, aspeto, caso, tempo, diatese, quantificação, determinação em geral) que indicam a construção de um acontecimento linguístico localizado em relação à situação de enunciação. A produção de um enunciado requer a concetualização de uma ideia, a formação gramatical e a codificação fonológica da mesma, o que obedecerá aos padrões de organização sintática da língua (Sim-Sim, 1998). De uma forma mais simplista, pode dizer-se que é uma unidade que constitui o discurso, numa linguagem em funcionamento, voltada para a comunicação (Craig & Washington, 2000). Neste estudo foca-se um tipo de segmentação de enunciados específico, a C-unit (“Communication Unit”), que equivale a uma oração subordinante e todas as subordinadas que dela dependem e trata-se de uma atualização do conceito de “T-unit”, proposto por Hunt (1965), citado por Bloom (1970). Através da contagem de palavras de uma “T-unit” o autor obtinha uma medida de aproximação à complexidade sintática.

Os enunciados são constituídos por palavras, que, segundo Sim-Sim (1998), “são as ferramentas básicas da linguagem; são símbolos que representam uma realidade” (p.121). Para Villalva (2008) a definição de palavra é uma questão complexa. Diferentes abordagens podem ou não convergir nas mesmas unidades, o que é alvo de controvérsia, pois a palavra pode representar uma unidade gráfica, unidade fonológica, unidade lexical, um elemento terminal de estrutura sintática, ou, ainda, definir-se na interface de dois domínios, como o caso da palavra enquanto unidade morfossintática (Azuaga, 1996). Perante a panóplia de possibilidades de estudo das palavras, e atendendo ao foco deste trabalho, considera-se relevante entender a palavra, assim como refere Villalva (2008), como um “átomo sintático” (p.20), ou seja, uma unidade que corresponde a uma categoria gramatical, passível de ocupar um nó terminal na estrutura frásica.

À luz do enquadramento previamente citado, queira-se entender o significado de Extensão Média de Enunciado (EME) como uma média do número de morfemas - EME-

m - ou de palavras -EME-p - que a criança produz num enunciado, isto é, o número total de morfemas ou de palavras, sobre o número total de enunciados produzidos (Shipley & McAfee, 2009).

$$\text{EME} - \mathbf{m} = \frac{\text{n}^\circ \text{ total de morfemas}}{\text{n}^\circ \text{ total de enunciados}}$$

$$\text{EME} - \mathbf{p} = \frac{\text{n}^\circ \text{ total de palavras}}{\text{n}^\circ \text{ total de enunciados}}$$

Muito tempo antes de Roger Brown ficar conhecido como precursor da EME, medida em morfemas (EME-m), foi o autor Nice, em 1925, que idealizou o “Mean Length of Response” (MLR), equiparável à EME-p, pois era medido com palavras. Mais tarde, o autor Brown (1973) estandardizou-o, para a língua inglesa, através de morfemas, pois estava ininteressado em conhecer de que forma a aquisição gramatical destes reflete o crescimento sintático, considerando que seria medido de forma mais precisa pela EME-m do que pela EME-p (MacWhinney, 2000). Com o seu conhecido e extenso trabalho longitudinal, levado a cabo com três crianças – Adam, Eve e Sarah – provou que as crianças têm ritmos de desenvolvimento linguístico diferentes, e é possível, de forma simples e fidedigna, medir o seu desenvolvimento linguístico, através da EME-m, pois, segundo o autor, é um índice que reflete novo conhecimento gramatical: à medida que surgem novas estruturas (e.g., número de papéis temáticos, adição de morfemas obrigatórios, adição de verbos auxiliares, flexão de nomes e verbos, adição de formas negativas e de coordenadas e subordinadas) aumenta o tamanho do enunciado (Brown, 1973). Assim, o autor propõe cinco etapas, que correspondem a diferentes níveis de desenvolvimento linguístico, isto é, encontrou uma ordem de aquisição de 14 morfemas da língua inglesa, à qual fez corresponder uma determinada EME e idade esperada (Tabela 1).

Tabela 1- Etapas de desenvolvimento, valores da EME e idade.
(Brown, 1973)

Estadio	MLU/ EME-m	Idade cronológica (em meses)
I	1.75	15 - 30
II	2.25	28 - 36
III	2.75	36 - 42
IV	3.50	40 - 46
V	4.00	42 - 52 +

Este autor retirou algumas conclusões importantes do seu estudo, nomeadamente que na etapa V a EME-m perde o seu valor de indicador gramatical, pois a criança começa a fazer variar as construções que já domina e será o caráter da interação que fará depender mais a EME-m, em detrimento do próprio conhecimento linguístico da criança. Por outro lado, considera que duas crianças com a mesma EME-m são mais propensas a apresentar semelhanças na complexidade de construção sintática do que duas crianças com a mesma idade cronológica, para o inglês, o que deve ser discutido noutras línguas. Concluiu, finalmente, que a EME-m é um justo indicador de desenvolvimento de morfemas gramaticais, pelo menos até ao estágio V (EME-m = 4), e enquanto a idade isoladamente é um índice pobre de desenvolvimento linguístico, quando conjugada com a EME-m poderá ser um bom preditor de desenvolvimento, ou, da mesma forma, apenas a EME-m isolada (Brown, 1973).

Atualmente, o cálculo da EME-m/EME-p pode realizar-se seguindo métodos rápidos e simples, já há muitos anos disponíveis, a partir de transcrições de linguagem (MacWhinney, 2000). Dois dos programas reconhecidos na transcrição de amostras de discurso são o SALT, de Miller e Chapman, 1987; e o CLAN do projeto CHILDES, de MacWhinney, 1990,1991, citados por Acosta *et al.* (2003). Não obstante, é necessária uma recolha e análise rigorosas. Por essa razão, e para permitir que comparações fossem realizadas, Brown (1973) partilhou os seus critérios, indicando o que incluir na contagem e o que excluir, que passaram a ser utilizados por inúmeros investigadores.

Nas últimas décadas muitos foram os autores que investigaram este índice, procurando verificar a existência de correlação com a idade e com a complexidade gramatical e/ou desenvolvimento linguístico. Rondal, Ghiotto, Bredart, e Bachelet (1987) descobriram que a EME está mais fortemente correlacionada com aumentos de complexidade gramatical nas etapas I e III, de Roger Brown, não existindo correlação a partir daí. Dois anos mais tarde, Klee, Schaffer, May, Membrino, e Mougey (1989) verificaram que a EME se correlaciona com a idade até aos 48 meses (4 anos). Um estudo seguinte, de Blake, Quartaro, e Onorati (1993), com um maior número de sujeitos, verificou uma relação da EME com o *Language Assessment, Remediation and Screening Procedure* (LARSP), que é um instrumento de avaliação gramatical, até ao “MLU” 4,5. Anos depois, Rescorla, Dahlsgaard, e Roberts (2000) verificaram, comparando com outros medidores mais recentemente estudados - *Index of Productive Syntax* (IPSyn) - que existe boa correlação entre os dois índices nos três anos em desenvolvimento típico; nos quatro anos esse aspeto só esteve patente no grupo de crianças com atraso expressivo.

Já mais consistentes com os resultados de Brown (1973) foram as pesquisas das décadas mais recentes. Araujo (2007) realizou um estudo comparativo entre um grupo de 35 participantes com desenvolvimento típico e um grupo, igualmente de 35 participantes, com diagnóstico de Perturbação Específica de Linguagem, de idades compreendidas entre três anos e um mês e seis anos e onze meses. Atendendo ao tamanho da amostra agrupou os de três e quatro anos e os de cinco e seis anos, para efetuar análises da EME-m e EME-p. A autora verificou aumentos da EME-m e EME-p conforme o aumento da idade e concluiu que, apesar de a EME-m e EME-p não serem indicadores de valores de desenvolvimento distintos, houve indícios de a EME-m se relacionar mais com desenvolvimento sintático e a EME-p com desenvolvimento linguístico global. Rice *et al.* (2006) provaram ser válido e fiável considerar a EME um índice de desenvolvimento da linguagem. Por outro lado, os resultados do estudo levado a cabo por Rice *et al.* (2010) deram, também, suporte à validade e fiabilidade da EME como um índice de aquisição normal da linguagem e marca de desenvolvimento atípico.

Ainda no âmbito da caracterização da relação da EME com a idade cronológica, Chapman e Miller (1981) estudaram largas amostras e verificaram uma correlação positiva entre a EME-m e a idade, tendo os últimos autores apresentando valores preditivos da EME em correspondência com idade cronológica, desde os 18 meses até

aos cinco anos de idade (Shiple & McAfee, 2009). Contudo, Klee e Fitzgerald (1985) encontraram evidência de fraca correlação entre EME-m e a idade, principalmente no grupo dos dois anos de idade, e explicaram este fenómeno mostrando que a relação entre EME e idade não é consistente ao longo de todo o leque de idades (Demuth, 1998). Mais tarde, contrariando este estudo, Rondal *et al.* (1987) provaram que as medidas da EME apresentam boa correlação com a idade e com o desenvolvimento gramatical.

Estudos de Araujo, 2003, e Araujo e Befi-Lopes, 2006, citados por Araujo (2007), realizados para o português brasileiro, evidenciam aumentos significativos da EME-m e EME-p com o aumento da idade cronológica da criança, refletindo produção linguística mais elaborada. Destes estudos resultaram, ainda, dados para os valores médios da EME-m e EME-p para crianças entre dois e seis anos, do português brasileiro, fornecendo informação do seu desenvolvimento morfossintático. Ainda no português brasileiro, Fensterseifer e Ramos (2003) estudaram esta medida num grupo de 93 crianças, de um ano e seis meses aos cinco e retiraram semelhantes conclusões, contudo, referem como importante uma utilização complementar de uma avaliação qualitativa da linguagem.

Em Portugal conhecem-se recentes estudos, conjuntos, (Cacela, 2013; Faria, 2013) cada um deles com amostras de 30 participantes, tendo elaborado pesquisas da EME-p, com o objetivo de obter valores padrão para três faixas etárias (englobando os quatro e cinco anos). Procuraram, ainda, relacionar os valores da EME-p com o género da criança e com a escolarização dos pais. Desconhecem-se os resultados do estudo dos cinco anos, no entanto, relativamente aos das faixas dos quatro as conclusões indicam diferenças, nomeadamente na influência da escolaridade dos pais na faixa dos quatro anos e sete meses aos quatro anos e onze meses, sendo a EME-p mais elevada quanto mais elevada a escolarização dos pais.

Como se pode constatar, alguns estudos supramencionados centraram-se tanto no cálculo da EME-m como no da EME-p. Segundo Araujo (2007) a EME-p apresenta sempre um valor mais baixo do que a EME-m e é preferida em estudos que pretendem comparar línguas diferentes, para minimizar os problemas de divergências morfológicas entre as línguas, e, também, porque se torna um meio mais rápido e fácil de analisar.

Um estudo recente (Wieczorek, 2010) verificou uma forte correlação entre a EME-m e EME-p num estudo longitudinal realizado com duas crianças inglesas, com e

sem patologia linguística. Outras pesquisas realizadas anteriormente, na Holanda, Irlanda e Islândia encontraram elevadas correlações entre EME-m e EME-p (Hickey, 1991; Thordardottir & Weismer, 1998), levando vários investigadores a crer que a EME-p será um melhor e mais credível índice para o cálculo da EME (Malakoff et. al., 1999, citado por Araujo, 2007).

Análises da EME em diferentes contextos também vêm referidas na literatura. Leadholm e Miller (1992) verificaram que as medidas de análise linguística, nomeadamente NDW, “Total Number of Words” (TNW) e EME, se correlacionam mais com o desenvolvimento, nas amostras de discurso narrativo, comparativamente com as amostras de discurso conversacional, em crianças em idade pré-escolar e escolar. Os mesmos autores realizaram pesquisas em diferentes contextos, que evidenciam que o discurso espontâneo tende a elicitar enunciados mais curtos do que o discurso narrativo (Leadholm & Miller, 1992). Mais recentemente Miles, Chapman, e Sindberg (2006) realizaram um estudo com 120 participantes de desenvolvimento típico, incluindo crianças, adolescentes e adultos, no qual confirmaram que, em todas as idades estudadas, houve evidências de maior complexidade sintática em discurso expositivo do que em conversacional.

Aplicabilidade e limitações da EME

O interesse em estudar a EME prende-se, não só, mas também com as diferentes vantagens e aplicações que estudos indicam possuir. Foi reconhecida por Chapman e Miller (1981) como importante no rastreio de crianças que necessitem de uma avaliação complementar mais profunda. Outro estudo indica a EME como índice de desenvolvimento gramatical, útil para diagnosticar crianças com perturbações da linguagem (Klee & Fitzgerald, 1985). Poderá ser, ainda, útil para determinar o nível de desenvolvimento de crianças em idade pré-escolar (Klee *et al.*, 1989), um indicador de dominância da língua em casos de bilinguismo (Yip & Matthews, 2006). Em Portugal foi utilizado, na sequência dos estudos anteriormente referidos para o desenvolvimento típico, para pesquisar valores da EME-p em populações de prematuros (Félix, 2014).

A EME revela-se, ainda, um índice valioso na investigação de crianças com perturbações de linguagem (Rice *et al.*, 2010), de etiologias díspares, tais como as perturbações do espectro do autismo (Lopes-Herrera & Almeida, 2008) e a síndrome de

Down, existindo estudos que indicam que esta medida objetiva poderá ser confiável e eficaz na caracterização do desenvolvimento linguístico de indivíduos desta população (Carvalho, 2012). Contudo Marques e Limongi (2011), apesar de apoiarem a afirmação anterior, consideram que se devem considerar todas as variáveis inerentes à EME para uma maior eficiência na identificação do desenvolvimento linguístico e análise das suas alterações. Foi também utilizada em estudos de avaliação da eficácia de tipos de tratamento em pré-escolares com Perturbação Específica da Linguagem (Yoder, Molfese, & Gardner, 2011).

A EME-m tem sido alvo de críticas porque o seu uso foi inicialmente concebido como um instrumento de descrição do inglês. A grande questão é usar essa medida para comparar línguas de morfologia rica, como o italiano ou o português, com línguas de morfologia pobre como o inglês. Realizar esse tipo de comparação, em virtude de o inglês apresentar poucos morfemas, poderia levar a pensar que crianças doutras línguas apresentariam um desenvolvimento linguístico mais adiantado.

É verdade que, ainda antes das publicações de Brown (1973), já o autor Blom (2010) havia concluído que há diferenças importantes a considerar numa análise sintática que se estendem para além de uma tradicional análise da extensão média de enunciado ou outros marcos de desenvolvimento.

Estudos levados a cabo com línguas com morfologia flexional mais rica do que o inglês (língua para a qual a EME tinha regras estipuladas), tais como espanhol, sueco, finlandês e alemão, mostraram dificuldades em adaptar as regras do inglês para o cálculo da EME. Este é um aspeto apontado pelos críticos, além do facto de a EME ser alvo de reduzida investigação e de ser uma medida que, logo que adotada, necessita de ser usada de forma muito consistente, para não incorrer em julgamentos superficiais no diagnóstico de perturbações da linguagem (Hickey, 1991). Para Klee e Fitzgerald (1985) a EME falha como medida de avaliação do nível sintático, a partir dos três anos de idade, comparando-o com outros scores mais válidos.

I. 2. Objetivos do estudo

De acordo com o quadro de referência supracitado, e seguindo o interesse espoletado pelo debate científico em torno do índice EME, não ignorando que existem reservas quanto à sua aplicabilidade em diferentes línguas, torna-se clara a necessidade de caracterizar a EME na língua portuguesa, dos três aos sete anos de idade, já que, de acordo com Sim-Sim (1998), até aos sete anos de idade dá-se um crescimento e consolidação de conhecimento morfológico, acompanhado de um crescente domínio da estrutura sintática.

Assim, tem-se como objetivos do presente trabalho:

Objetivo 1 – Verificar de que forma se correlaciona a EME-p com a idade da criança e comparar as médias da EME-p nas faixas etárias dos três aos sete anos.

Objetivo 2 – Analisar diferenças no valor da EME-p em diferentes contextos discursivos (narrativo *versus* espontâneo).

Objetivo 3 – Comparar a EME-p no grupo pré-escolar e no escolar.

Objetivo 4 – Verificar se a EME-p se correlaciona positivamente com o desenvolvimento linguístico de crianças monolíngues do português europeu.

Objetivo 5 – Comparar a correlação entre a EME-p e a idade da criança com a correlação entre a EME-p e o desenvolvimento linguístico, e verificar qual delas é mais forte, entre grupos etários.

I. 3. Hipóteses

Atendendo aos objetivos um a cinco do presente estudo, e tomando em linha de conta o estado de arte supracitado, consideram-se as seguintes hipóteses:

H1: A EME-p encontra-se correlacionada positivamente com a idade.

Considera-se a formulação desta hipótese sobretudo atendendo a estudos de Rice *et al.* (2006), que revelaram correlação entre a EME-p e a idade dos três aos dez anos.

Pesquisas elaboradas para a EME-p no português brasileiro também provaram esta tendência Araujo (2007).

H2: O desempenho da criança, ao nível da EME-p em discurso narrativo é superior ao seu desempenho em discurso espontâneo.

Pesquisas realizadas em diferentes contextos evidenciam diferenças entre o espontâneo e o narrativo, tendendo o discurso espontâneo a elicitar enunciados mais curtos (Leadholm & Miller, 1992).

H3: No grupo do pré-escolar a EME-p é inferior à verificada no escolar.

Uma vez que a EME se correlacionará fortemente com a idade, um grupo de crianças mais novas apresentarão uma EME-p inferior às crianças em idade escolar; além disso, os estudos de aquisição da linguagem parecem consensuais em assumir que se adquirem estruturas precoces, depois dos seis anos, as tardias, ocorrendo consolidação de conhecimento morfológico e crescendo o domínio da estruturação frásica (Gonçalves *et al.*, 2011; Sim-Sim, 1998).

H4: A EME-p correlaciona-se positivamente com o desenvolvimento linguístico de crianças monolíngues do português europeu.

Estudos de Araujo, 2003, e Araujo e Befi-Lopes, 2006, citados por Araujo (2007), e, ainda, Fensterseifer e Ramos (2003), que estudaram esta medida para o português brasileiro, evidenciam aumentos significativos da EME-p com o aumento da idade cronológica da criança, refletindo produção linguística mais elaborada.

Para corresponder ao objetivo cinco do presente estudo explorar-se-ão possíveis diferenças na força da correlação da EME-p com a idade e na correlação da EME-p com o desenvolvimento da linguagem, partindo do interesse desencadeado pelos estudos de Brown (1973), que assumiu que a EME-m seria um melhor preditor de desenvolvimento linguístico do que a própria idade.

CAPÍTULO II: METODOLOGIA

O presente capítulo encerra a caracterização dos participantes, considerações éticas inerentes ao estudo, materiais utilizados e procedimento até à constituição dos *corpora*.

A metodologia deste trabalho foi desenhada conjuntamente com outra investigadora, Raquel João Campos, cujos objetivos incidiam sobre a mesma amostra populacional, com a finalidade de estudar o Rácio Type-Token e Índice D.

II. 1. Participantes e considerações éticas

Na presente investigação recorreu-se a um tipo de amostragem não probabilística, nomeadamente uma amostragem por conveniência geográfica. Este é um procedimento de seleção de participantes, segundo o qual cada elemento da população não tem uma probabilidade igual de ser escolhido para formar a amostra (Fortin, 1999).

Após a obtenção de autorização para implementação do estudo, por parte da Diretora do Agrupamento de Escolas de Santa Comba Dão (Apêndice A), procedeu-se à distribuição homogénea de 94 documentos para declaração de consentimento informado, e respetiva explicação dos projetos em curso (Apêndice B) por todos os grupos do Ensino Pré-escolar e turmas do 1.º ano do Ensino Básico. Pretendia-se assegurar os princípios éticos da beneficência, respeito pela dignidade humana e justiça.

Assim, os educadores e professores foram elucidados sobre os objetivos do trabalho e foi-lhes solicitado um rastreio prévio, de acordo com critérios de inclusão inicial no estudo:

- crianças monolíngues do português europeu;
- idades compreendidas entre os três anos e 0 meses e os sete anos e onze meses;
- desenvolvimento típico de linguagem.

Posteriormente, foram recolhidos os consentimentos já preenchidos, num total de 84, dos quais apenas 79 se encontravam devidamente preenchidos, portanto, considerados como válidos. Elaborou-se uma pré-seleção dos 50 participantes que melhor diversificavam a amostra, no que concerne à idade, em meses, ao género e ao Centro

Educativo que frequentavam. Contudo, durante a recolha de dados houve necessidade de proceder a alguns ajustes, eliminando-se participantes que se verificou não preencherem os critérios de inclusão prévios, já referidos (desenvolvimento típico de linguagem), e outros definidos posteriormente: gravação de 30 minutos em diferentes situações (duas recolhas), sugerido pelos autores Bloom & Lahey, 1978; Crystal, Fletcher, e Garman, 1978; citados por Lund e Duchan (1983), e um mínimo de 100 enunciados analisáveis, como defendem Lee, 1974; e Tyack e Gottsleben, 1974; citados por Lund e Duchan (1983).

Ao nível dos resultados da avaliação da linguagem foram excluídos dois participantes e dialogou-se com o profissional responsável pelo grupo/turma para proceder a um encaminhamento. Quanto ao número de enunciados totais produzidos excluíram-se três participantes da amostra, e foi, ainda, excluído um outro por não ter comparecido à escola. Estes foram substituídos por seis novos participantes, correspondentes à mesma faixa etária e cumpridores dos critérios necessários, constituindo a amostra populacional final de 50 participantes, caracterizada na Tabela 2. A idade mínima da amostra é três anos e cinco meses e a máxima sete anos e dez meses.

Tabela 2 - Caracterização da amostra da população.

Faixa etária	Número de participantes	Média da idade	Género		Nível de Escolaridade	
			M	F	PE	E
[3;00 – 3;11]	10	3A 8M	1	9	10	0
[4;00 – 4;11]	10	4A 5M	6	4	10	0
[5;00 – 5;11]	10	5A 6M	5	5	10	0
[6;00 – 6;11]	10	6A 5M	6	4	5	5
[7;00 – 7;11]	10	7A 3M	6	4	0	10

II. 2. Procedimento de recolha e materiais

Para a recolha de dados atendeu-se a alguns aspetos cruciais para a análise de enunciados (Lund & Duchan, 1983) nomeadamente: criação de um contexto favorável à interação, conteúdo de interesse para a criança, material adequado à faixa etária, representatividade dos *corpora* e gravação. Para a recolha foi dada ênfase aos procedimentos utilizados pelo proponente da EME (Brown, 1973), mas também outros reconhecidos na literatura internacional (Demuth, 1998; Lund & Duchan, 1983; Miller, 1981; Retherford, 1993), e, ainda, aos procedimentos do estudo de Araujo (2007).

Neste estudo transversal foram realizadas duas sessões de recolha com cada participante, que decorreram na escola frequentada pela criança, em sala própria, à porta fechada, tendo-se procurado eliminar o máximo de estímulos visuais e auditivos. O tempo entre a primeira recolha e a segunda não excedeu sete dias (Apêndice C).

Os materiais utilizados para a recolha de dados incluíram equipamentos de gravação e instrumentos formais de avaliação da linguagem na criança, ambos requisitados na Escola Superior de Saúde (ESS) do Instituto Politécnico de setúbal (IPS); e materiais para elicitación de discurso, adquiridos pelas investigadoras.

Para cada momento de gravação audiovisual, com duração de 30 minutos, em que a investigadora foi a única interlocutora da criança, foram utilizadas duas câmaras digitais HDD da marca Sony e modelo DCR-SR 58 E; respetivos tripés, um da marca Sony e modelo VC R840 e o outro da marca Vivitar e modelo V2400 GX. Para efetuar a gravação áudio utilizaram-se dois gravadores áudio digitais portáteis da marca Sony e modelo ICD-P 520.

No segundo momento de recolha, procedeu-se, também, à administração de testes de avaliação formal da linguagem, adequados à faixa etária da criança, a fim de obter uma medida de avaliação da linguagem independente da EME-p. Escolheu-se o segundo momento, pois a criança já estaria mais ambientada ao local e ao investigador. Os testes formais administrados foram selecionados pelo facto de avaliarem competências sintáticas, de forma formal e padronizada, cujos resultados eram passíveis de converter num valor-referência para o desenvolvimento linguístico (DL) da criança:

- Teste de Avaliação da Linguagem na Criança (Kay & Tavares, 2006)

- Grelha de Observação da Linguagem – Nível Escolar (Kay & Santos, 2003)

Assumi-se o resultado total dos Subtestes da Expressão no Teste de Avaliação da Linguagem na Criança (TALC) (Kay & Tavares, 2006), para as crianças até aos cinco anos, inclusive, e o resultado global dos Subtestes da Estrutura Morfo-sintáctica da Grelha de Observação da Linguagem - Nível Escolar (GOL_E) (Kay & Santos, 2003), para as crianças de seis e sete anos. No TALC a “Parte II: Expressão” engloba os Subtestes “Vocabulário”, “Frases absurdas”, “Constituintes morfossintáticos” e “Intenções comunicativas”, sendo a pontuação máxima 54 pontos. Na GOL_E a “Estrutura Morfo-sintáctica” integra os Subtestes “Reconhecimento de frases agramaticais”, Coordenação e subordinação de Frases”, “Ordem de palavras na frase” e “Derivação de palavras”, constituído cada um deles por 10 itens, podendo equivaler a uma pontuação máxima de 50 pontos.

A preparação das investigadoras para a recolha de dados foi um processo moroso, no ponto de vista de definição de critérios para a interação com a criança, que permitissem reduzir ao máximo a discrepância nas referidas recolhas, e criassem uma atmosfera propícia à produção de muitos enunciados, uniformizando estratégias de comunicação. Seguindo a linha dos autores Lund e Duchan (1983) e Retherford (1993) foram selecionadas atividades mais lúdicas, adaptadas ao interesse da criança, durante as quais se evitou falar muito ou fazer perguntas para preenchimento de pausas, potenciando, ao máximo, uma condução da conversação por parte do participante. Se a criança não iniciava o tópico, utilizaram-se pistas não verbais para estimular um novo turno comunicativo ou promoviam-se questões de resposta aberta ou realizavam-se comentários, recorrendo à espera estruturada e evitando interromper a verbalização da criança.

A interação com o participante envolveu diferentes atividades: diálogo estruturado, jogo simbólico com brinquedos diversificados e narração de histórias com base em imagens sequenciais e livros de imagens. O diálogo estruturado pretendeu ser um momento de integração da criança nas atividades posteriores, bem como uma forma de promover a relação comunicativa. Abordaram-se diferentes temáticas, tendo-se como tópicos pré-definidos alguns inerentes à data da recolha, como a comemoração escolar do Dia da Criança e as eminentes férias de verão, mas, sobretudo, procurando seguir o interesse da criança. Posteriormente, no jogo simbólico, procurou-se dar a escolher à criança a atividade a realizar, de forma a torná-la mais confortável e, assim, permitir a

elicitação de maior quantidade e complexidade de linguagem, tal como afirmam Longhurst e Grubb, 1974, citados por Lund e Duchan (1983). Também o autor Miller (1981) considera que os brinquedos constituem a ferramenta apropriada para elicitar uma amostra mais representativa, pois não canaliza tanto o diálogo, com um jogo de tomada de vez ou um puzzle, ou mesmo a descrição de imagens, que tende a provocar frases feitas.

Apresentaram-se oito conjuntos/*kits* de objetos de diferentes temáticas (Anexo I), tendo o participante selecionado três à sua escolha, para brincar. A investigadora ia exemplificando com os objetos o que se pretendia que a criança fizesse, realizando jogo simbólico e verbalizando o que estava a acontecer (Lund & Duchan, 1983). Na elicitação de discurso narrativo, cada participante foi conduzido a narrar uma história sobre o que observava em sequências de imagens e em livros ilustrados (Anexo I), pois uma série de imagens em sequência potencia mais história do que uma única imagem (Lund & Duchan, 1983).

Divididas pelos dois momentos de gravação, aleatoriamente ou seguindo a opção do participante, utilizaram-se seis sequências de imagens, que divergiram quanto à temática e número de cartões. Sempre que necessário, a investigadora fornecia ajuda na ordenação das mesmas. Nos livros, num total de dois, foi ocultada a componente escrita, evitando influenciar os participantes leitores. Primeiro as crianças viam todas as ilustrações, só depois, retomando ao início, começavam a narrativa oral. O livro mais pequeno, “A Festa de Anos do Tomás”, foi utilizado até aos cinco anos, e o mais extenso, “É Tudo Verdade”, com os participantes mais velhos. Assim, procurou-se que os materiais fossem adequados ao nível de desenvolvimento dos participantes, dirigindo alguns para idades mais precoces e outros para mais tardias. Sempre que a criança se fazia acompanhar de objetos ou jogos significativos, eles foram valorizados e utilizados para promover a interação linguística (Lund & Duchan, 1983).

II. 3. Constituição dos *corpora*

Após a sua recolha, cada investigadora procedeu à transcrição dos dados de 25 participantes, através do software Computerized Language Analysis (CLAN), em ficheiros com formato CHAT, específico para o registo de amostras de fala espontânea. Este programa enquadra-se num sistema de arquivo internacional, denominado Child

Language Data Exchange System (CHILDES) e tem vindo a ser utilizado noutros estudos.

Previamente, e para potenciar a fiabilidade dos resultados, houve necessidade de uniformizar critérios para a definição dos *corpora*, baseados em literatura reconhecida na área da avaliação da linguagem da criança e nos objetivos de cada uma das investigadoras.

Assim, das recolhas efetuadas rejeitaram-se os cinco minutos iniciais da gravação (Araujo, 2007; Brown, 1973; Miller, 1981), transcrevendo-se os 25 minutos restantes, incluindo o discurso da criança e da investigadora (Miller, 1981; Oosthuizen & Southwood, 2009). Foram incluídos apenas os participantes que produziram um mínimo de 100 enunciados analisáveis (Araujo, 2007; Lund & Duchan, 1983; Marques & Limongi, 2011; Miller, 1981; Oosthuizen & Southwood, 2009).

Tal como foi possível verificar no capítulo do enquadramento teórico, tanto o conceito de palavra como de enunciado são complexos. Neste estudo a fronteira de palavra foi definida pela ortografia, à luz do Novo Acordo Ortográfico da Língua Portuguesa, com exceção para as palavras compostas por justaposição e os pronomes clíticos, que foram contabilizados como únicas palavras. Estas exceções encontram-se fundamentadas pelo conceito de palavra à luz da sintaxe, referido por Villalva (2008), pois os pronomes clíticos e as palavras compostas ocupam posições terminais na estrutura sintática, correspondendo a palavras morfológicas. No que concerne, especificamente, aos compostos morfossintáticos, uma interpretação semântica dos mesmos exige que ocupem uma posição sintática terminal, apesar de serem sintaticamente analisáveis, pelo que se consideram uma única palavra, ou expressão semântica lexicalizada, de acordo com a mesma autora. Também Brown (1973) considerou as palavras compostas de forma isolada, na sua pesquisa morfológica, já que não há evidência científica em aquisição de linguagem que justifique o oposto.

Para a tarefa de segmentação de enunciados, assumiram-se perspetivas de diferentes autores para uma segmentação equilibrada, que procurasse cruzar as necessidades inerentes aos estudos de cada investigadora, com critérios utilizados em trabalhos anteriores. Assim, constituíram um único enunciado as frases simples. Fez-se corresponder um enunciado a uma “C-unit” de acordo com o Programa de Transcrição SALT (Andriachii, Nockerts, & Miller, 2012), considerando algumas exceções que se focam adiante.

Neste âmbito de classificação de enunciados, consideraram-se enunciados as frases complexas, coordenadas e subordinadas; integraram-se num enunciado as interrogativas-tag (e.g., “...pois não?”, “..., não foi?”, “..., não é?”), assim como estruturas parentéticas (e.g., “...penso eu”, “acho eu”...etc.) e marcadores de discurso (e.g., “olha,...”). Ao contrário do que preconizam os autores referidos anteriormente, as respostas elíticas e respostas de “sim/não” foram consideradas enunciados dependentes e agregaram-se as frases independentes coordenadas, à exceção das coordenadas com a copulativa “e” como se explanará de seguida.

Houve, ainda, necessidade de tomar decisões em algumas situações específicas. Quando a criança introduziu discurso indireto, segmentou-se o enunciado apenas quando aquela não utilizou uma subordinada completiva (1).

E.g.,

(1) 1

*CHI: e os outros disseram.

*CHI: (es)tamos de folga.

(2) 2

*CHI: e disseram que (es)tava delicioso!

Tal como referido anteriormente, criou-se uma exceção para as frases coordenadas com a copulativa “e”, que constituíram enunciados independentes, sempre que surgia a conjunção (3). Na coordenação de sintagmas nominais ou verbais esta separação de enunciado não ocorreu (4), a não ser que houvesse quebra entoacional antes da conjunção coordenativa.

E.g.,

(4) 3

*CHI: e depois (es)tava a mexei [: mexer] o bolo.

*CHI: e depois a [///] o menino (es)tava a pô(r) est(r)elas.

¹ *Fcheiro 35A*

² *Fcheiro 38A*

³ *Fcheiro 10B*

(5) 4

*CHI: <ligaram o rádio> [/-] depois ligaram o rádio <e foram> [/] e foram ver televisão.

Noutros casos de indecisão na terminação das fronteiras dos enunciados, pelo uso dos critérios anteriormente referidos, atendeu-se aos contornos entoacionais do discurso, assim como às pausas superiores a dois segundos, já que 80% das pausas superiores a dois segundos ocorrem entre enunciados (Miller, 1981).

Foram excluídos da análise da EME-p alguns elementos, que o próprio software CLAN já exclui, por defeito, visto que se baseia nos critérios adotados por Brown (1973), nomeadamente as contagens, sequências (e.g., nomeações), ou outras enumerações, que não correspondessem a um sintagma nominal. Estes critérios também são defendidos por outros autores (Lund & Duchan, 1983; Miller, 1981; Shipley & McAfee, 2009). Por outro lado, ainda na linha de Brown (1973), excluíram-se os enunciados incompletos, sempre que não foram retomados pelo participante; as repetições de palavras ou enunciados, a não ser que fossem utilizados de forma intencional, para enfatizar a informação veiculada; a repetição exata de enunciados proferidos previamente pela investigadora, e produções desviantes, contendo palavras ou sequências ininteligíveis (Wieczorek, 2010), disfluências, assim como as interjeições e onomatopeias que não têm funcionamento de palavra. Não foram incluídas, ainda, produções consideradas supérfluas para a amostra, tais como o canto ou situações pontuais de leitura.

Considerou-se importante a marcação especial de desvios sintáticos e morfofonológicos, tendo-se recuperado a produção-alvo sempre que possível, de modo a possibilitar análises posteriores. No entanto, estes foram contemplados na análise da EME-p, pois Bloom (1970) concluiu no seu estudo que produções acidentais e agramaticais são raras no conjunto de dados de discurso espontâneo da criança, não sendo um fator dominante nas suas produções. Por outro lado, desvios face à gramaticalidade do adulto no seu discurso, que poderão ser sistemáticos, não são considerados erros (p.221). Também se especificaram situações contextuais, além dos enunciados produzidos pela interlocutora, nomeadamente aspetos pragmáticos e não verbais (Retherford, 1993) que poderiam vir a ser úteis futuramente.

⁴ Ficheiro *I4B*

As opções de codificação utilizadas na transcrição com o CLAN, inerentes aos critérios anteriormente descritos (Apêndice D), foram exploradas a partir do Manual do Software (MacWhinney, 2000) e tiveram em conta necessidades específicas de cada estudo assim como a possibilidade de contabilizar alguns elementos discursivos para o cálculo de apenas um dos índices e vice-versa.

De forma a acautelar a maior precisão possível nas transcrições, realizou-se um treino prévio, testando, sucessivamente, a taxa de concordância entre as transcrições até se atingirem, pelo menos, 90%. Utilizando-se trechos de três minutos, de forma aleatória, cada investigadora realizava a sua transcrição, confrontando-se, de seguida, os resultados. A taxa de concordância foi obtida através do rácio entre os resultados dos diferentes índices (EME-p; RTT; D) obtidos por cada uma das investigadoras. Assim, obteve-se 99% de concordância para o RT, 95% para a EME-p e 97% para o D.

Após a transcrição, durante a qual houve tomadas de decisão conjuntas em casos de maior dúvida, constituíram-se os *corpora*, integrando 100 ficheiros em formato CHAT5, dois de cada um dos 50 participantes, correspondentes aos dois momentos de recolha. Ressalva-se a dificuldade verificada no processo de transcrição, ao nível da precisão e atenção exigida, mas, principalmente, por requerer muito tempo e muito rigor na aplicação dos critérios.

⁵ Disponíveis no CD-ROM.

CAPÍTULO III: RESULTADOS

No presente capítulo são apresentados os procedimentos de análise de dados e descrição dos resultados obtidos.

A análise segmentou-se em duas grandes partes. A primeira (análise dos *corpora*) visou a obtenção de resultados para constituição da base de dados, nomeadamente através do programa CLAN e da cotação dos testes de linguagem aplicados. A segunda (descrição dos resultados) direcionou-se para o tratamento estatístico dos dados, recorrendo ao software Statistical Package for Social Sciences (SPSS), com o objetivo de testar as seis hipóteses avançadas previamente, através da estatística indutiva.

III.1 – Análise dos *corpora*

Relativamente à primeira parte, o programa CLAN permitiu, de forma automática, obter os valores da EME-p necessários para este estudo, dando origem a um conjunto de dados, como os constantes da Figura 1.

Figura 1 – Output do programa CLAN, com os resultados do valor da EME-p (mlu).

```
From file <trans1A.cha>
From file <trans1B.cha>
MLU for Speaker: *CHI:
  MLU (xxx, yyy and www are EXCLUDED from the utterance and morpheme counts):
    Number of: utterances = 409, words = 1577
    Ratio of words over utterances = 3.856
    Standard deviation = 2.194
```

Após uma longa pesquisa, chegou-se aos comandos necessários, de forma a contemplar todos os critérios de pesquisa (MacWhinney, 2000):

a) **mlu +t*CHI -t%mor -s"[+ esc]" +u *.cha** – obtenção da EME-p total para cada participante.

Este comando (a) permitiu obter a EME-p (mlu) para os enunciados produzidos na linha ortográfica da criança (+ t*CHI), excluindo-se a análise da linha MOR (utilizada para transcrição morfé mica) inexistente nas transcrições (-t%mor), retirando enunciados que se pretendia excluir desta pesquisa (-s"[+ esc]") e englobando os dois momentos/ficheiros de recolha (+ u), constantes da pasta onde se está a proceder à análise (*.cha).

b) **gem +t*CHI +d1 +n +snarrativo *.cha | mlu -t%mor -s"[+esc]"** – obtenção da EME-p em discurso narrativo.

c) **gem +t*CHI +d1 +n +sespontâneo *.cha | mlu -t%mor -s"[+esc]"** – obtenção da EME-p em discurso espontâneo.

Estes comandos (b e c) permitiram identificar segmentos do discurso, contemplando apenas enunciados produzidos na linha ortográfica da criança (+ t*CHI) e transformá-los num ficheiro com formato CHAT (+d1), agrupando todos os segmentos (+n) delimitados na transcrição, respetivamente com "@G: narrativo " ou "@G: espontâneo " (+snarrativo/+sespontâneo), referente a todos os ficheiros na pasta a analisar (uma por cada participante) (*.cha). Sobre esse ficheiro extrai-se a EME-p (mlu), excluindo-se a análise da linha MOR (utilizada para transcrição morfé mica), inexistente nas transcrições (-t%mor), retirando enunciados que se pretenderam excluir desta pesquisa (-s"[+ esc]").

Verificou-se que, na obtenção da EME-p em diferentes contextos discursivos, naturalmente houve uma redução do número de enunciados analisados, contudo apenas sete participantes deixaram de cumprir o critério definido para o tamanho da amostra de discurso. Assim, os participantes 6, 13, 26, 31, 41 e 42 apresentaram uma amostra de discurso narrativo inferior a 100 enunciados e o mesmo se verificou no discurso espontâneo do participante 24.

No fim desta primeira análise, os dados extraídos do CLAN foram organizados numa base de dados, no SPSS, reunindo as diferentes variáveis em estudo (Apêndice E).

III. 2. Descrição dos Resultados

O tratamento estatístico dos dados foi realizado a partir da estatística descritiva, analisando-se o desempenho dos sujeitos com recurso a medidas de tendência central (média) e medidas de dispersão (Mínimo, Máximo e Desvio Padrão (DP), para a obtenção de valores; e através da estatística inferencial no sentido de testar as hipóteses avançadas previamente e realizar algumas análises adicionais. Estabeleceu-se um $\alpha = 0.05$, para considerar o resultado significativo, e um $\alpha = 0.10$, para considerar um resultado marginalmente significativo⁶.

Para dar cumprimento ao primeiro objetivo deste trabalho, procurando verificar se o desempenho da EME-p se correlaciona positivamente com a idade da criança, procedeu-se a uma correlação linear r de Pearson, entre a idade, em meses, e a EME-p (considerando a EME-p global e a EME-p para cada um dos contextos discursivos - narrativo e espontâneo).

Tabela 3 - Resultados do coeficiente de correlação linear r de Pearson entre os índices em análise e a idade da criança.

	Idade	EME-p	EME-p-DN	EME-p-DE
Idade (meses)	---	0,733**	0,672**	0,597**

Nota: ** $p < 0,01$

Os resultados da Tabela 3 mostram uma correlação positiva entre a EME-p, quer global, quer em discurso narrativo e em espontâneo, e a idade, $p < 0.001$. Para um $n = 50$ pode observar-se uma associação estatisticamente significativa entre a idade dos

⁶ Sempre que as variáveis são quantitativas (EME-p; idade; DL) e o $n > 30$ assumiu-se a normalidade das mesmas, invocando o Teorema do Limite Central (TLC). Quando $n < 30$ testou-se a normalidade através do Teste Shapiro-Wilk. Assumida uma distribuição normal da variável, recorreu-se a testes paramétricos e em caso contrário aos seus equivalentes não paramétricos (Laureano, 2011; Maroco, 2007).

participantes e a EME-p ($r = 0,733$), indicador de que há uma tendência forte para o aumento do desempenho na EME-p com a idade. A mesma tendência se verifica na EME-p em diferentes contextos discursivos ($r_{\text{EME-p-DN}} = 0,672$; $r_{\text{EME-p-DE}} = 0,597$). Comparando estas duas correlações, através do teste Fisher Z, não se verificam diferenças estatisticamente significativas entre elas ($p = 0.542$), o que significa que podem ser consideradas de intensidade equivalente.

Uma vez que existe uma associação entre a EME-p (global e em contextos discursivos específicos) em estudo e a idade, foi-se comparar a distribuição dos valores médios da EME-p global, no discurso narrativo e no discurso espontâneo, por faixa etária. O Gráfico 1 permite visualizar a evolução da EME-p global (EME-p), da EME-p em discurso narrativo (EME-p-DN) e em discurso espontâneo (EME-p-DE), segundo as faixas etárias em análise.

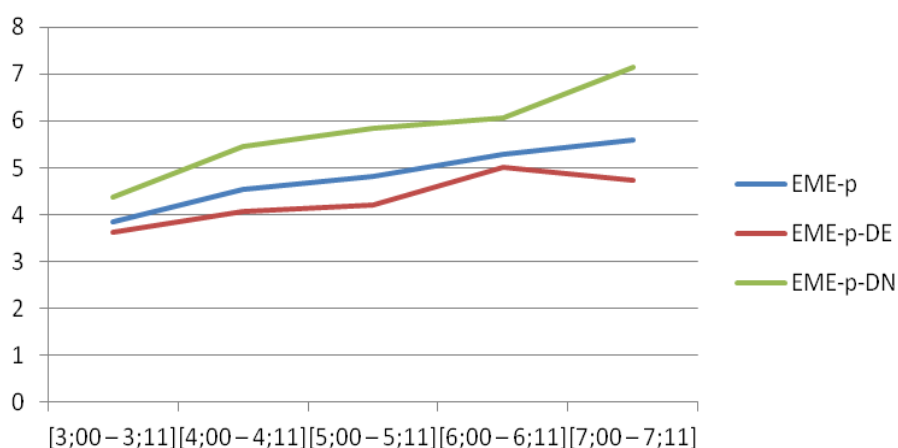


Gráfico 1 – Padrão dos diferentes índices analisados, de acordo com a faixa etária.

Como se pode observar, ainda sem recurso à estatística inferencial, os valores da EME-p situam-se numa posição medial, face aos valores nos diferentes tipos de discurso, sendo sempre o desempenho no narrativo superior ao desempenho no espontâneo.

Para verificar, se há diferenças no valor dos índices entre as várias faixas etárias (10 participantes por faixa etária)⁷ realizou-se o Teste Kruskal-Wallis, que indica existirem diferenças estatisticamente significativas no caso da EME-p, $H(4) = 30,237$, $p < 0,001$; bem como EME-p em discurso narrativo $H(4) = 27,953$, $p < 0,001$, e na EME-p em discurso espontâneo $H(4) = 22,833$, $p < 0,001$.

Com o intuito de verificar entre que faixas residem essas diferenças, comparando os rankings das médias da EME-p (global), realizaram-se comparações *pairwise*.

Tabela 4 - Resultados da EME-p (global) de acordo com a faixa etária.

Faixa etária	EME-p	DP	MIN	MAX
[3;00 – 3;11]	3,856(a,b)	0,475	2,989	4,619
[4;00 – 4;11]	4,533(a,c)	0,541	3,449	5,375
[5;00 – 5;11]	4,825(b ⁸ ,c,d)	0,479	4,064	5,534
[6;00 – 6;11]	5,284(c,d)	0,428	4,594	5,756
[7;00 – 7;11]	5,593(d)	0,677	4,913	7,150

Nota: As médias estatisticamente diferentes, para um $p < .05$ (comparações *pairwise*), estão identificadas com letras diferentes. Letras iguais, nas faixas etárias, indicam que não existem diferenças estatisticamente significativas entre esses grupos.

No caso da EME-p, tal como se pode verificar na Tabela 4⁹, é a faixa dos três anos que mais se distingue das restantes (médias que partilham uma letra não diferem entre si). A EME-p nos três anos ($M_{EME-p [3;00 - 3;11]} = 3,856$) é inferior à da faixa etária dos cinco, $M_{EME-p [5;00 - 5;11]} = 4,825$ (de forma marginalmente significativa, $p = 0,087$); inferior à dos

⁷ A distribuição da variável EME-p, em cada faixa etária ($n=10$) é significativamente diferente de uma distribuição normal na faixa dos sete anos ($SW(10) = 0,838$; $p = 0,041$). Também não segue uma distribuição normal a variável EME-p em discurso narrativo, apenas na faixa dos cinco anos ($SW(10) = 0,771$; $p = 0,006$) e em discurso espontâneo na faixa dos sete anos ($SW(10) = 0,822$; $p = 0,027$).

⁸ Marginalmente significativo ($p = 0,087$)

⁹ Apesar do teste Kruskal-Wallis ser não paramétrico decidiu-se indicar no texto as médias e os desvios-padrão na Tabela 2, de forma a facilitar a interpretação dos resultados e comparação com outros estudos. No Anexo II poderá aceder-se aos resultados das médias de ordens, ou seja, os valores que entram no cálculo deste teste estatístico.

seis ($M_{EME-p\ 6;00-6;11} = 5,284$); e inferior à dos sete anos ($M_{EME-p\ 7;00-7;11} = 5,593$). Por outro lado, a EME-p na faixa etária dos quatro anos apenas se distingue da dos sete, sendo-lhe inferior ($M_{EME-p\ 4;00-4;11} = 4,533$).

Considere-se, agora, cada um dos contextos discursivos, analisando a Tabela 5 (médias seguidas da mesma letra não diferem entre si).

Tabela 5 - Resultados da EME-p, em diferentes contextos discursivos.

Faixa etária	EME-p-DE	DP	MIN	MAX	EME-p-DN	DP	MIN	MAX
[3;00 – 3;11]	3,637 (a) (I)	0,424	2,912	4,434	4,387 (a) (I)	0,805	3,063	5,480
[4;00 – 4;11]	4,078 (a, b) (I)	0,696	3,174	5,495	5,464 (a,b) (II)	0,620	4,030	6,106
[5;00 – 5;11]	4,202 (a, c) (I)	0,538	3,153	4,907	5,853 (b,c) (II)	0,623	4,258	6,423
[6;00 – 6;11]	5,021 (c) (I)	0,414	4,280	5,589	6,051 (b,c) (II)	0,799	4,587	7,431
[7;00 – 7;11]	4,731 (b, c) (I)	0,699	4,066	5,809	7,132 (c) (II)	1,335	5,509	10,289

Nota: As médias estatisticamente diferentes, em coluna, para um $p < .05$ (comparações *pairwise*), estão identificadas com letras diferentes. Assim, letras iguais, nas faixas etárias, indicam que não existem diferenças estatisticamente significativas entre esses grupos. As médias estatisticamente diferentes, em linha, para um $p < .05$ (teste Post-hoc de Tukey), estão identificadas com números romanos diferentes.

Verifica-se que a EME-p-DE na faixa dos três anos ($M_{EME-p-DE\ 3;00-3;11} = 3,637$) é inferior à dos seis ($M_{EME-p-DE\ 6;00-6;11} = 5,021$), $p < 0,001$ e à dos sete anos ($M_{EME-p-DE\ 7;00-7;11} = 4,731$), $p = 0,011$; e a dos seis anos é superior à dos quatro ($M_{EME-p-DE\ 4;00-4;11} = 4,078$), $p = 0,031$.

Quanto à EME-p-DN, existem diferenças estatisticamente significativas que envolvem a faixa etária dos três ($M_{EME-p-DN\ 3;00-3;11} = 4,387$), sendo esta inferior a todas as faixas não contíguas ($M_{EME-p-DN\ 5;00-5;11} = 5,853$, $p = 0,019$; $M_{EME-p-DN\ 6;00-6;11} = 6,051$, $p = 0,005$; $M_{EME-p-DN\ 7;00-7;11} = 7,132$, $p < 0,001$). A EME-p-DN nos sete anos é também superior à dos quatro ($M_{EME-p-DN\ 4;00-4;11} = 5,464$, $p = 0,016$).

Para testar a **H2**, realizou-se uma ANOVA de medidas mistas comparando o desempenho dos participantes em discurso narrativo e em espontâneo (fator repetido ou intra-sujeitos), segundo a faixa etária (fator independente ou inter-sujeitos), que permitiu

constatar que existem diferenças significativas no valor da EME-p nos dois contextos discursivos, sendo a EME-p superior em discurso narrativo: $M_{EME-p-DN} = 5,778$, $DP = 1,229$ vs. $M_{EME-p-DE} = 4,334$, $DP = 0,736$, $F(1, 45) = 165,528$, $p < 0,001$.

Com esta análise, verificou-se, ainda, um efeito de interação entre o tipo de discurso e a faixa etária, $F(4, 45) = 6,418$, $p < 0,001$, indicador de que a média da EME-p difere nos dois tipos de discurso, segundo a faixa etária. Para interpretar as diferenças entre os grupos, utilizou-se o teste Post-hoc de Tukey, cujo resultado está patente na Tabela 5 e é conclusivo quanto à existência de diferenças entre os dois tipos de discurso em todas as faixas etárias, exceto na dos três anos.

Uma análise complementar da correlação r de Pearson (Tabela 6) mostra que a EME-p está fortemente correlacionada com a EME-p-DN ($r = 0,885$) e com a EME-p-DE ($r = 0,883$). O Teste Fisher Z revelou não existirem diferenças estatisticamente significativas entre as duas correlações ($p = 1,000$), pelo que não podemos afirmar que uma correlação seja mais forte do que a outra.

Tabela 6 - Resultados do coeficiente de correlação linear r de Pearson entre os índices analisados.

	EME-p	EME-p-DN	EME-p-DE
EME-p	---	0,885**	0,883**
EME-p-DN		---	0,633**

Nota: ** $p < 0,01$

Com o intuito de testar a **H3**, e verificar se a média da EME-p no grupo de participantes do pré-escolar (PE) é inferior à média no nível escolar¹⁰ (E), realizou-se um

10 A EME-p segue uma distribuição normal no nível escolar ($SW(15) = 0,904$; $p = 0,109$), da mesma forma que a EME-p-DN ($SW(15) = 0,903$; $p = 0,104$) e EME-p-DE ($SW(15) = 0,890$; $p = 0,066$).

Teste t para amostras independentes. Os resultados indicam que existem diferenças estatisticamente significativas (Tabela 7), quer para o valor global da EME-p, $t(48) = -4,318$, $p < 0,001$, quer para discurso narrativo, $t(48) = -3,756$, $p < 0,001$, quer espontâneo, $t(48) = -2,212$, $p = 0,002$.

Tabela 7 - Resultados da EME-p de acordo com o nível de escolaridade.

	EME-p	DP	EME-p-DE	DP	EME-p-DN	DP
Pré-escolar (n=35)	4,546	0,694	4,133	0,693	5,398	0,979
Escolar (n=15)	5,452	0,641	4,802	0,627	6,664	1,328

Pode-se depreender que, em média, os alunos do pré-escolar tiveram um desempenho inferior na EME-p ($M_{\text{EME-p PE}} = 4,546$; $DP = 0,694$ vs. $M_{\text{EME-p E}} = 5,452$; $DP = 0,641$).

Com vista a testar a **H4**, nomeadamente caracterizando a correlação da EME-p com o DL da criança e explorar possíveis diferenças relativas à correlação da EME-p com a idade¹¹, recorreu-se ao R ó de Spearman (Tabela 8). Uma vez que foram administrados dois testes de avaliação da linguagem diferentes, em função das idades dos participantes, foi necessário filtrar a amostra e realizar correlações separadamente para os participantes mais novos (faixa dos três, quatro e cinco anos) e para os mais velhos (seis e sete anos).

¹¹ Para um $n = 20$, a EME-p não segue uma distribuição normal ($SW(20) = 0,903$, $p = 0,047$).

Tabela 8 - Resultados da correlação Ró de Spearman entre a idade, a EME-p e o resultado do teste de avaliação da linguagem.

	Idade (meses)		EME-p		Desenvolvimento da Linguagem	
	Mais novos (3,4,5) (n=30)	Mais velhos (6,7) (n=20)	Mais novos (3,4,5) (n=30)	Mais velhos (6,7) (n=20)	Mais novos (3,4,5) (n=30)	Mais velhos (6,7) (n=20)
Idade (meses)	---	---	0,599**	0,194	0,809**	0,526*
EME-p			---	---	0,676**	0,134
DL					---	---

Notas: ** $p < 0,01$ * $p < 0,05$

É possível verificar que, apenas dos três aos cinco anos, a EME-p se correlaciona positivamente com o desenvolvimento da linguagem ($r_s = 0,676$). Portanto, a correlação entre a EME-p e o desenvolvimento linguístico só é válida para crianças mais novas. Esta partição em grupos, de acordo com os testes de linguagem aplicados, permitiu refinar o resultado da correlação da EME-p com a idade, concluindo-se que só é válida em crianças mais novas, entre os três e cinco anos, pois não é estatisticamente significativa para os seis e sete.

Para comparar a intensidade das duas correlações apresentadas, de forma exploratória, cumprindo-se, assim, o último objetivo do presente estudo, utilizou-se o Teste Fisher Z, tendo-se concluído que não existem diferenças estatisticamente significativas entre as duas correlações ($p = 0,631$), pelo que não podemos afirmar que uma correlação seja mais forte do que a outra.

CAPÍTULO IV: DISCUSSÃO

Ao longo deste capítulo, far-se-á a discussão dos resultados obtidos, à luz do enquadramento teórico que se foi apresentando, e apontar-se-á algumas limitações do estudo e possibilidades de trabalho futuro.

IV. 1. Discussão dos resultados

No sentido de contribuir para a descoberta de indicadores de medida de linguagem em Portugal, o principal objetivo deste estudo foi verificar se o índice EME-p, habitualmente considerado um indicador de desenvolvimento morfossintático, pode ser considerado um indicador de desenvolvimento da linguagem na criança monolíngue do português europeu, através da correlação do desempenho dos participantes na EME-p com o resultado obtido pelos mesmos em Subtestes, provenientes de testes de linguagem padronizados: o TALC e a GOL_E. A EME-p, neste estudo, foi considerada de forma global, contabilizada a partir de *corpora*, contendo discurso narrativo e espontâneo.

Até se chegar ao objetivo final, foram realizadas análises, não menos importantes, para caracterizar, primeiro, a correlação da EME-p com a idade e, posteriormente, verificar diferenças nas médias dos valores por faixa etária; de seguida, a caracterização da EME-p em diferentes tipos de discurso, narrativo e espontâneo, e em grupos de escolaridade, pré-escolar e escolar; bem como, seguidamente, correlacionar este índice com o desenvolvimento linguístico dos participantes, para, de forma exploratória, equiparar-se esta mesma correlação com a verificada com a idade.

De seguida serão sistematizadas as conclusões a que foi possível chegar e discutir-se-ão os resultados com base nas hipóteses formuladas, para perceber em que medida estas receberam apoio empírico. Por outro lado procurar-se-á fazer uma análise comparativa dos resultados com outros estudos, salvaguardando-se, desde já, a complexidade e diversidade metodológica do presente estudo e de estudos similares.

H1: A EME-p encontra-se correlacionada positivamente com a idade.

Os resultados da correlação r de Pearson indicaram uma associação forte entre a EME-p e a idade da criança ($r = 0,733$), que permite concluir que há um aumento de desempenho da EME-p com a idade. Estes dados vão ao encontro dos resultados de Araujo (2007), cuja metodologia de recolha foi equivalente em vários aspetos. A mesma autora cita, na sua obra, estudos de Araujo, (2003), e Araujo e Befi-Lopes (2006), que chegaram ao mesmo resultado para a EME-p, assim como Fensterseifer e Ramos (2003), que confirmaram que a EME-m tem aumentos significativos com a idade, refletindo produção linguística mais elaborada.

Ao verificar as correlações da EME-p-DN e da EME-p-DE com a idade, concluiu-se que apresentam também uma correlação positiva e que a intensidade destas correlações é equivalente. Assim, a EME-p em diferentes contextos discursivos também cresce consoante a idade.

Após obtenção da EME-p de acordo com a faixa etária, verificou-se que existiam diferenças significativas, tanto na EME-p global como nos diferentes contextos discursivos. A EME-p na faixa dos três anos é menor que nas restantes, excepto na dos quatro, e a faixa dos quatro apresenta um valor significativamente menor que a dos sete. Parece, assim, existir um padrão que distingue crianças de faixas mais novas das mais velhas. Quanto às médias da EME-p-DN e EME-p-DE, verifica-se que as principais diferenças, por faixa etária, também envolvem, maioritariamente, a faixa dos três anos em relação às mais distantes. Verificou-se que as diferenças entre faixas etárias da da EME-p global são mais coincidentes com as da EME-p-DN do que com as da EME-p-DE. A Tabela 9 faz uma síntese informativa desses valores médios.

Tabela 9 - Resultados médios da EME-p global, em discurso narrativo e em espontâneo.

Faixa etária	EME-p	EME-p-Dn	EME-p-DE
[3;00 – 3;11]	3,856	4,387	3,637
[4;00 – 4;11]	4,533	5,464	4,078
[5;00 – 5;11]	4,825	5,853	4,202
[6;00 – 6;11]	5,284	6,051	5,021
[7;00 – 7;11]	5,593	7,132	4,731

Estes dados vêm completar estudos anteriores de crianças portuguesas, que não haviam contemplado as faixas dos três anos, dos seis anos e dos sete, nem os diferentes tipos de discurso. Contudo, uma vez que não houve diferenças significativas entre todas as faixas, não é possível assumir que os valores encontrados sejam padrão ou referência para o português europeu e, por outro lado, torna-se pouco pertinente efetuar comparação com médias obtidas noutros estudos. Será importante aumentar os participantes por cada faixa etária para potenciar a representatividade da amostra e, assim, a fiabilidade dos dados.

A hipótese um foi, assim, confirmada através dos resultados obtidos, porque a EME-p apresenta uma correlação positiva forte com a idade. Apesar disso, pôde constatar-se que a análise por faixa etária nem sempre revelou diferenças significativas entre grupos.

H2: O desempenho da criança, ao nível da EME-p em discurso narrativo é superior ao seu desempenho em discurso espontâneo.

Considerando-se a média da EME-p obtida pelos participantes em cada tipo de discurso ($M_{EME-p-DN} = 5,778$, $DP = 1,229$ vs. $M_{EME-p-DE} = 4,334$, $DP = 0,736$), a ANOVA de medidas mistas, veio confirmar que, no discurso narrativo, a média é superior. Também em cada faixa etária isso se verificou, à exceção da faixa dos três anos de idade. Estes dados vêm, assim, na maioria dos casos, confirmar a hipótese avançada, tendo-se comprovado que nos quatro, cinco, seis e sete anos a criança apresenta uma EME-p mais elevada quando produz discurso narrativo do que durante o espontâneo. É sabido que com o aumento da idade a criança tende a realizar narrações mais complexas (Bento & Befi-Lopes, 2010) sendo que tipo de sequência de imagem fornecida facilita a produção de discursos mais ou menos complexos. Neste estudo procurou-se diversificar o tipo de imagem para eliciar discurso narrativo (Anexo I), considerando diferentes materiais-sequências e livros – e diferentes temáticas – aniversário, passeio e aventura, acidente, alimentação, praia e pesca, etc...

A correlação r de Pearson permitiu verificar altos níveis de associação entre a EME-p e a EME-p DN ($r = 0,885$) e entre a EME-p e a EME-p-DE ($r = 0,883$), contudo não se verificaram diferenças entre a força destas correlações, ao contrário do que

evidenciou a pesquisa de Leadholm e Miller (1992) verificando que a EME se correlaciona mais com o desenvolvimento no discurso narrativo.

H3: No grupo do pré-escolar a EME-p é inferior à verificada no escolar.

Quando se compararam os participantes do grupo pré-escolar com os do grupo escolar quanto ao seu desempenho na EME-p global e em diferentes tipos de discurso, o Teste *t* permitiu confirmar que a média no grupo dos primeiros é inferior à dos segundos ($M_{\text{EME-p PE}} = 4,546$; $DP = 0,694$ vs. $M_{\text{EME-p E}} = 5,452$). Esta pesquisa vai ao encontro do que foi descrito no enquadramento teórico e confirma a hipótese avançada, que partia da constatação de que as crianças em idade pré-escolar produzem frases menos elaboradas sintaticamente do que as que se encontram em idade escolar, momento em que se desenvolvem as estruturas sintáticas tardias e se aperfeiçoa a morfossintaxe (Gonçalves *et al.*, 2011; Sim-Sim, 1998).

H4: A EME-p correlaciona-se positivamente com o desenvolvimento linguístico de crianças monolíngues do português europeu.

Para testar esta última hipótese avançada e cumprir um dos principais objetivos do presente trabalho, houve necessidade de dividir os grupos em crianças mais novas (três, quatro e cinco anos) e mais velhas (seis e sete anos), uma vez que não se conhecia um teste de avaliação normativa, padronizado para o português europeu, que permitisse atribuir um valor numérico ao desenvolvimento expressivo, morfossintático ou linguístico global. A correlação de Spearman indicou uma correlação positiva entre a EME-p e o desenvolvimento gramatical dado pelo teste, apenas nas crianças mais novas ($r_s = 0,676$), não sendo a correlação nos mais velhos significativa. Estes achados estão em conformidade com os de Araujo (2007). Assim, a última hipótese formulada não obteve apoio empírico na sua totalidade.

Na sequência da última análise, e apesar de se ter verificado a hipótese um, decidiu-se explorar a correlação da EME-p com a idade nos grupos agora divididos (mais novos e mais velhos). Verificou-se que a EME-p se correlaciona positivamente com a idade apenas nos mais novos (dos três aos cinco anos). Para discutir este novo dado apresentam-se dois aspetos que podem explicar por que razão não se verificou uma correlação da idade com a EME-p nos mais velhos. Por um lado, com a segmentação em

grupos, verificou-se um decréscimo na amostra submetida à correlação, de 50 para 20, no caso dos mais velhos, e alguns autores referem que quanto maior a amostra de discurso, mais válido será o valor encontrado (Brown, 1973; Lund & Duchan, 1983). Por outro lado, este resultado parece ter alguma compatibilidade com aquele que nos indica que apenas nas crianças mais novas podemos encarar a EME como indicador de desenvolvimento linguístico. Os cinco anos têm sido descritos como um marco de desenvolvimento morfossintático, surgindo enunciados mais extensos e complexos nesta faixa etária (Gonçalves *et al.*, 2011), o que parece estar em concordância com os dados apresentados. Chapman e Miller (1981), através de amostras de 123 participantes, conseguiram estabelecer uma relação preditiva entre a EME e a idade cronológica, o que reforça a existência de uma boa correlação.

Finalmente, e de forma exploratória, quando se realizou o teste Fisher Z para descobrir se havia alguma correlação mais forte do que a outra, no que concerne às correlações entre a EME-p e a idade, por um lado, e entre a EME-p e o desenvolvimento linguístico, por outro, não se verificaram diferenças. Brown (1973) admitiu que a EME-m estaria mais associada ao desenvolvimento linguístico do que a própria idade. Contudo, não se pode esquecer que o seu estudo se referia ao inglês e analisava morfemas. Para além disso, os procedimentos metodológicos foram diferentes.

A análise dos resultados leva a considerar que o índice EME-p poderá ser válido para complementar uma avaliação linguística, na medida em que se reconhece, mais uma vez, o seu valor como indicador de desenvolvimento linguístico. Contudo, e visto que não houve diferenças significativas entre todas as faixas etárias, à luz do presente estudo, não poderá ser um critério de diagnóstico para terapeutas da fala. De acordo com Crystal (2002), e atendendo à necessidade de capacitar o clínico de conhecimento e confiança para alcançar as decisões mais acertadas, uma análise gramatical nos seus diferentes domínios (semântica lexical e composicional, morfologia, sintaxe, fonologia e pragmática – e suas interfaces) permitirá clarificar, descrever, fazer um diagnóstico diferencial, avaliar a natureza e gravidade das dificuldades e intervir de forma mais adequada.

O processo metodológico desta dissertação ocupou um lugar de destaque em todo o trajeto, pois demorou mais tempo do que o expectável e desejável, principalmente ao nível da transcrição. Torna-se difícil enquadrar este tipo de trabalho no dia a dia de um terapeuta da fala, não só pela morosidade, como também pela

dificuldade em definir critérios uniformes, tal como já foi apontado na literatura como limitação na utilização deste índice.

IV. 2. Limitações do estudo e trabalho futuro

Considera-se que este estudo foi um contributo valioso, no que concerne ao fornecimento de novos dados em avaliação da linguagem. No entanto, neste momento do presente trabalho é importante reconhecer-lhe algumas limitações que podem, de alguma forma, ter condicionado os resultados obtidos.

Os estudos com base em recolhas orais são já muito antigos, assim como a discussão das limitações inerentes a esta metodologia. Atualmente existe, ainda, elevada inconsistência na metodologia utilizada nas diferentes investigações (Finestack *et al.*, 2014). Neste trabalho, a presença de ruído em algumas recolhas, na escola, proveniente dos intervalos e, ainda, algumas interrupções pontuais de funcionários, pode ter sido um fator condicionante. Demuth (1998) exemplifica diversas situações como potenciais bloqueios do canal de comunicação e num estudo realizado por Dijk e Geert (2005) analisaram-se desvios na qualidade dos resultados da EME devido a dificuldades na interpretação das produções linguísticas, que resultam em frases incompletas ou sequências ininteligíveis, devido a fatores inerentes aos cenários da recolha. Visto que, neste estudo, foram excluídas da amostra frases incompletas e sequências ininteligíveis pode considerar-se uma limitação na análise da EME. Por outro lado, apesar do uso de estratégias verbais controladas, e pré-definidas à luz das que foram utilizadas por Rice *et al.* (2010), não existiu um treino prévio entre investigadoras. Por outro lado, não houve um conhecimento prévio do participante, o que poderia ter favorecido o estabelecimento de relação.

Outra das limitações prende-se com a técnica de amostragem utilizada, por conveniência geográfica, que apresenta a desvantagem de ser menos representativa do que a amostragem probabilística (Fortin, 1999).

Há, ainda, outro aspeto que se prende com o desenho do estudo. A utilização de testes de linguagem diferentes pode, de alguma forma, ter condicionado os dados, pois

um dos testes avaliou a expressão da linguagem, e outro a expressão morfossintática, tendo-se feito equivaler ambos ao desenvolvimento da linguagem.

Perante as limitações apontadas, considera-se importante fazer algumas sugestões de melhoria em estudos posteriores, bem como propostas de realização de trabalhos complementares. Neste sentido, seria importante expandir a amostra, não só em tamanho mas também promovendo um alargamento geográfico.

Os *corpora*, contendo recolhas superiores a 100 enunciados de cada um dos 50 participantes, apresentam riqueza para possibilitar análises futuras, o que, de acordo com Retherford (1993) constitui uma mais-valia associada às recolhas de produções orais, visto que se poderá ganhar tempo em investigações futuras. Uma vez que foi utilizado o software CLAN, pertencente ao projeto CHILDES, que, por sua vez, integra uma base de dados internacional, poderá constituir trabalho futuro a proposta de inclusão dos *corpora* do presente estudo, nesta base de dados.

A partir das transcrições realizadas, seria interessante realizar um estudo qualitativo das produções das crianças, não só para verificar se existe correlação entre o desenvolvimento morfossintático e EME-p, como também caracterizar frases agramaticais que já se encontram codificadas. Seria, ainda, pertinente acrescentar uma codificação em morfemas às linhas transcritas, de forma a poder comparar o desempenho da EME-p com EME-m e verificar qual seria um melhor indicador de desenvolvimento linguístico. No presente trabalho, atendendo à morfologia do português, esta análise foi colocada de parte, pelas limitações de tempo.

CONCLUSÃO

Reconhecendo, enquanto terapeuta da fala, a necessidade de amplificação e de validação de novos instrumentos ou índices de desenvolvimento, e atendendo à escassez de dados no português europeu, foi natural a escolha da EME como tema da presente dissertação. Trata-se de uma medida já reconhecida noutros países como um índice válido na avaliação da linguagem, cuja aplicabilidade é extensível a estudos científicos, no âmbito da normalidade e da patologia, e a processos de intervenção terapêutica.

O presente estudo permitiu obter novos dados acerca da EME-p da criança, apesar de não ter fornecido valores que se possam considerar referência para profissionais, como o terapeuta da fala, se basearem nos mesmos como critério e diagnóstico.

Apesar de ser discutível a representatividade da amostra, foram confirmadas todas as hipóteses, salvaguardando-se algumas limitações na pesquisa. Foi obtido apoio empírico para a hipótese um, tendo-se verificado uma associação forte entre a idade e a EME-p; contudo, salvaguardando limitações na amostra, numa pesquisa em grupos de mais novos e mais velhos constatou-se que apenas nos mais novos essa correlação existia. Será, ainda, importante referir que não houve diferenças significativas entre muitas faixas etárias, situando-se nos três anos as principais. Também a segunda hipótese formulada foi confirmada, pois o valor da EME-p em discurso narrativo é estatisticamente superior ao seu valor em discurso espontâneo; da mesma forma que se validou a hipótese três, visto que no grupo pré-escolar a EME-p é inferior à obtida no grupo de escolares.

Finalmente, os resultados indicam uma correlação positiva entre a EME-p e o desenvolvimento da linguagem de crianças monolíngues do português europeu, até aos cinco anos de idade, pelo que não se pode validar totalmente a hipótese quatro. Apesar disso concluiu-se que a EME-p é um indicador de desenvolvimento linguístico expressivo, correlacionado com os valores obtidos pela parte expressiva do Teste de Avaliação da Linguagem na Criança (Kay & Tavares, 2006).

Os *corpora* constituídos, contendo recolhas superiores a 100 enunciados de cada um dos 50 participantes, apresentam riqueza para possibilitar análises futuras e, possivelmente, integrar a base de dados internacional CHILDES.

Assim, os resultados conduziram a conclusões importantes para o português europeu, podendo espoletar a curiosidade científica em dar continuidade ao estudo da EME em Portugal. Enquanto terapeuta da fala, considera-se que foi uma mais valia aprofundar conhecimentos no âmbito da avaliação da linguagem através da recolha de produções orais, considerando importante, nesta etapa final, reforçar a necessidade, já apontada por alguns autores, de se realizar uma avaliação qualitativa do perfil linguístico da criança para uma avaliação menos restrita. A metodologia de obtenção dos valores da EME pode não ser acessível a todos os terapeutas da fala em Portugal, pela morosidade e pela dificuldade de implementação e uniformização de critérios.

Em suma, considera-se que a presente dissertação constituiu um contributo valioso, permitindo dar mais um passo na busca de indicadores de medida objetivos, que poderão vir a ser utilizados pelo terapeuta da fala ou outro profissional habilitado para a avaliação da linguagem na criança.

REFERÊNCIAS BIBLIOGRÁFICAS

- Acosta, V. M., Moreno, A., Ramos, V., Quintana, A., & Espino, O. (2003). *Avaliação da Linguagem – Teoria e Prática do Processo de Avaliação do Comportamento Linguístico Infantil*. São Paulo: Livraria Santos Editora.
- Albuquerque, F. (2000). *À Descoberta da Palavra Redondinha*. Porto: Porto Editora.
- Andriachii, K., Nockerts, A., & Miller, J. F. (2012). *Assessing Language Production Using SALT Software*.
- Aram, D. M., Morris, R., & Hall, N. E. (1993). Clinical and Research Congruence in Identifying Children With Specific Language Impairment. *Journal of Speech and Hearing Research, 36*(3), 580-591.
- Araujo, K. (2007). *Desempenho Gramatical de Criança em Desenvolvimento Normal e com Distúrbio Específico de Linguagem.*, Universidade de São Paulo, São Paulo.
- Azuaga, L. (1996). Morfologia. In I. H. Faria, E. R. Pedro, I. Duarte & C. A. M. Gouveia (Eds.), *Introdução à Linguística Geral e Portuguesa*. Lisboa: Caminho.
- Bento, A. C. P., & Befi-Lopes, D. M. (2010). Organização e narração de histórias por escolares em desenvolvimento típico de linguagem. *Pró-Fono Revista de Atualização Científica, 22* (4), 503-508.
- Blake, J., Quartaro, G., & Onorati, S. (1993). Evaluating quantitative measures of grammatical complexity in spontaneous speech samples. *Journal of Child Language, 20*(1), 139-152.
- Blom, E. (2010). Effects of input on the early grammatical development of bilingual children. *International Journal of Bilingualism, 14* (4), 422-446.
- Bloom, L. (1970). *Language Development*. Cambridge (MA): MIT Press.
- Brain, M. D. S. (1963). The ontogeny of English phrase structure: The first phase. *Language, 39*, 1-13.
- Brown, R. (1973). *A First Language*. Cambridge (MA): Harvard University Press.

- Cacela, M. (2013). *Extensão média do enunciado : obtenção de valores padrão para o português europeu em crianças de 4 anos (4;00-4;05)*. (Dissertação de Mestrado), Universidade Católica Portuguesa, Lisboa.
- Campos, M. H. C., & Xavier, M. F. (1991). *Sintaxe e Semântica do Português*. Lisboa: Universidade Aberta.
- Carvalho, A. M. d. A. (2012). *Extensão Média do Enunciado em crianças com síndrome de Down*. (Mestre), Universidade de São Paulo, São Paulo.
- Chapman, R. (1981). Computing Mean Length of Utterances in Morphemes. In J. F. Miller (Ed.), *Assessing Language Production in Children*. Austin, Texas: PRO-ED.
- Chapman, R., & Miller, J. F. (1981). The Relation Between Age and Mean Length of Utterance in Morphemes. *Journal of Speech, Language and Hearing Research*, 24, 154-161.
- Costa, J., & Santos, A. L. (2003). *A falar como os bebés*. Lisboa.: Caminho.
- Craig, H. K., & Washington, J. A. (2000). An Assessment Battery for Identifying Language Impairments in African American Children. *Journal of Speech, Language, and Hearing Research*, 43(2), 366-379.
- Crystal, D. (2002). Clinical Linguistics. In M. Aronoff & J. Rees-Miller (Eds.), *The Handbook of Linguistics* (pp. 673-682). Oxford: Blackwell.
- Demuth, K. (1998). Collecting Spontaneous Production Data. In D. McDaniel, C. McKee & H. S. Cairns (Eds.), *Methods for Assessing Children's Syntax*. London: MTT Press.
- Diessel, H. (2009). Corpus linguistics and language acquisition: An International Handbook. In A. Ludeling & M. Kyto (Eds.), *Corpus Linguistics* (Vol. II, pp. 1197-1212). Berlin: Walter de Gruyter GmbH & Co. KG.
- Dijk, M. v., & Geert, P. v. (2005). Disentangling behavior in early child development: Interpretability of early child language and its effect on utterance length measures. *Infant Behavior & Development*, 28, 99-117.

- Faria, S. d. S. B. L. d. (2013). *Extensão Média do Enunciado: obtenção de valores padrão para o português europeu em crianças de 4 anos (4;07 - 4;11)*. (Dissertação de Mestrado), Universidade Católica Portuguesa, Lisboa.
- Félix, J. F. L. (2014). *Extensão Média do Enunciado em Palavras (EME-P) de Crianças com 4 e 5 anos Nascidas Pré-termo*. (Mestrado), Universidade Católica Portuguesa, Lisboa.
- Fensterseifer, A., & Ramos, A. P. F. (2003). Extensão média de enunciados em crianças de 1 a 5 anos. *Pró-fono*, 15, 251-258.
- Finestack, L. H., Payesteh, B., Disher, J. R., & Julien, H. M. (2014). Reporting Child Language Sampling Procedures. *Journal of Speech, Language and Hearing Research*, 57, 2274-2279.
- Fortin, M.-F. (1999). *O Processo de Investigação - Da concepção à realização: Lusociência*.
- Gonçalves, F., Guerreiro, P., & Freitas, M. J. (2011). *O Conhecimento da Língua: Percursos de Desenvolvimento*. Lisboa: Ministério da Educação - Direção-Geral de Inovação e Desenvolvimento Curricular
- Hickey, T. (1991). Mean length of utterance and the acquisition of Irish. *Journal of Child Language*, 18(3), 553 - 569.
- Kay, E. S., & Santos, M. E. (2003). *Grelha de Observação da Linguagem, Nível Escolar*: Escola Superior de Saúde do Alcoitão.
- Kay, E. S., & Tavares, M. D. (2006). *Teste de Avaliação da Linguagem na Criança (TALC)*: Oficina Didáctica.
- Klee, T., & Fitzgerald, M. D. (1985). The relation between grammatical development and mean length of utterance in morphemes. *Journal of Child Language*, 12, 251-269.
- Klee, T., Schaffer, M., May, S., Membrino, I., & Mougey, K. (1989). A comparison of the age-MLU relation in normal and specifically language-impaired preschool children. *Journal of Speech and Hearing Disorders*, 54(2), 226-233.
- Laureano, R. M. S. (2011). *Testes de Hipóteses com o SPSS*. Lisboa: Edições Sílabo.
- LDA, L. D. A.: LDA, Learning Development Aids.

- LDA, L. D. A. Sequential Thinking, Level 4: LDA, Learning Development Aids.
- LDA, L. D. A. Sequential Thinking, Level 5: LDA, Learning Development Aids.
- Leadholm, B. J., & Miller, J. F. (1992). *Language Samples Analysis: The Wisconsin Guide*: Wisconsin Department of Public Instruction.
- Lopes-Herrera, S. A., & Almeida, M. A. (2008). O uso de habilidades comunicativas verbais para aumento da extensão média de enunciados no autismo de alto funcionamento e na síndrome de Asperger. *Pró-Fono Revista de Atualização Científica*, 20 (1), 37-42.
- Lund, N. J., & Duchan, J. F. (1983). *Assessing Children's Language in Naturalistic Contexts*. Englewood Cliffs, New Jersey: Prentice-Hall.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3.^a ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Maroco, J. (2007). *Análise Estatística Com Utilização do SPSS* (3.^a ed.): Edições Sílabo.
- Marques, S. F., & Limongi, S. C. O. (2011). A extensão média do enunciado (EME) como medida do desenvolvimento de linguagem de crianças com síndrome de Down. *Jornal da Sociedade Brasileira de Fonoaudiologia*, 23(2), 157-157.
- Miles, S., Chapman, R., & Sindberg, H. (2006). Sampling Context Affects MLU in the Language of Adolescents With Down Syndrome. *Journal of Speech, Language, and Hearing Research*, 49, 325-337.
- Miller, J. F. (1981). *Assessing Language Production in Children*. Austin, Texas: PRO-ED.
- Most, N. (2011). *É tudo verdade!* Estoril: Minutos de Leitura.
- Olswang, L. B., & Carpenter, R. L. (1978). Elicitor effects on the language obtained from young language-impaired children. 43, 76-88.
- Oosthuizen, H., & Southwood, F. (2009). Methodological Issues in the Calculation of Mean Length of Utterance. *South African Journal of Communication Disorders*, 56(1), 76-87.
- Preece, A. (1987). The range of narrative forms conversationally produced by young children. *Journal of Child Language*, 14, 353-373.

- Prieto, P., Estrella, A., Thorson, J., & Vanrell, M. D. M. (2011). Is prosodic development correlated with grammatical and lexical development? Evidence from emerging intonation in Catalan and Spanish. *Journal of Child Language*, 28(1), 1-37.
- Rescorla, L., Dahlsgaard, K., & Roberts, J. (2000). Late-talking toddlers: MLU and IPSyn outcomes at 3:0 and 4:0. *Journal of Child Language*, 27(3), 643-664.
- Retherford, K. S. (1993). *Guide to Analysis of Language Transcripts* (2.^a ed.). Eau Claire (WI): Thinking Publications.
- Rice, M. L., Redmond, S. M., & Hoffman, L. (2006). Mean Length of Utterance in Children With Specific Language Impairment and in Younger Control Children Shows Concurrent Validity and Stable and Parallel Growth Trajectories. *Journal of Speech, Language, and Hearing Research*, 49, 793-808.
- Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010). Mean Length of Utterance Levels in 6-Month Intervals for Children 3 to 9 Years With and Without Language Impairments. *Journal of Speech, Language and Hearing Research*, 53, 333 - 349.
- Rondal, J. A., Ghiotto, M., Bredart, S., & Bachelet, J.-F. (1987). Age-relation, reability and grammatical validity of measures of utterance length. *Journal of Child Language*, 14(3), 433-446.
- Santos, M. E., Afonso, C., Taveira, L., Carvalho, T. N. d., Neto, N., Pimenta, C., . . . Mina, S. P. d. (2006). *Levantamento de instrumentos de avaliação de linguagem usados em Portugal, no âmbito da Terapia da Fala*. Unidade Curricular "Avaliação da Linguagem Oral e Escrita". Escola Superior de Saúde do Alcoitão e Universidade Católica Portuguesa Alcoitão.
- Schubi. Obrigado, KIM!
- Shipley, K. G., & McAfee, J. G. (2009). *Assessment in Speech-Language Pathology* (4.^a ed.). Delmar: Clifton Park.
- Sim-Sim, I. (1998). *Desenvolvimento da Linguagem*. Lisboa: Universidade Aberta.
- Simas, H. (2003). *A Festa de Anos do Tomás*. Estoril: Minutos de Leitura.

- Thordardottir, E. T., & Weismer, S. E. (1998). Mean length of utterance and other language sample measures in early Icelandic. *First Language*, 18, 1-32.
- Villalva, A. (2008). *Morfologia*. Lisboa: Universidade Aberta.
- Wieczorek, R. (2010). Using MLU to Study Early Language Development in English. *Psychology of Language and Communication*, 14 (2), 59-69.
- Yip, V., & Matthews, S. (2006). Assessing Language Dominance Bilingual Acquisition: A Case for Mean Length Utterance Differentials. *Language Assessment Quarterly*, 3(2), 97-116.
- Yoder, P. J., Molfese, D., & Gardner, E. (2011). Initial Mean Length of Utterance Predicts the Relative Efficacy of Two Grammatical Treatments in Preschoolers With Specific Language Impairment. *Journal of Speech, Language and Hearing Research*, 54, 1170 - 1181.

LISTA DE TABELAS

Tabela 1 - Etapas de desenvolvimento, valores da EME e idade.	10
Tabela 2 - Caracterização da amostra da população.....	18
Tabela 3 - Resultados do coeficiente de correlação linear r de Pearson entre os índices em análise e a idade da criança.....	28
Tabela 4 - Resultados da EME-p (global) de acordo com a faixa etária.	30
Tabela 5 - Resultados da EME-p, em diferentes contextos discursivos.	31
Tabela 6 - Resultados do coeficiente de correlação linear r de Pearson entre os índices analisados.	32
Tabela 7 - Resultados da EME-p de acordo com o nível de escolaridade.....	33
Tabela 8 - Resultados da correlação R_s de Spearman entre a idade, a EME-p e o resultado do teste de avaliação da linguagem.....	34
Tabela 9 - Resultados médios da EME-p global, em discurso narrativo e em espontâneo.	36
Tabela 10 - Caracterização dos participantes na recolha de dados.....	viii
Tabela 11 - Opções de codificação no programa CLAN.....	x
Tabela 12 - Base de dados do estudo.....	xv
Tabela 13 - Resultado das médias das ordens da EME-p, de acordo com a faixa etária.	xxii

LISTA DE FIGURAS

Figura 1 – Output do programa CLAN, com os resultados do valor da EME-p (mlu)...	26
Figura 2 - Tabuleiro de base para o Kit dos bombeiros, construtores e polícias.....	xvii
Figura 3 - Conjunto de brinquedos inerentes à temática das construções.....	xvii
Figura 4 - Conjunto de brinquedos inerentes à temática dos bombeiros.....	xvii
Figura 5 - Conjunto de brinquedos inerentes à temática dos polícias	xvii
Figura 6 - Conjunto de brinquedos inerentes à temática do parque.....	xvii
Figura 7 - Conjunto de brinquedos inerentes à temática dos polícias.....	xvii
Figura 8 - Conjunto de brinquedos inerentes à temática da casa de bonecas.....	xvii
Figura 9 - Conjunto de brinquedos inerentes à temática dos médicos.....	xvii
Figura 10 - Conjunto de brinquedos inerentes à temática das ferramentas.....	xvii
Figura 11 – Livros “É Tudo Verdade” e “A Festa de Anos do Tomás”.....	xviii
Figura 12 – Sequência de oito cartões (ida ao médico).....	xviii
Figura 13 – Sequência de seis cartões (acidente na estrada).....	xix
Figura 14 – Sequências de cinco cartões (na praia).....	xix
Figura 15 – Sequências de cinco cartões (na pesca).....	xx
Figura 16 – Sequências de 6 cartões (os biscoitos).....	xx
Figura 17 – Sequências de 6 cartões (acabou o sumo).....	xx

LISTA DE GRÁFICOS

Gráfico 1 – Resultados da EME-p, de acordo com a faixa etária.....	29
---	----

APÊNDICE A

Consentimento informado à Diretora do Agrupamento

Exma. Senhora Diretora
do Agrupamento de Escolas de Santa Comba Dão, Dra. Madalena Dinis

Assunto: Pedido de Colaboração em Projetos de Mestrado

Ana Rita Coimbra Pina de Campos, Terapeuta da Fala com a cédula profissional n.º C-023473185, e Raquel João Pinheiro Marques de Campos, Terapeuta da Fala com a cédula profissional n.º C-023116188, alunas do segundo ano do curso de mestrado em Desenvolvimento e Perturbações da Linguagem na Criança, da Escola Superior de Saúde do Instituto Politécnico de Setúbal e da Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa, vêm, por este meio, solicitar a V/ autorização para proceder à recolha de amostras de discurso e aplicação de uma prova de avaliação formal da linguagem oral, em alunos com desenvolvimento típico da linguagem do Agrupamento de Escolas de Santa Comba Dão, no âmbito dos projetos de dissertação de Mestrado subordinados aos temas “Rácio Type-Token e D como indicadores de desenvolvimento linguístico no português europeu.” e “Extensão Média de Enunciado como indicador de desenvolvimento linguístico no português europeu.”, sob orientação da professora Doutora Maria Lobo e coorientação da professora Professora Eileen Sua Kay.

Sumariamente, os projetos supracitados têm como objetivos a verificação da utilidade dos parâmetros diversidade vocabular e extensão média de enunciado, na avaliação do desenvolvimento da linguagem da criança e, conseqüentemente, na prática profissional dos terapeutas da fala. Pretende-se contribuir para a obtenção de valores referenciais dos índices D, Rácio Type-token e Extensão Média de Enunciado, para o português europeu. De modo a atingir os objetivos propostos será necessário efetuar a recolha de amostras de discurso de 50 crianças com idades compreendidas entre os três anos e 0 meses e os sete anos e onze meses, assegurando-se a assinatura prévia de um consentimento informado por parte do Encarregado de Educação. Para uma explicação mais detalhada do estudo consultar documento em anexo.

Desde já, agradecemos a atenção dispensada.

Com os melhores cumprimentos,

(Ana Rita Coimbra Pina de Campos (Raquel João Pinheiro Marques de Campos)

Concedo autorização

Não concedo autorização

(Assinatura da Diretora do Agrupamento)

APÊNDICE B

Consentimento informado aos Encarregados de Educação

Declaração de Consentimento Informado

Participação nos Projetos de Mestrado

“Rácio Type-Token e D como indicadores de desenvolvimento linguístico no português europeu” e “Extensão Média de Enunciado como indicador de desenvolvimento linguístico no português europeu”

Eu, abaixo-assinado(a) _____, na qualidade de encarregado(a) de educação do(a) aluno(a) _____, data de nascimento ____/____/____, a frequentar o ____º ano, na escola/jardim de infância (riscar o que não interessa) _____, declaro que autorizo de livre vontade a participação do(a) meu(minha) educando(a) nos estudos descritos anteriormente.

Compreendi a informação que me foi dada, tive a oportunidade de fazer perguntas e as minhas dúvidas foram esclarecidas.

Foi-me garantido que todos os dados relativos à identificação dos participantes neste estudo são confidenciais e que serão mantidos no anonimato.

Sei que posso recusar-me a autorizar a participação ou interromper a qualquer momento a participação no estudo, sem nenhum tipo de penalização por este facto.

Concordo que seja efetuada a necessária recolha audiovisual de dados. Também autorizo a divulgação dos resultados obtidos no meio científico.

_____, ____ de _____ de 2011

(Assinatura do(a) encarregado(a) de educação)

(Assinatura das investigadoras)

Explicação dos Projetos de Mestrado

“Rácio Type-Token e D como indicadores de desenvolvimento linguístico no português europeu” e “Extensão Média de Enunciado como indicador de desenvolvimento linguístico no português europeu”

Os projetos de mestrado subordinados aos temas *“Diversidade Vocabular como medida de desenvolvimento no português europeu”* e *“Extensão Média de Enunciado como medida de desenvolvimento no português europeu”* são da responsabilidade, respetivamente, das terapeutas da fala Raquel João Pinheiro Marques de Campos, cédula profissional n.º C-023116188, contacto 967176792, e Ana Rita Coimbra Pina de Campos, cédula profissional n.º C-023473185, contacto 967148043, alunas do segundo ano do curso de mestrado em Desenvolvimento e Perturbações da Linguagem na Criança, da Escola Superior de Saúde do Instituto Politécnico de Setúbal e da Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa, sob orientação da professora Doutora Maria Lobo e coorientação da Professora Eileen Sua Kay.

Os projetos supracitados têm como objetivos a verificação da utilidade dos parâmetros Diversidade Vocabular e Extensão Média de Enunciado, na avaliação do desenvolvimento da linguagem da criança e, conseqüentemente, na prática profissional dos terapeutas da fala. Pretende-se contribuir para a obtenção de valores referenciais dos índices D, Rácio type-token e Extensão Média de Enunciado, para o português europeu.

De modo a atingir os objetivos propostos, serão efetuadas gravações audiovisuais de amostras de discurso de crianças com idades compreendidas entre os três anos e 0 meses e os sete anos e onze meses, com desenvolvimento típico da linguagem. As gravações decorrerão em dois momentos que terão a duração aproximada de 30 minutos, cada.

Pretende-se que, em interação com apenas uma das investigadoras, cada criança construa uma narrativa, através da apresentação de imagens, converse sobre temas previamente definidos (e.g., férias de Natal, atividades de tempos livres,...) e participe num jogo conjunto, com recurso a brinquedos apropriados à sua idade.

Para além disso, após a recolha do discurso espontâneo, será solicitada a colaboração das crianças em algumas tarefas de avaliação formal da linguagem oral, no âmbito dos aspetos a analisar (semântica lexical e sintaxe), numa sessão de, aproximadamente, 30 minutos.

As produções das crianças serão posteriormente transcritas e analisadas, com recurso a software específico.

APÊNDICE C

Caracterização dos participantes na recolha de dados

Tabela 10 - Caracterização dos participantes na recolha de dados.

Participante	Data de Nascimento	Idade	Género	1. ^a Recolha	2. ^a Recolha	
1	AB	01/07/2008	[3A;11M]	F	08/06/2012	14/06/2012
2	EC	05/01/2009	[3A; 5M]	F	06/06/2012	12/06/2012
3	LA	10/10/2008	[3A;7M]	F	06/06/2012	13/06/2012
4	LD	15/09/2008	[3A;9M]	F	18/06/2012	18/06/2012
5	LC	26/08/2008	[3A;9M]	F	05/06/2012	11/06/2012
6	LD	18/07/2008	[3A;10M]	F	06/06/2012	12/06/2012
7	MM	15/08/2008	[3A;9M]	F	08/06/2012	14/06/2012
8	MA	01/11/2008	[3A;7M]	F	08/06/2012	14/06/2012
9	SA	06/10/2008	[3A;8M]	M	12/06/2012	15/06/2012
10	TS	19/07/2008	[3A;10M]	F	08/06/2012	13/06/2012
11	AC	25/05/2008	[4A;0M]	F	05/06/2012	11/06/2012
12	GP	20/03/2008	[4A;2M]	M	06/06/2012	13/06/2012
13	GL	22/06/2007	[4A;11M]	M	08/06/2012	14/06/2012
14	LD	22/09/2007	[4A;8M]	F	06/06/2012	13/06/2012
15	LD	21/01/2008	[4A;4M]	M	06/06/2012	12/06/2012
16	LR	23/10/2007	[4A;7M]	M	06/06/2012	12/06/2012
17	MC	30/08/2007	[4A;9M]	M	05/06/2012	11/06/2012
18	MN	06/02/2008	[4A;4M]	F	18/06/2012	18/06/2012
19	RF	04/12/2007	[4A;6M]	M	06/06/2012	13/06/2012
20	SC	07/12/2007	[4A;6M]	F	08/06/2012	14/06/2012
21	AC	25/08/2006	[5A;9M]	F	08/06/2012	14/06/2012
22	AL	09/08/2006	[5A;9M]	M	05/06/2012	11/06/2012
23	CN	18/11/2006	[5A;6M]	F	05/06/2012	12/06/2012
24	CP	05/07/2006	[5A;11M]	M	14/06/2012	14/06/2012
25	FM	13/03/2007	[5A;2M]	M	08/06/2012	12/06/2012
26	IC	14/09/2006	[5A;8M]	F	05/06/2012	12/06/2012
27	IR	21/03/2007	[5A;2M]	F	05/06/2012	12/06/2012
28	JC	01/11/2006	[5A;7M]	M	05/06/2012	11/06/2012
29	LS	27/02/2007	[5A;3M]	M	06/06/2012	13/06/2012
30	MB	26/08/2006	[5A;9M]	F	08/06/2012	14/06/2012
31	AA	20/01/2006	[6A;4M]	M	05/06/2012	11/06/2012
32	AA	03/03/2006	[6A;3M]	F	13/06/2012	18/06/2012
33	CA	24/04/2006	[6A;1M]	F	08/06/2012	14/06/2012
34	ES	17/06/2005	[6A;11M]	F	06/06/2012	08/06/2012
35	GS	06/04/2006	[6A;2M]	M	08/06/2012	13/06/2012
36	GC	31/03/2006	[6A;2M]	M	08/06/2012	12/06/2012
37	GG	21/06/2005	[6A;11M]	M	05/06/2012	11/06/2012
38	IM	19/07/2005	[6A;10M]	F	06/06/2012	13/06/2012
39	MC	19/10/2005	[6A;7M]	M	05/06/2012	11/06/2012
40	MA	16/10/2005	[6A;7M]	M	08/06/2012	12/06/2012
41	CF	12/07/2004	[7A;10M]	F	05/06/2012	11/06/2012
42	DC	27/09/2004	[7A;8M]	M	05/06/2012	11/06/2012
43	DS	19/09/2004	[7A;8M]	M	05/06/2012	11/06/2012
44	HH	06/04/2005	[7A;2M]	M	06/06/2012	08/06/2012
45	LC	24/05/2005	[7A;0M]	F	06/06/2012	13/06/2012
46	LM	05/01/2005	[7A;5M]	F	06/06/2012	12/06/2012
47	LV	28/04/2005	[7A;1M]	M	06/06/2012	12/06/2012
48	MF	06/08/2004	[7A;9M]	F	05/06/2012	11/06/2012
49	TP	14/03/2005	[7A;2M]	M	06/06/2012	13/06/2012
50	TC	19/04/2005	[7A;1M]	M	06/06/2012	13/06/2012

APÊNDICE D

Opções de codificação no programa CLAN

Tabela 11 - Opções de codificação no programa CLAN.

Estrutura do Ficheiro	Códigos utilizados	Descrição	Exemplos
A. Cabeçalho	@Begin	Primeira linha obrigatória pela qual começa o ficheiro.	<p>@Begin</p> <p>@Languages: por</p> <p>@Participants: CHI LD Target_Child, INV Rita Investigator</p> <p>@ID: por ritaraquel CHI 3;9.03 female Target_Child </p> <p>@ID: por ritaraquel INV Investigator </p> <p>@Birth of CHI: 15-SEP-2008</p> <p>@Date: 18-JUN-2012</p> <p>@Transcriber: Rita Campos</p> <p>@Location: Santa Comba Dão, Portugal</p> <p>@Situation: Narração de histórias a partir de sequências e de livros de imagens. Diálogo entre a investigadora e a criança no contexto dos tópicos abordados.</p> <p>@G: Tipo de discurso.</p> <p>@End</p>
	@Languages:	Língua(s) utilizada(s) representadas de acordo com as normas do ISO 639-3.	
	@Participants:	Linha obrigatória que lista os elementos participantes na interação transcrita. CHI e INV (identificação única de cada participante – criança e investigadora, respetivamente - no corpo da transcrição) seguido das iniciais do nome e apelido e papel desempenhado na interação - Target_Child (criança) e Investigator (investigadora.)	
	@ID:	Linha obrigatória atribuída a cada participante, na qual se efetua uma caracterização mais pormenorizada através do preenchimento dos campos considerados mais pertinentes.	
	@Birth of CHI:	Fornece informação específica da data de nascimento da criança, no formato dia-mês-ano	
	@Date:	Indica a data da interação, no formato dia-mês-ano.	
	@Transcriber:	Identifica a pessoa que transcreveu e codificou o ficheiro.	
	@Location:	Indica o local (cidade e país) em que se realizou a interação	
	@Situation:	Descreve a interação no que concerne às atividades, participantes e materiais utilizados, de forma a contextualizar o leitor.	
	@G:	Permite segmentar o corpo da transcrição de acordo com o que se pretende analisar: tipo de discurso.	
@End	Última linha obrigatória que finaliza o ficheiro.		

Estrutura do Ficheiro	Códigos utilizados	Descrição	Exemplos
B. Transcrição B.1. Linha ortográfica	CHI: INV:	Cada linha ortográfica corresponde a um enunciado e inicia-se com a identificação do falante, seguida de um espaço “Tab”.	*INV: está tudo muito longe.
	· ! ?	Sinais de pontuação utilizada para marcar o fim de uma linha ortográfica, de acordo com a entoação melódica.	*INV: pronto! *CHI: o que é isto?
	[/]	Repetição exata de palavra(s), sem correção.	*CHI: depois [/] depois o [/] o namorado tinha uma espada e partiu o vidro.
	[//]	Repetição parcial de palavra(s), com correção	*CHI: <isto é> [//] isto tem aqui letras?
	[///]	Reformulação total de palavra(s), sem ocorrer repetição.	*CHI: sim [///] não, os bombeiros.
	<>	Delimita uma expressão que vai ser repetida ou reformulada.	*CHI: <e o> [/] e o out(r)o cão levava <&u duas &pe> [//] uma p(r)enda.
	+...	Marcador final de enunciado incompleto, não interrompido.	*CHI: e a ta(r)taruga que tinha sempre o coelhinho às costas
	+,	Marcador inicial de enunciado retomado para completar um enunciado previamente interrompido.	+... *CHI: +, e a toupeira +... *CHI: +, estavam lá a ler umm jornal.
	+//.	Marcador final de enunciado incompleto resultante de uma auto-interrupção.	*CHI: eu queria ver uma &gru +//. [+ esc] *CHI: aqui dois homens iguais.
	+/.	Marcador final de enunciado incompleto, interrompido pelo interlocutor.	*CHI: eu não trouxe assim +/. [+ esc] *INV: espera, que eu não percebi. *INV: do Ben Ten? *CHI: sim, Alien Force.
	[-]	Inicia um novo enunciado, após um falso começo, com mudança de tópico e mantendo a mesma linha ortográfica.	*CHI: <a &ou> [-] mas essa já vimos.
	Xxx	Codifica partes do discurso ininteligíveis.	*CHI: aqui não vejo xxx.
www	Codifica produções consideradas supérfluas (e.g.: canto, leitura, repetições exatas de enunciados proferidos previamente pela investigadora, ...) para a transcrição, constituindo uma linha independente.	*INV: e a médica chamou. *CHI: www. *CHI: e depois ab(r)iu a boca pá [: para a] médica ver.	

Estrutura do Ficheiro	Códigos utilizados	Descrição	Exemplos
B. Transcrição B.1. Linha ortográfica	_	Substitui o hífen nas palavras compostas por justaposição.	*CHI: com umm saca_rolhas.
	0palavra omitida	Omissão de palavra marcada por 0, imediatamente seguido da palavra que o investigador considera estar em falta.	*CHI: e depois temos que 0jogar este chogo [: jogo]. [+ bch]
	[:]	Utilizado para substituir palavras resultantes de processos de coarticulação e produções consideradas desviantes do português-europeu padrão, sempre que se consegue atribuir um referente.	*CHI: foi dois camiões [: camiões]. [+ bch]
	()	Permite recuperar segmentos fónicos omitidos, resultantes de processos de coarticulação ou de produções desviantes, excluindo a utilização do código anterior.	*CHI: nós não (es)tavamos a jogar.
	&	Antecede, imediatamente, fragmentos fónicos resultantes de disfluências ou interrupções do discurso, interjeições e/ou onomatopeias que não têm funcionamento de palavra.	*CHI: &ca a mota &miaapu.
	@o	Codifica uma produção onomatopeica com funcionamento de palavra, localizando-se no fim da mesma.	*CHI: fui aos memés@o com o meu pai.
	umm	Substitui a palavra “um” que é excluída, por defeito, pelo comando do MLU.	*CHI: e eles acreditaram e depois começaram todos a correr de umm lado pro [: para o] outro.
	[+ esc]	Marcador final de enunciados incompletos, exceto os que são retomados posteriormente; de contagens de elementos; de enumerações, quando excedem um sintagma nominal.	*CHI: dó ré mi fá só lá si dó ré mi dó ré mi fá sol lá. [+ esc]
	[+ bch]	Marcador final de enunciados que contêm desvios sintáticos e morfofonológicos.	*CHI: e depois porem as malas no comboio. [+ bch]
	Sem codificação específica		O nome da criança transcreve-se utilizando as iniciais em maiúsculas, separadas por um espaço.
		As siglas são representadas apenas com a letra inicial maiúscula.	*CHI: &ah escolhia jogador de Nba .
		Os pronomes clíticos separam-se do verbo com um espaço, excluindo-se o hífen.	*CHI: acabou se a história!

Estrutura do Ficheiro		Códigos utilizados	Descrição	Exemplos
B. Transcrição	B.3.Linha Independente	%act	Constitui uma linha dependente para descrição da ação de um dos participantes	<p>%act: ruído vindo do exterior.</p> <p>@G: espontâneo.</p> <p>*INV: hum hum, muito bem.</p> <p>*INV: está?</p> <p>*INV: e tu tens medo de ir ao médico?</p> <p>*CHI: eu?</p> <p>*CHI: eu vou depois das férias porque tenho aqui borbulhas.</p>
	B.2.Linha Dependente	@G: narrativo. @G: espontâneo.	Marca o início de um novo contexto discursivo, espontâneo ou narrativo.	<p>*INV: oh então?</p> <p>*INV: como é que isso apareceu?</p> <p>*CHI: não sei.</p> <p>*INV: não?</p> <p>@G: narrativo.</p> <p>*CHI: então, era uma vez dois irmãos que queriam beber sumo de laranja e não tinham.</p> <p>*CHI: e o pai tinha muitas garrafas de sumo de laranja.</p>

APÊNDICE E

Base de dados do estudo

Tabela 12 - Base de dados do estudo.

Participante	Idade em meses	Gênero	Nível de Escolaridade	Média EME-p	Média EME-p-DN	Média EME-p-DE	Resultado do Teste de Linguagem
1	47	F	PE	3,856	4,486	3,528	50
2	41	F	PE	4,116	4,787	3,788	41
3	43	F	PE	3,406	3,335	3,370	41
4	45	F	PE	2,989	3,063	2,912	38
5	45	F	PE	4,16	4,781	4,434	44
6	46	F	PE	4,312	5,298	4,101	44
7	45	F	PE	4,619	5,480	3,832	48
8	43	F	PE	3,714	4,184	3,464	38
9	44	M	PE	3,500	3,712	3,399	35
10	46	F	PE	3,928	4,744	4,546	49
11	48	F	PE	4,306	6,019	3,469	46
12	50	M	PE	4,284	5,096	3,599	46
13	59	M	PE	4,892	5,867	4,627	52
14	56	F	PE	4,864	5,523	4,627	52
15	52	M	PE	3,449	4,030	3,174	41
16	55	M	PE	4,517	5,689	3,910	51
17	57	M	PE	4,426	6,106	3,737	47
18	52	F	PE	5,375	5,268	5,495	46
19	54	M	PE	4,174	5,142	3,760	45
20	54	F	PE	5,043	5,901	4,378	49
21	69	F	PE	5,534	6,423	4,833	51
22	70	M	PE	4,786	5,619	4,175	51
23	66	F	PE	5,271	6,213	4,907	52
24	71	M	PE	4,064	4,258	3,854	50
25	62	M	PE	5,093	6,328	4,647	50
26	68	F	PE	4,165	6,045	3,698	50
27	62	F	PE	5,031	6,126	4,132	52
28	67	M	PE	4,388	6,080	3,153	53
29	63	M	PE	4,888	5,826	4,240	50
30	69	F	PE	5,027	5,615	4,380	53
31	76	M	PE	5,297	7,431	4,930	32
32	75	F	PE	4,902	5,238	4,761	28
33	73	F	PE	5,333	6,190	4,837	25
34	83	F	E	4,678	5,598	4,280	34
35	74	M	PE	5,741	6,208	5,589	35
36	74	M	PE	5,718	6,814	5,379	23
37	83	M	E	5,459	6,146	5,293	38
38	82	F	E	5,360	5,847	5,103	36
39	79	M	E	5,756	6,455	5,440	44
40	79	M	E	4,594	4,587	4,600	40
41	82	F	E	7,150	10,289	5,809	36
42	92	M	E	5,915	7,138	5,778	37
43	92	M	E	5,192	7,216	4,259	43
44	86	M	E	4,913	6,128	4,066	33
45	84	F	E	4,991	5,509	4,119	43
46	89	F	E	5,831	7,619	4,803	42
47	85	M	E	5,143	7,058	4,170	33
48	93	F	E	5,804	7,884	4,262	40
49	86	M	E	5,872	6,388	5,470	36
50	85	M	E	5,116	6,095	4,573	44

ANEXO I

Materiais utilizados na elicitación de discurso

Brinquedos



Figura 2 - Tabuleiro de base para o *Kit* dos bombeiros, construtores e polícias



Figura 3 - Conjunto de brinquedos inerentes à temática das construções



Figura 4 - Conjunto de brinquedos inerentes à temática dos bombeiros



Figura 5 - Conjunto de brinquedos inerentes à temática dos polícias

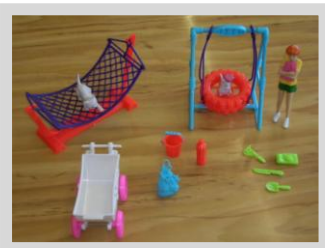


Figura 6 - Conjunto de brinquedos inerentes à temática do parque



Figura 7 - Conjunto de brinquedos inerentes à temática dos polícias



Figura 8 - Conjunto de brinquedos inerentes à temática da casa de



Figura 9 - Conjunto de brinquedos inerentes à temática dos médicos



Figura 10 - Conjunto de brinquedos inerentes à temática das ferramentas

Livros



Figura 11 – Livros “É Tudo Verdade” e “A Festa de Anos do Tomás”

(Most, 2011; Simas, 2003)

Sequências de imagens

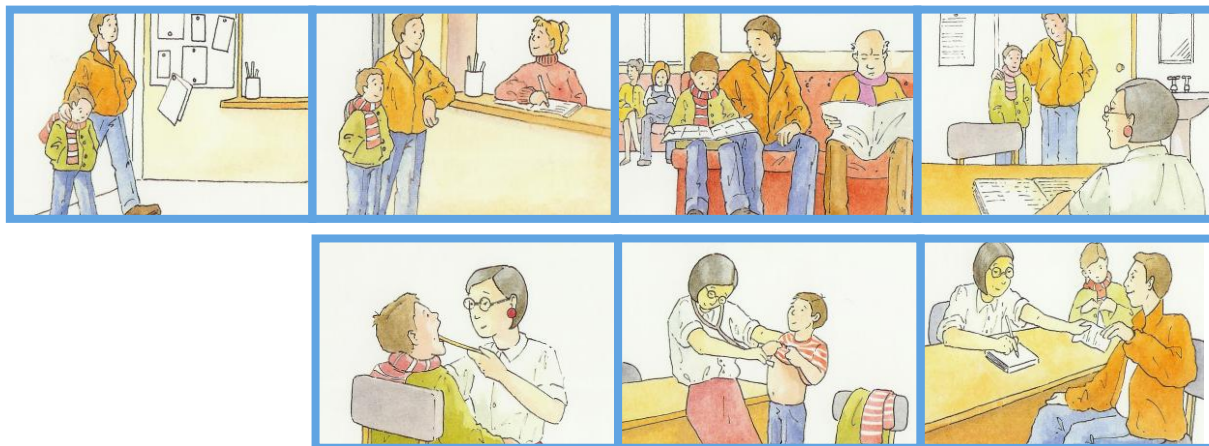


Figura 12 – Sequência de oito cartões (ida ao médico) (LDA)

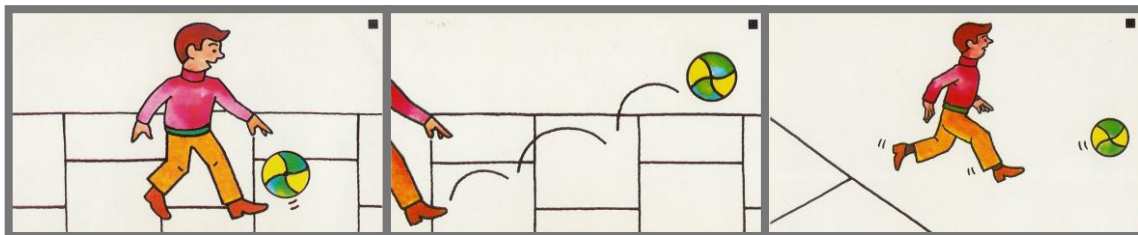


Figura 13 – Sequência de seis cartões (acidente na estrada). (LDA)



Figura 14 – Sequências de cinco cartões (na praia)
(LDA)



Figura 15 – Sequências de cinco cartões (na pesca)
(LDA)



Figura 16 – Sequências de 6 cartões (os biscoitos)
(Schubi)

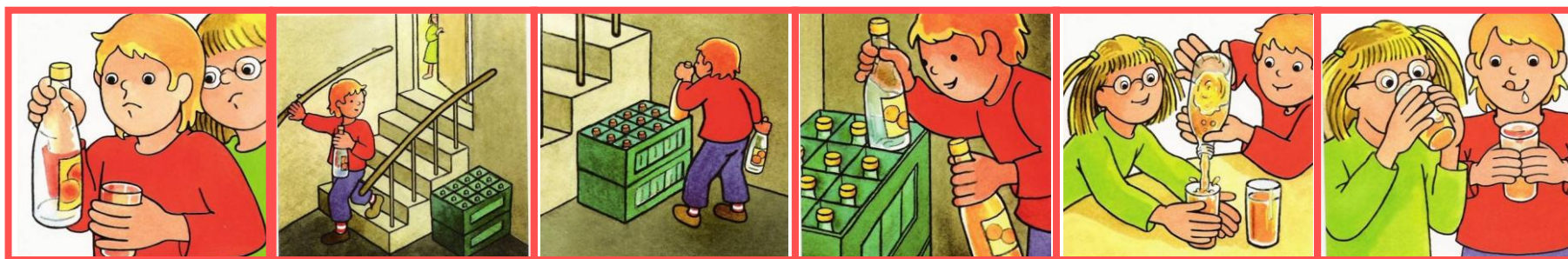


Figura 17 – Sequências de 6 cartões (acabou o sumo)
(Schubi)

ANEXO II

Resultado de teste estatístico

Tabela 13 - Resultado das médias das ordens da EME-p, de acordo com a faixa etária.

Faixa etária	EME-p	EME-p-DN	EME-p-DE
[3;00 – 3;11]	7,90	8,10	11,10
[4;00 – 4;11]	19,60	19,90	20,40
[5;00 – 5;11]	25,00	28,30	23,90
[6;00 – 6;11]	35,30	30,70	39,70
[7;00 – 7;11]	39,70	40,50	32,40