



Raquel João  
Pinheiro Marques de  
Campos

**Rácio Type-Token e D como  
indicadores de desenvolvimento  
linguístico no português  
europeu**

Dissertação de Mestrado em Desenvolvimento e  
Perturbações da Linguagem na Criança  
Área de Especialização em Terapia da Fala e  
Perturbações da Linguagem

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 científica da Doutora 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 ti, pequena Leonor,  
por me presentares com as tuas descobertas  
e colorires a minha vida com os teus sorrisos ternurentos.*

## AGRADECIMENTOS

Agora que se aproxima a conclusão de mais uma importante etapa do meu percurso acadêmico, impõe-se um agradecimento sentido e profundo a todos aqueles que contribuíram decisivamente para a concretização desta dissertação.

Expresso a minha imensa gratidão à Professora Doutora Maria Lobo, pela forma sublime como orientou este estudo, com um interesse e dedicação constantes. Agradeço o sentido de responsabilidade que me foi inculcando em todas as fases de trabalho e, acima de tudo, a partilha de conhecimentos que despertou o meu desejo de procurar saber mais, para fazer melhor.

Um agradecimento especial à coorientadora, Doutora Eileen Sua Kay, pelas valiosas contribuições e, fundamentalmente, por ter acreditado neste projeto.

Uma palavra de apreço ao Professor Doutor Hélder Alves, pela preciosa ajuda na análise estatística, pela atenção, paciência e encorajamento.

À Professora Doutora Ana Castro, coordenadora do curso de mestrado em Desenvolvimento e Perturbações da Linguagem na Criança, agradeço ter acompanhado este projeto.

Apresento, também, os meus agradecimentos à Diretora do Agrupamento de Escolas de Santa Comba Dão, por ter autorizado a realização das recolhas e aos professores e educadores por terem acolhido este estudo, bem como pela colaboração na seleção da amostra, aos encarregados de educação por terem autorizado a participação dos seus educandos neste projeto e, principalmente, a todas as crianças que participaram nesta investigação e prestaram uma contribuição fundamental, para que este estudo fosse possível.

Às crianças que acompanho e aos respetivos pais, em especial à Bárbara e à Sandra, o meu reconhecimento pelos momentos de partilha e de crescimento pessoal e profissional. Por acreditarem nas minhas competências e fomentarem, em mim, a busca do saber.

À minha amiga e colega de investigação, Rita Campos, por ter estado sempre presente neste percurso. Obrigada pelas gargalhadas, pelo companheirismo nos momentos de desespero, por nunca me deixar desistir. Foi e será sempre um privilégio trabalhar contigo.

A todos os meus amigos, pelo apoio e compreensão nos momentos em que não pude estar presente, à Catarina Santos pela amizade e afeto, por me fazer encarar este trabalho com motivação, desde o início.

Ao João, um especial agradecimento pelo amor e confiança em todos os momentos e por respeitar e compreender a minha ausência.

Finalmente, à minha família, particularmente aos meus Pais, Irmã e Cunhado, o meu eterno agradecimento pelo amor e apoio incondicional, pelas palavras de transmissão de força e por me terem proporcionado as condições necessárias à realização desta dissertação.

## RESUMO

### RÁCIO TYPE-TOKEN E D COMO INDICADORES DE DESENVOLVIMENTO LINGUÍSTICO NO PORTUGUÊS EUROPEU

RAQUEL JOÃO PINHEIRO MARQUES DE CAMPOS

PALAVRAS-CHAVE: Desenvolvimento Linguístico; Diversidade Vocabular; Rácio Type-Token; Índice D; CLAN; CHAT; Escolaridade; Contexto narrativo; Contexto Espontâneo.

A Diversidade Vocabular (DV) relaciona a variedade de palavras que são utilizadas por um indivíduo num texto, seja ele oral ou escrito, refletindo assim a complexidade do seu conhecimento vocabular. Apesar de a importância da DV ser reconhecida, não existe ainda uma unanimidade quanto à medida que permite a obtenção de resultados mais válidos. O Rácio Type-Token (RTT) e o D são os índices mais conhecidos e os mais aplicados em diversas áreas da investigação linguística, nomeadamente na aquisição e desenvolvimento da linguagem oral, aprendizagem de segunda língua, perturbações específicas da linguagem, afasias, demências e alterações comportamentais. Não são, contudo, conhecidos estudos desenvolvidos para o português europeu, pelo que se desenvolveu este trabalho que teve como principais objetivos descrever a DV, medida pelo RTT e D, em crianças com idades compreendidas entre os 3 e os 7 anos de idade, bem como verificar a relação destes índices com o desenvolvimento linguístico.

Neste estudo foram gravadas amostras de discurso espontâneo e narrativo de 50 crianças com desenvolvimento típico de linguagem, monolíngues, falantes nativas do português europeu e a frequentar o ensino pré-escolar ou primeiro ciclo, recolhidas em dois períodos de 30 minutos que contiveram, cada um, momentos de conversação, jogo simbólico, descrição de imagens e livros de histórias. Todas as amostras de discurso foram transcritas no programa *Computerized Language Analysis* (CLAN), em formato CHAT, e a DV, medida pelos índices RTT e D, foi calculada no mesmo programa através dos comandos *freq* e *vocd*, para cada faixa etária e contexto de discurso, tendo sido depois realizada a análise estatística dos dados.

Concluiu-se que i) o RTT não é capaz de refletir a evolução da DV, entre os 3 e 7 anos de idade e o D apenas reflete diferenças entre os 3 e os 6 anos e entre os 3 e 7 anos; ii) as crianças que frequentam o ensino escolar apresentam uma DV superior à das crianças que frequentam o pré-escolar; iii) apenas o índice D é indicador de desenvolvimento linguístico.

## ABSTRACT

### TYPE-TOKEN RATIO AND D AS INDICATORS OF LINGUISTIC DEVELOPMENT IN EUROPEAN PORTUGUESE

RAQUEL JOÃO PINHEIRO MARQUES DE CAMPOS

KEYWORDS: Language Development; Vocabulary Diversity; Type-Token Ratio; Index D; CLAN; CHAT; Education; Narrative Context; Spontaneous Context.

The variety of words used by a person in both oral and written text represents the Vocabulary Diversity (VD). Therefore, the complexity of person's vocabulary knowledge is reflected by VD. Although the importance of VD has been recognized, there is still no unanimity regarding the most suitable parameter to obtain more valid results. Nowadays, indexes such as Type-Token Ratio (RTT) and D are the best-known and the most widely used in several areas of linguistic research, namely in acquisition and development of oral language, second language learning, specific language disorders, aphasia, dementia and behavioural disorders. However, for European Portuguese, there were no studies available. Thus, the purpose of this work was to describe the VD measured by RTT and D indexes in children aged between 3 to 7 years old and to relate the obtained indexes with language development measures.

For this study, we used spontaneous and narrative discourse of 50 monolingual children, native speakers of European Portuguese, with typical language development, attending pre-school or first cycle. Firstly, the speech samples were recorded in two 30-minute periods, both containing moments of conversation, symbolic play, image description and book storytelling. Then, we transcribed the speech samples following Computerized Language Analysis program (CLAN), in CHAT format. For each age group and speech context, we calculated VD (measured by RTT and D indexes) using *freq* and *vocd* commands of CLAN. In a final stage, we conducted the statistical analysis of the obtained data.

Finally, it was concluded that i) RTT is unable to describe the VD evolution in children aged between 3 to 7 years old and D index only describes the differences between 3 and 6 to 7 years; ii) children attending school education have a higher VD when compared with children attending pre-school; and iii) only D index is an indicator of language development.

# ÍNDICE

|  |    |
|--|----|
| INTRODUÇÃO .....   | 1  |
| CAPÍTULO I: ENQUADRAMENTO TEÓRICO .....  | 3  |
| I. 1. Aquisição da linguagem oral: da formação de conceitos às primeiras produções com significado ..... | 3  |
| I. 1. 1. Operações mentais subjacentes à aquisição e produção de palavras .....                          | 5  |
| I. 1. 2. Etapas do desenvolvimento vocabular.....  | 10 |
| I. 1. 3. A relação entre desenvolvimento lexical e domínio vocabular.....                                | 13 |
| I. 1. 3. 1. A Diversidade Vocabular como preditora de sucesso escolar .....                              | 15 |
| I. 1. 3. 2. Fatores que influenciam a expansão vocabular .....   | 16 |
| I. 1. 4. Diversidade Vocabular: quantificação e avaliação .....  | 18 |
| I. 1. 4. 1. Item lexical: o problema da definição de palavra.....  | 18 |
| I. 1. 4. 2. A importância da definição de critérios no cálculo da Diversidade Vocabular .....            | 19 |
| I. 1. 4. 3. Cálculo da Diversidade Vocabular: medidas tradicionais.....                                  | 22 |
| I. 1. 4. 4. Rácio Type-Token .....   | 24 |
| I. 1. 4. 5. D .....  | 28 |
| I. 1. 4. 6. Procedimentos de avaliação.....  | 31 |
| CAPÍTULO II: METODOLOGIA .....   | 35 |
| II. 1. Questões de investigação e tipo de estudo.....  | 35 |
| II. 2. Hipóteses e Variáveis.....  | 36 |
| II. 3. Caracterização da amostra .....   | 37 |
| II. 4. Material e Procedimentos da Recolha de Dados.....   | 38 |
| II. 5. Tratamento e Análise dos dados .....  | 44 |
| CAPÍTULO III: APRESENTAÇÃO DOS RESULTADOS .....  | 52 |
| III. 1. Descrição dos resultados .....   | 52 |

|  |     |
|--|-----|
| III. 2. Valores médios e desvios-padrão de RTT e D, segundo faixa etária.....  | 53  |
| III. 3. Comparação dos desempenhos no D e RTT, segundo as faixas etárias .....   | 54  |
| III. 4. Comparação dos desempenhos no D e RTT, segundo os níveis de escolaridade .   | 55  |
| III. 5. Comparação dos desempenhos no D e RTT, segundo os contextos discursivos..  | 56  |
| III. 6. Correlação entre RTT, D, idade e desenvolvimento linguístico .....   | 58  |
| CAPÍTULO IV: DISCUSSÃO DOS RESULTADOS.....   | 60  |
| IV. 1. Discussão dos Resultados.....   | 60  |
| IV. 2. Limitações do Estudo .....  | 67  |
| CONCLUSÃO .....  | 68  |
| BIBLIOGRAFIA .....   | 70  |
| Lista de Figuras .....   | 78  |
| Lista de Gráficos .....  | 79  |
| Lista de Tabelas.....  | 80  |
| Apêndice A: Autorização da Diretora do Agrupamento de Escolas de Santa Comba Dão.....  | i   |
| Apêndice B: Explicação do Projeto de Mestrado entregue aos Encarregados de Educação<br>.....   | iii |
| Apêndice C: Declaração de Consentimento Informado entregue aos Encarregados de<br>Educação .....   | v   |
| Apêndice D: Brinquedos utilizados na recolha de amostras de discurso espontâneo.....   | vii |
| Apêndice E: Sequências de imagens utilizadas na recolha de amostras de discurso narrativo,<br>da coleção “ <i>LDA Language Cards – Sequential Thinking</i> ” ..... | xi  |
| Apêndice F: Sequências de imagens utilizadas na recolha de amostras de discurso narrativo,<br>da coleção “Obrigado, Kim!” .....                                    | xvi |
| Apêndice G: Formato do ficheiro CHAT .....   | xix |
| Apêndice H: Critérios de Transcrição do Programa CLAN .....  | xxi |

## **LISTA DE ABREVIATURAS**

A - Anos

ANOVA – Análise de Variância

CHILDES - Child Language Data Exchange System

CLAN - Computerized Language Analysis

DIVA - Directions Into Velocities of Articulators

*DP* - Desvio-padrão

DV - Diversidade Vocabular

EME - Extensão Média de Enunciado

*F* - ANOVA

GOLE - Grelha de Observação da Linguagem – Nível Escolar

H - Hipótese

*H* - Teste não Kruskal-Wallis

*M* - Média

M – Meses

MCDI - MacArthur-Bates Communicative Development Inventory

MTLD - Measure of Textual Lexical Diversity

n- número de crianças

NDW - Number of Different Words

*p* – Nível de Significância

PEDL - Perturbações Específicas do Desenvolvimento da Linguagem

*r* – Coeficiente de Correlação de Pearson

RTT - Rácio Type-Token

SALT - Systematic Analysis of Language Transcripts

SPSS - Statistical Package for the Social Sciences

TALC - Teste de Avaliação da Linguagem na Criança

TNW – Total Number of Words

U – Teste de Mann-Whitney

## INTRODUÇÃO

A produção da primeira palavra constitui um marco importante no desenvolvimento linguístico de uma criança. As crianças produzem as suas primeiras palavras entre os 10 e 13 meses (Puyuelo & Rondal, 2007), no entanto, este processo tem o seu início aproximadamente aos 6 meses, momento em que é despoletada a sensibilidade para a variedade de propriedades fonológicas das palavras da língua materna.

É através do uso das palavras que a criança começa a representar a realidade. Segundo Sim-Sim (1998), as primeiras produções lexicais das crianças apresentam muitas semelhanças, uma vez que começam por representar pessoas e objetos com importância para a própria criança. A aquisição lexical do ponto de vista da produção é lenta, após o surgimento da primeira palavra, aumentando consideravelmente a partir dos 2 anos e 6 meses (Costa & Santos, 2003). Os primeiros anos de vida são cruciais no que concerne ao desenvolvimento vocabular, afetando o futuro escolar, social e profissional do indivíduo (Pikulski & Templeton, 2004). Considerando a Diversidade Vocabular (DV) como “um fator determinante para a qualidade da escrita”, Duarte, Colaço, Freitas, e Gonçalves (2011) referem que esta influencia indiretamente o sucesso escolar, social e profissional, uma vez que a escrita assume um papel extremamente importante na vida escolar, enquanto meio preferencial de avaliação, e, na vida social e profissional, enquanto meio de comunicação (Mata, 2003). A DV revela-se igualmente importante na qualidade das mensagens orais, nomeadamente na partilha de ideias, experiências e conhecimentos (Duarte *et al.*, 2011). De acordo com Malvern, Richards, Chipere, e Durán (2004), a DV é um bom indicador da interação existente entre a dimensão do vocabulário de um indivíduo e o uso que faz do mesmo. A DV corresponde à variedade de itens utilizados pelos falantes numa conversação e pode ser calculada através de diversos índices, tais como o RTT e D. A necessidade de obter uma medida que permita determinar, com rigor, a DV conduziu à realização de diversos estudos, não se verificando ainda consenso relativamente ao processo que confere resultados mais credíveis.

Deste modo, e atendendo à importância deste parâmetro linguístico, a presente dissertação tem como objetivo principal a identificação de uma medida capaz de representar este parâmetro linguístico e de se relacionar com o desenvolvimento linguístico. Pretende-se também verificar a utilidade destes índices na prática profissional de um Terapeuta da Fala, já que só conhecendo o desenvolvimento normal é que se pode detetar,

compreender e atuar sobre a patologia, proporcionando, se possível, valores de referência para o português europeu, para as faixas etárias compreendidas entre os 3 e os 7 anos. Para além destes objetivos, procurar-se-á analisar se o desempenho das crianças, a este nível, é influenciado pelo contexto em que ocorre (narrativo e/ou espontâneo), bem como verificar se o processo de aprendizagem escolar obrigatória exerce alterações na DV.

No seguimento dos objetivos propostos, este trabalho foi organizado da forma que aqui se descreve. Após este ponto introdutório, efetua-se um enquadramento teórico da DV, onde constam os conceitos relacionados com o desenvolvimento linguístico e com o cálculo da DV (Capítulo I). Posteriormente, apresenta-se a metodologia utilizada neste estudo, focando as questões de investigação e o tipo de estudo realizado, as variáveis e as hipóteses formuladas (Capítulo II). É também apresentada e caracterizada a amostra do estudo e são descritos todos os materiais e procedimentos adotados na recolha, tratamento e análise de dados. Seguidamente apresentam-se os resultados decorrentes da análise estatística (Capítulo III), efetua-se a devida discussão e referem-se as limitações associadas ao estudo (Capítulo IV). Finda-se com a apresentação das conclusões e reflexão sobre a importância do estudo.

## **CAPÍTULO I: ENQUADRAMENTO TEÓRICO**

No presente capítulo serão abordados aspetos relacionados com o desenvolvimento linguístico, mais concretamente com o processo de produção de palavras. Será também apresentado o conceito de DV, assim como as respetivas formas de avaliação e quantificação, através de uma descrição detalhada das medidas consideradas importantes para o presente estudo.

### **I. 1. Aquisição da linguagem oral: da formação de conceitos às primeiras produções com significado**

A linguagem representa o nível mais alto na comunicação e faz parte do que define os indivíduos como seres humanos, sendo por isso uma propriedade exclusiva do Homem. De acordo com a teoria de Chomsky, abordada por Costa e Santos (2003) e Guasti (2002), o ser humano já nasce com uma capacidade inata para adquirir linguagem que é desenvolvida em interação com o meio envolvente, ou seja, através do contacto com outros falantes. É através da linguagem que o ser humano concretiza o seu pensamento sobre tudo o que o rodeia e, assim, se mantém conectado com os seus pares. A forma mais utilizada na comunicação entre indivíduos é, sem dúvida, a oral. A aquisição da linguagem oral constitui um dos marcos mais significativos da primeira infância, provavelmente até, de toda a vida de um indivíduo e, em condições consideradas normais, tem o seu início nos primeiros meses de vida e desenvolve-se até ao final da vida (Costa & Santos, 2003). Pressupõe a aquisição de um sistema, a língua, que, apesar de apresentar características universais, se torna específico dentro de cada comunidade linguística.

A produção de uma palavra não é mais do que a expressão de um conceito, ou seja, a produção de uma sequência fónica à qual é atribuído um significado preciso. A capacidade de produzir palavras implica necessariamente um conhecimento prévio dessas mesmas palavras e, por conseguinte, o estabelecimento de uma associação correta ao conceito ao qual está ligado, tornando-se para isso essencial a estimulação linguística.

O processo de produção das primeiras palavras tem o seu início aos 6 meses (aproximadamente), momento em que é despoletada a sensibilidade para a variedade de propriedades fonológicas das palavras da língua materna. As crianças apresentam um conhecimento bastante sofisticado da língua materna e uma capacidade inata de perceber os padrões mais frequentes, que funcionam, neste caso, como pré-requisito no

processo de aquisição lexical. O processo de aquisição lexical é rápido e inicia-se com o “*mapping*” que consiste no reconhecimento de uma palavra nova enquanto palavra, seguido da entrada no léxico (Carey, 1978) e, de acordo com Guasti (2002), envolve duas tarefas, que devem ser realizadas separadamente pela criança: (1) segmentação do discurso em palavras e (2) associação de significado às palavras. Uma das hipóteses explicativas desta segunda tarefa defende que a criança aprende o significado de uma nova palavra ao identificar os elementos comuns às diferentes situações em que esta ocorre. Porém, esta hipótese apresenta limitações, na medida em que, de uma determinada situação podem ser extraídas múltiplas interpretações, e, para além disso, não explica a aprendizagem de palavras cujo referente não é diretamente observável, tal como nomes abstratos e verbos.

Guasti (2002) refere, na sua obra, que a aprendizagem de um conceito depende de um conjunto de pressupostos: a assunção do objeto completo, a exclusividade mútua e a assunção taxonómica. De acordo com a assunção do objeto completo, perante uma palavra nova e um dado objeto, a criança assume que o seu referente é o objeto no seu todo e não as partes que o constituem. Contudo, pelo pressuposto da exclusividade mútua, quando é apresentado mais do que um objeto, dos quais só um é que é desconhecido, a palavra nova referir-se-á a esse objeto, da mesma forma que, pelo mesmo pressuposto, perante duas palavras, sendo apenas uma delas desconhecida, a criança supõe que a palavra nova se refere a uma parte desse mesmo objeto. Quando já conhece várias palavras, a criança é capaz de estabelecer as conexões necessárias que permitam reconhecer outras palavras que se relacionem com a nova, construindo assim uma rede e, por conseguinte, o seu significado (Carey, 1978). Pela assunção taxonómica, a aprendizagem do significado da nova palavra resulta de uma relação de categoria entre as palavras, porque os objetos que designam têm as mesmas características e pertencem ao mesmo grupo taxonómico. Sim-Sim (1998) refere que a inclusão de um elemento numa determinada categoria, designada por categorização, implica uma seleção apropriada das características ou atributos relevantes e torna-se facilitada quando esses atributos são salientes. A saliência dos atributos determina, assim, a facilidade da categorização e, por conseguinte, a rapidez de aquisição do conceito.

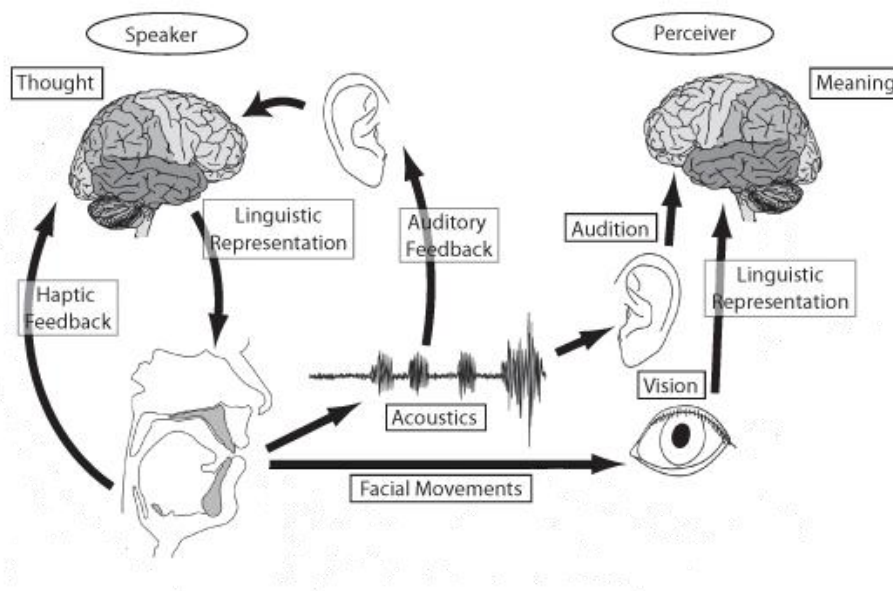
A utilização de referenciais explica o facto observado na maioria dos estudos realizados, que serão aqui descritos adiante, no âmbito das primeiras produções da criança, com significado. Verifica-se que, na generalidade, as primeiras palavras da criança são nomes, utilizados para representar as pessoas e os objetos que a rodeiam (Guasti, 2002; Sim-Sim, 1998). Segundo Sim-Sim (1998), as primeiras produções lexicais das crianças

apresentam muitas semelhanças, uma vez que começam por representar membros da família, alimentos preferidos, brinquedos e animais. Só posteriormente começam a utilizar, para além de nomes, verbos.

O momento preciso em que a criança produz a primeira palavra pode ser difícil de determinar, na medida em que surge durante a fase do balbúcio, caracterizado pelas produções repetidas de sílabas constituídas pelo padrão consoante/vogal, ainda sem qualquer significação. A repetição de monossílabos constitui uma etapa importante, uma vez que a sua constituição se aproxima daquilo que serão as primeiras palavras da criança. A primeira produção com significado distingue-se das restantes a partir do momento em que ocorre espontaneamente, sempre no mesmo contexto e na presença do mesmo referente. Apesar de a sua forma ainda se distinguir da forma adulta, os sons utilizados são semelhantes aos sons constituintes da palavra que a criança pretende produzir e denotam uma simplicidade fonética e fonológica (Costa & Santos, 2003; Sim-Sim, 1998).

### I. 1. 1. Operações mentais subjacentes à aquisição e produção de palavras

O processo de produção e compreensão de discurso é bastante complexo e envolve vários níveis da atividade humana, representados por uma sucessão de acontecimentos entre o falante e o ouvinte, tal como se pode observar na Figura 1.



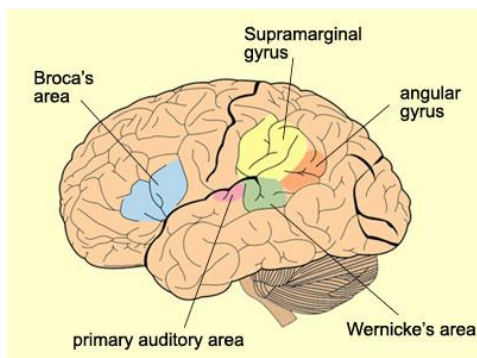
**Figura 1** – Representação dos mecanismos envolvidos na produção e percepção de fala (Yeung, 2013).

Quando um falante tem a intenção de produzir uma mensagem que alcance o interlocutor, necessita primeiro de organizar os seus pensamentos, decidir o que pretende transmitir e posteriormente selecionar e ordenar corretamente as palavras de acordo com as regras gramaticais da língua (nível linguístico), de modo a produzir um enunciado com significado. Esta mensagem, que tem a sua origem no cérebro, é conduzida através de impulsos nervosos para os músculos dos órgãos fonatórios e articulatórios (nível fisiológico) que, por sua vez, dão corpo à mensagem sob a forma de ondas sonoras (nível acústico). A mensagem chega ao ouvinte e ao falante através dessas ondas sonoras que, ao ativarem o sistema auditivo, provocam a produção de impulsos nervosos conduzidos até ao cérebro, através do nervo auditivo (nível fisiológico). Este processo ocorrido no falante é designado por *feedback* auditivo e possibilita-lhe uma monitorização da qualidade e inteligibilidade das suas produções. Para além da via auditiva, a mensagem chega simultaneamente ao ouvinte pela via visual. A capacidade de efetuar uma leitura dos movimentos labiais do falante facilita a perceção da mensagem, principalmente em ambientes ruidosos. É, então, no cérebro que ocorre o reconhecimento da mensagem do falante e lhe é atribuído um significado (nível linguístico) (Denes & Pinson, 1993).

Diversos têm sido os estudos realizados no âmbito da neuroanatomia, com o objetivo de se particularizar a atividade cerebral envolvida na compreensão e produção da linguagem, todavia não existe ainda uma posição consensual relativamente à sua localização e ao modo como se processa.

Se, por um lado, há autores que continuam a atribuir a função do processamento da linguagem ao hemisfério esquerdo do cérebro, outros há, como Hickok e Poeppel (2000, 2004, 2007), que propuseram recentemente um modelo explicativo que reconhece a importância e intervenção de ambos os hemisférios cerebrais nesse processamento.

A teoria que tem prevalecido localiza a linguagem numa região do córtex perissilviano que é constituída pela área de Wernicke, área de Broca, gyrus supramarginal e gyrus angular (Figura 2) (Mazoyer *et al.*, 1993; Webb & Adler, 2008).

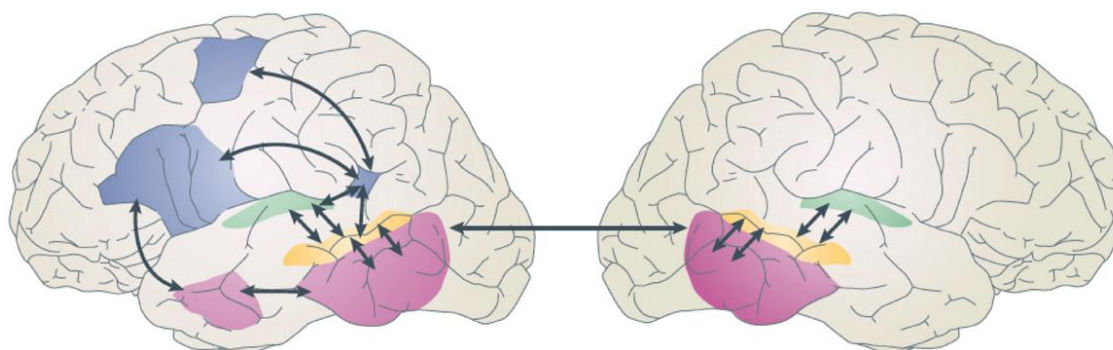


**Figura 2** – Esquema do hemisfério esquerdo do cérebro humano com representação da área de Wernicke, área de Broca, giro supramarginal e giro angular (*In*: [http://thebrain.mcgill.ca/flash/a/a\\_10/a\\_10\\_cr/a\\_10\\_cr\\_lan/a\\_10\\_cr\\_lan.html](http://thebrain.mcgill.ca/flash/a/a_10/a_10_cr/a_10_cr_lan/a_10_cr_lan.html)).

A área de Broca, localizada no lobo frontal, é responsável pela programação motora dos movimentos necessários à articulação verbal. A parte anterior e ventral desta área estão envolvidas no processamento semântico, enquanto na parte posterior ocorre o processamento sintático e fonológico, assim como o controlo motor da fala. À área de Wernicke, localizada no lobo temporal, está associada a compreensão da linguagem oral e a formulação de conceitos. Os conceitos formados nesta área são transmitidos à área de Broca, para que depois ocorra a programação motora e a expressão da linguagem (Webb & Adler, 2008).

Estudos linguísticos contemporâneos demonstraram que o processo de mapeamento das palavras ao respetivo significado pode ser explicado pelo modelo da dupla via. De acordo com este modelo, representado anatomicamente na Figura 3, na presença de um estímulo de fala, o processamento do material linguístico inicia-se com a ativação bilateral da região dorsal do giro temporal superior e do sulco temporal superior, divergindo depois em duas vias, uma ventral e outra dorsal. A via ventral está organizada bilateralmente, verificando-se a existência de duas correntes paralelas em ambos os hemisférios; envolve as estruturas localizadas nas porções superior e média do lobo temporal. A esta corrente está afeto o reconhecimento da fala. É, por conseguinte, responsável pela transformação do sinal acústico em representações que possam ser mapeadas com o léxico mental. À via dorsal, que envolve algumas estruturas localizadas essencialmente no hemisfério esquerdo (região posterior do lobo frontal, região posterior do lobo temporal e opérculo parietal), está atribuída a função de transformar os sinais acústicos em representações articulatórias que possam ser aí conservadas de forma a

poderem ser recrutadas em operações específicas (Hickok, 2012; Hickok & Poeppel, 2007; Jr., 2012).

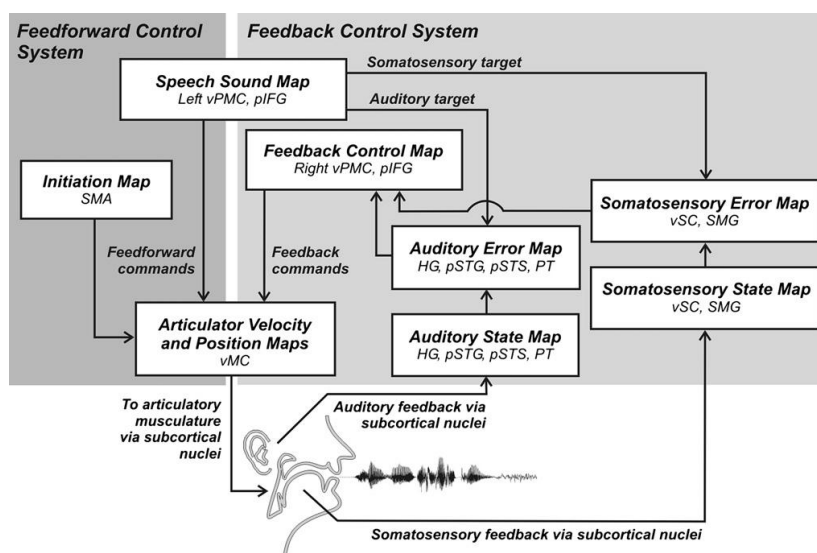


**Figura 3** – Representação anatômica do modelo da dupla via. Áreas marcadas a verde e amarelo representam, respetivamente, o giro temporal superior e o sulco temporal superior. As áreas identificadas bilateralmente a cor-de-rosa representam a via ventral; a área azul que surge apenas no hemisfério esquerdo corresponde à via dorsal (adaptado de Hickok (2012)).

O conhecimento concetual também designado por memória semântica diz respeito ao conhecimento que cada indivíduo tem dos objetos, acontecimentos, pessoas e palavras sem que, para isso, seja necessária uma ligação a um local ou momento; é partilhado pelos membros de uma dada cultura, contudo depende sempre da experiência individual. Correntes teóricas mais recentes defendem que a construção do conhecimento concetual depende de fatores como a percepção e ação, pelo que a sua representação no cérebro estará associada a uma rede onde, para além da região responsável pela linguagem, constam também as regiões associadas aos sistemas sensorial e motor (Patterson, Nestor, & Rogers, 2007; Webb & Adler, 2008). Deste modo, cada conceito corresponde a uma entrada distinta no cérebro que contém informação relativamente ao significado e à forma como se relaciona com outros conceitos. É cerebralmente representado por redes, distribuídas por diferentes áreas, conforme as propriedades semânticas que lhe são atribuídas. Assim, se a um determinado conceito está associada uma forte componente visual, a rede concentrar-se-á principalmente no córtex visual, da mesma forma que um conceito associado a movimento ativar-se-á principalmente neurónios nas áreas motoras (Setola & Reilly, 2005).

Na produção de fala estão igualmente envolvidos mecanismos de interação entre os sistemas acústico, somatossensorial e motor que envolvem estruturas corticais e subcorticais. Deste modo, para que unidades linguísticas sejam veiculadas de um falante

para um ouvinte, é necessário que se produzam sinais acústicos moldados pela forma do trato vocal, através da informação fornecida pelos recetores proprioceptivos e táteis e determinados pelos movimentos dos articuladores resultantes dos comandos motores (Guenther, 1994). O modelo DIVA (*Directions Into Velocities of Articulators*), proposto por (Guenther, 1994; Guenther & Vladusich, 2012), procura explicar estes mecanismos e encontra-se representado na Figura 4.



**Figura 4** – Representação esquemática do modelo DIVA<sup>1</sup> (Guenther, 1994; Guenther & Vladusich, 2012).

De acordo com este modelo, as primeiras produções de fala resultam de mapeamentos ocorridos num estágio precoce do desenvolvimento da criança, em que os sons da fala estão ainda ausentes, pelo que se considera que decorrem do processo de integração de relações sensoriomotoras. Assim, os sons da fala que foram previamente percebidos encontram-se representados nos neurónios que serão posteriormente ativados no momento da produção (“*speech sound map*”). A produção dos primeiros segmentos depende primariamente dos mecanismos de *feedback*, uma vez que ainda não

<sup>1</sup> Este modelo esclarece que a produção de sons de fala inicia com a ativação dos neurónios a que estão associados (“*speech sound map*”). Esta área localiza-se na área de Broca e na porção ventral do córtex pré-motor. A ativação desta área gera comandos motores que são conduzidos através de dois subsistemas de controlo. Um subsistema de *feedforward* em que ocorre uma projeção do comando diretamente para o cerebelo e córtex motor primário, responsáveis pela coordenação do movimento e velocidade e posicionamento dos articuladores, respetivamente, e um sistema de *feedback* auditivo e somatossensorial que envolve projeções para as áreas sensoriais do córtex, onde são integradas informações sobre os parâmetros acústicos de cada som. Os controlos de *feedback* permitem melhorar os mecanismos automáticos de *feedforward*, através da deteção de erros (“*auditory state map*” e “*somatosensory state map*”) e respetiva correção (“*auditory error map*” e “*somatosensory error map*”).

ocorreu o mapeamento dos movimentos necessários à realização dos mesmos. À medida que pratica o balbucio, a criança vai recebendo informações motoras, auditivas e somatossensoriais que vão sendo utilizadas para regular o sistema de *feedforward*, automatizando assim os movimentos.

A continuidade deste processo, que se vai aprimorando através de fenômenos de *feedback*, faz com que, numa fase posterior, os sons da fala sejam produzidos em maior quantidade, resultando também daí um aumento da complexidade dos mesmos, em termos articulatórios.

### **I. 1. 2. Etapas do desenvolvimento vocabular**

Tal como previamente referido, a tarefa de identificação da primeira palavra produzida pela criança não é simples de sinalizar cronologicamente. Não existindo um marco definido para este acontecimento, sabe-se que as produções das primeiras palavras com significado surgem aproximadamente entre os 9 e os 13 meses, existindo contudo diferenças individuais entre crianças (Albuquerque, 2000; Costa & Santos, 2003; Rescorla, Mirak, & Singh, 2000). Os dados existentes neste âmbito resultam da análise de inventários realizados pelos pais ao vocabulário das crianças. São conhecidos alguns estudos realizados nas mais diversas línguas, nomeadamente inglês, espanhol, alemão e mandarim. Resultam da análise de *checklists* realizadas pelos pais, onde são inventariadas, por categorias, as palavras produzidas pelas crianças. Nos estudos que adiante se destacam foi utilizado o inventário *MacArthur-Bates Communicative Development Inventory (MCDI)*, nas versões originais ou adaptadas.

Nos primeiros meses de aquisição de novas palavras o ritmo é lento, verificando-se, em média, a produção de duas novas palavras por mês (Papaeliou & Rescorla, 2011). Entre os 0;8 e o 1;4 observa-se uma grande variabilidade nas dimensões do vocabulário. Entre os 0;8 e 1;1 a maioria das crianças não produz ou produz poucas palavras (entre 0 e 3 palavras), sugerindo portanto que as primeiras palavras surgem nesta faixa (Bates *et al.*, 1994). O crescimento das primeiras produções ocorre a um ritmo lento, atingindo um valor médio de 1,8 aos 8 meses, 10 aos 12 meses, 64 aos 16 meses. Aos 30 meses o número médio de palavras produzidas é de 534. Contudo, estes valores são apenas tendências, verificando-se variabilidade. Por exemplo, aos 12 meses o número de palavras pode variar entre 0 e 24 (Bates, Dale, & Thal, 1995). Os resultados obtidos no estudo efetuado para o espanhol indicaram que a média de palavras produzidas foi de 0 palavras entre os 0;7 e 0;8,

aumentando para 4 palavras, entre os 0;11 e 1;00 e, posteriormente, 13,5 palavras entre 1;3 e 1;4 (Jackson-Maldonado, Thal, Marchman, Bates, & Gutierrez-Clellen, 1993).

Aos dois anos a criança com desenvolvimento linguístico típico produz um mínimo de 50 palavras (Rescorla *et al.*, 2000). Após atingir as 50 palavras, no inglês entre 1;6 e 1;9, verifica-se uma súbita aceleração na produção de novas palavras, verificando-se que a criança produz novas palavras todos os dias (Bloom, 1973; Nelson, 1973; Rescorla, 1980 citados por Rescorla *et al.* (2000) e Clark (1993)). No mandarim, este fenómeno acontece mais cedo, 1;4 (Tardif, Fletcher, Liang, & Kaciroti, 2009). Bates *et al.* (1995) assumem que esta variabilidade vocabular aumenta após os 13 meses, período em que se inicia esse rápido crescimento vocabular. Constataram no seu estudo, por exemplo, que aos 24 meses a média de vocabulário é de 312 palavras, atingindo um máximo de 534 e um mínimo de 89. Este fenómeno é designado por “*vocabulary burst*” ou “*vocabulary spurt*” e corresponde, portanto, a um aumento acentuado na taxa de produção de novas palavras, marcando a transição entre a produção de poucas palavras para uma produção aumentada (D’Odorico, Carubbi, Salerni, & Calvo, 2001; Hamilton, Plunkett, & Schafer, 2000; Huttenlocher, 1998). De acordo com D’Odorico *et al.* (2001), as crianças que vivenciam este fenómeno mais cedo, encontram-se em vantagem no que concerne ao ritmo de crescimento vocabular. Contudo, há diversos autores que refutam a existência de um “*vocabulary burst*”, alegando que o “*burst*” não é universal. Em algumas crianças o crescimento é contínuo, não sendo possível identificar um ponto de abrupta aceleração, e noutras o desenvolvimento vocabular é caracterizado por diversos “*bursts*”.

Bornstein, Leach, e Haynes (2004) observaram que há diferenças nas competências vocabulares entre crianças do sexo masculino e crianças do sexo feminino, sendo melhores nestas últimas. D’Odorico *et al.* (2001) verificaram que as crianças do sexo feminino atingiram o patamar das 50 palavras primeiro do que as crianças do sexo masculino. À semelhança dos estudos realizados para o inglês, também Tardif *et al.* (2009) concluíram que, no mandarim e cantonês, entre o 1;0 e 1;2, as competências das raparigas, a este nível, são superiores às dos rapazes. Papaeliou e Rescorla (2011), apesar de concordarem com a existência de diferenças entre géneros, consideram que esta é pouco significativa.

Relativamente ao padrão de palavras que utilizam, não são verificadas discrepâncias; tendencialmente, e independentemente da língua materna, a criança começa por utilizar palavras que representem o ambiente que a rodeia. De acordo com Sim-Sim (1998), as primeiras palavras selecionadas pela criança são simples do ponto de vista fonológico e refletem aquilo que é mais relevante no meio em que está inserida; por norma, pertencem à

classe dos nomes e estão associadas a entidades que se movam ou sejam passíveis de ser movimentadas. A predominância dos nomes está fortemente relacionada com a frequência com que são utilizados pelos adultos quando se dirigem à criança (Costa & Santos, 2003; Puyuelo & Rondal, 2007; Sim-Sim, 1998). Quanto maior a frequência a que uma criança é exposta a uma determinada palavra, mais cedo esta é adquirida, isto porque um maior número de apresentações em contextos extralinguísticos variados facilita o acesso ao significado.

Goodman, Dale, e Li (2008) procuraram, num estudo realizado para o inglês, verificar a existência de relação entre a frequência a que as crianças são expostas a determinadas classes de palavras e a idade de aquisição. Concluíram, por oposição ao defendido por Sim-Sim (1998), que os nomes são os menos produzidos pelos principais interlocutores, normalmente os pais, e são os primeiros a ser produzidos pela criança. As palavras de classe fechada são as mais frequentes nas produções dos interlocutores, mas das mais tardias em termos de produção. Constataram que dentro de cada classe analisada (nomes comuns, verbos, adjetivos, etc.) as palavras mais ouvidas eram efetivamente aquelas que a criança produz primeiro.

Até às 200 palavras, aproximadamente, a classe de palavras mais produzida pela criança é a classe dos nomes comuns, atingindo uma percentagem de 55,2% entre as 101 e as 200 palavras, momento em que as outras categorias começam a aumentar (Bates *et al.*, 1995; Bates *et al.*, 1994; Goodman *et al.*, 2008). Bates *et al.* (1995) referem que há uma preferência no uso de nomes associada ao crescimento lexical até aos 1;8-2;0. Esta percentagem diminui atingindo os 42% às 600 palavras, aproximadamente (Bates *et al.*, 1995; Bates *et al.*, 1994), sendo posteriormente observado um aumento lento, mas regular, no decorrer do desenvolvimento, de verbos. No período entre 1 e 50 palavras, os verbos correspondem a apenas 7,6%. A partir das 200 palavras aproximadamente, o número de verbos de ação, adjetivos e classes fechadas de palavras aumenta e a partir das 400 palavras as classes fechadas de palavras, que correspondem a 6% do total de produções, começam a aumentar proporcionalmente às outras categorias, até serem alcançados os 14% de produções (Bates *et al.*, 1995; Bates *et al.*, 1994; Goodman *et al.*, 2008; Tardif *et al.*, 2009).

De acordo com um estudo realizado por Rescorla (1980), quando a criança produz entre 72 a 247 palavras, a maioria corresponde a nomes, particularmente animais e alimentos. Clark (1993) e Goodman *et al.* (2008) verificaram que, para além dos nomes, as primeiras produções das crianças dizem respeito a aspetos relacionais e sociais. As crianças que aprendem mais nomes apresentam um crescimento vocabular mais rápido e as crianças

cujos vocabulário aumenta mais rapidamente têm uma percentagem maior de nomes (Bates *et al.*, 1995; Rescorla *et al.*, 2000).

Kim, McGregor, e Thompson (2000) observaram que, na marca das 50 palavras, tanto as crianças coreanas como as inglesas produziam mais nomes do que verbos. Contudo, as crianças coreanas produziam verbos em maior quantidade do que as inglesas. Num estudo realizado com crianças gregas, constatou-se que nas 50 primeiras palavras produzidas, 50% eram nomes (Papaeliou & Rescorla, 2011). No italiano, entre as 50 e 200 palavras os itens lexicais mais produzidos foram nomes, contudo foi observado um aumento mais acentuado nos verbos e adjetivos nesta etapa, comparativamente com os nomes, podendo este facto dever-se à estrutura da língua italiana (D'Odorico *et al.*, 2001).

De acordo com Tardif *et al.* (2009), a combinação de palavras, assim como o uso de formas flexionadas surge apenas após um longo período de produção isolada de palavras, verificando-se que essas palavras apresentam pouco ou nenhum traço de flexão. É também durante este período que se assinalam alterações no vocabulário, ocorrendo uma passagem da simples atribuição de um nome, para a predicação e, posteriormente, para a combinação de palavras. As primeiras combinações de palavras, ocorrem na maioria das crianças entre as 50 e as 200 palavras e a complexidade dos enunciados apresenta um aumento pronunciado quando ultrapassadas as 400. Tardif *et al.* (2009) relatam a existência de uma forte correlação entre as dimensões do vocabulário e as primeiras combinações de palavras normalmente entre as 50 e 100 palavras, assim como com a extensão média de enunciado (EME), entre as 300 e 600 palavras aproximadamente, sugerindo que “o desenvolvimento gramatical depende da criação de uma base lexical fundamental”.

Apesar de se verificar que o vocabulário apresenta um crescimento maior ao longo da infância, este desenvolvimento ocorre ao longo de toda a vida como resultado das novas vivências, invenções, oportunidades e tendências sociais (Read, 2000).

### **I. 1. 3. A relação entre desenvolvimento lexical e domínio vocabular**

A dissociação entre compreensão e produção é considerada um fenómeno robusto e generalizado no desenvolvimento linguístico, sendo por isso consensual que a compreensão precede a produção. Assim sendo, é indiscutível a existência de uma discrepância entre a quantidade de palavras reconhecidas por um indivíduo e a quantidade de palavras que este produz. Este facto é evidente no momento em que as crianças produzem as primeiras palavras. Bates *et al.* (1995) concluíram, num estudo realizado com

recurso ao MCDI, que uma criança não produzia qualquer palavra apesar de já compreender mais de 200.

A designação destas competências ainda não é clara e unânime, verificando-se a utilização de diferentes termos para a compreensão e produção de palavras, nomeadamente vocabulário passivo/ vocabulário ativo, compreensão/produção lexical, léxico/vocabulário, verificando-se também o uso indiscriminado do termo vocabulário ou léxico aplicado tanto à compreensão como expressão. Neste estudo, o conjunto de palavras que são compreendidas por um indivíduo será denominado por léxico, distinguindo-se daquelas que são utilizadas pelo mesmo e que constituem o vocabulário. Deste modo, o léxico é aquele que está relacionado com as ações de ouvir e ler, enquanto o vocabulário corresponde ao uso das palavras nos atos de fala e escrita.

Apesar desta distinção, aparentemente simples, não se consegue ainda determinar até que ponto estas competências se relacionam, ou seja, a partir de que momento uma determinada palavra passa do domínio compreensivo para o produtivo (Ma, 2009; Melka, 1997).

Após a primeira exposição a uma palavra, permanecem determinados traços desta no cérebro que, não sendo atualizados através de outras exposições, não integram o léxico. Assim, para que esta transferência ocorra, é necessário que a palavra seja apresentada várias vezes para que se dê o seu reconhecimento e a construção das respetivas redes semânticas. Este processo de reconhecimento vai-se consolidando e tornando tão rápido que permite que uma palavra seja localizada no léxico ainda antes de ter sido ouvida na totalidade (Melka, 1997).

As perspetivas mais convencionais apontam para uma progressão regular no processo de transição do reconhecimento de uma palavra para o respetivo uso. Num estudo, recentemente realizado por Gershkoff-Stowe e Hahn (2013), tornou-se mais uma vez evidente a existência de um hiato entre a compreensão e produção de novas palavras, no entanto, foram detetados diversos padrões no processo de transição. Os mesmos autores concluíram que tanto as crianças como os adultos experimentam paragens temporárias na aquisição de palavras, não sendo possível afirmar que há uma progressão linear nos estádios de aquisição de palavras, ou seja, numa determinada etapa compreendem e produzem novas palavras que mais tarde deixam de compreender e utilizar. Deste modo se pode afirmar que a produção de novas palavras não depende de um determinado nível de compreensão, mas da frequência com que foram anteriormente

produzidas, dos efeitos da competição com outras palavras com as quais se relacionam e da capacidade de resistirem à interferência das palavras com que as crianças estão familiarizadas. Para além disso, Gershkoff-Stowe e Hahn (2013) verificaram que as crianças com maior domínio vocabular são mais rápidas nas tarefas de compreensão, mas levam mais tempo para produzir palavras que fazem parte do seu léxico, do que aquelas cujo domínio vocabular é menor. Este padrão parece ser explicado pelo facto de ocorrer um aumento da densidade das redes semânticas associado à velocidade de aprendizagem de novas palavras, o que implica, posteriormente, uma maior ativação para gerar uma palavra, ato que consome, por sua vez, mais tempo. O processo de evocação de palavras exige, portanto, inicialmente algum esforço por parte do falante, porém da prática resulta alguma estabilidade no léxico e nas operações de seleção das palavras e, por conseguinte, uma automatização e rapidez na produção das palavras.

### **I. 1. 3. 1. A Diversidade Vocabular como preditora de sucesso escolar**

A linguagem oral assume extrema relevância no percurso académico de um aluno, uma vez que é utilizada não só para adquirir mas também para partilhar conhecimentos, relacionando-se indubitavelmente com a aprendizagem da leitura e escrita. Todas as modalidades usadas na comunicação – ouvir, falar, ler e escrever – correlacionam-se e influenciam-se mutuamente. Assim, da presença de alterações na linguagem oral podem advir graves dificuldades na aquisição de competências de leitura e escrita e, conseqüentemente, no cumprimento das metas curriculares (Leadholm & Miller, 1992).

O processo de escrita é complexo e constitui, nos dias que correm, uma necessidade básica. É a última das competências linguísticas a ser aprendida e exige, entre outros aspetos, um determinado nível de proficiência da linguagem oral, nomeadamente de domínio vocabular. Há um grande interesse por parte da comunidade científica em estudar as competências vocabulares da criança, atendendo à sua estreita relação com a aprendizagem da linguagem escrita. As melhores ferramentas que se podem fornecer a uma criança para que venha a obter sucesso, não só académico, mas na vida em geral, são um vasto e rico léxico e competências para utilizá-lo. A seleção vocabular constitui uma parte importante no processo de escrita, na medida em que a escolha da palavra certa assegura a transmissão da ideia que se pretende fazer chegar ao leitor (Olinghouse & Leaird, 2009). Por isso, um indivíduo que possua um vocabulário mais variado, isto é, que apresente nas suas produções uma maior DV, será mais profícuo nesta tarefa. Para Fergadiotis e Wright

(2010) a DV corresponde à variedade de palavras produzidas num texto e reflete a capacidade de um falante em aceder às palavras-alvo e assim construir unidades linguísticas maiores. A DV reflete o nível de competência linguística de um falante (Jarvis, 2013) e constitui um bom preditor do desenvolvimento de competências no âmbito da linguagem escrita; aqueles que têm um vocabulário mais amplo apresentam melhores competências no âmbito da compreensão oral, leitura e escrita (Pikulski & Templeton, 2004), diferenciando-se claramente das crianças que têm um vocabulário mais reduzido (Rebello, 1993). A DV exerce uma forte influência na avaliação qualitativa que os leitores fazem de um texto escrito, confirmando a ideia de que, ao efetuar a correção de uma atividade escrita, os professores têm em conta a DV aplicada (Engber, 1995).

Sabe-se que a aprendizagem da leitura, em idade escolar, é influenciada pelo desempenho vocabular das crianças em idade pré-escolar. Segundo, Wise, Sevcik, Morris, Lovett, e Wolf (2007) e Foy e Mann (2012) as competências vocabulares estão significativamente relacionadas com as capacidades que funcionam como pré-requisito para a aprendizagem da linguagem escrita, embora menos do que as competências lexicais. Para além disso, referem que o domínio vocabular é um importante preditor das capacidades de identificação de palavras no processo da leitura, mais do que o domínio lexical. Deste modo, crianças com reduzida DV podem apresentar mais dificuldades em identificar palavras, uma vez que não possuem representações fonológicas suficientes para serem mapeadas com a palavra escrita. Este facto confirma a conclusão a que chegaram Chiappe, Chiappe, e Gottardo (2004), que indicam que a performance em tarefas que exigem competências vocabulares está fortemente correlacionada com a performance em tarefas de competência fonológica e, por conseguinte, maus leitores apresentam um vocabulário mais reduzido.

### **I. 1. 3. 2. Fatores que influenciam a expansão vocabular**

As competências vocabulares das crianças são influenciadas pelas experiências de vida e estímulos a que são sujeitas, assim como pelo tipo de discurso que lhes é dirigido pelos interlocutores habituais. Estes fatores estão, obviamente, relacionados com o estatuto socioeconómico do agregado familiar, medido pela profissão, rendimentos e nível de escolaridade que possuem os seus elementos. Verifica-se, portanto, que durante a infância, as crianças cujo agregado familiar apresenta um estatuto socioeconómico e nível de escolaridade elevados são competentes no que concerne ao vocabulário e, para além disso,

o crescimento vocabular é mais rápido comparativamente com as crianças cujos agregados têm um estatuto socioeconómico mais baixo. (Hoff-Ginsberg, 1991; Hoff, 2003; Hoff, Laursen, & Tardif, 2002; Ninio, 1980) Huttenlocher (1998) analisou o vocabulário de crianças de 4 anos de nível socioeconómico médio e baixo e verificou que as crianças de nível baixo utilizam menos de dois terços de palavras diferentes do que as crianças de nível médio.

O estatuto socioeconómico é determinado por diversos fatores, sendo que apenas alguns afetam o desenvolvimento linguístico da criança. No caso concreto do vocabulário, verifica-se que o discurso maternal interfere no desenvolvimento vocabular da criança, enquanto variável do estatuto socioeconómico. A variabilidade no estatuto traduz-se numa variabilidade no discurso maternal que, por sua vez, conduz a uma variabilidade no tamanho do vocabulário da criança. As mães de estatutos socioeconómicos superiores produzem mais enunciados e mais extensos, sendo que este fator funciona como preditor do desenvolvimento vocabular da criança. A criança que é exposta a enunciados mais extensos desenvolve mais rapidamente o seu vocabulário. Uma vez que a enunciados mais longos produzidos pela mãe está associado um vocabulário mais diversificado, maior será a diversidade de palavras que a criança aprende e posteriormente produz (Hoff, 2003). Este autor concluiu que há uma forte correlação entre o nível de escolaridade da mãe e o número de palavras produzidas por crianças de 2 anos. O fator que influencia esta relação foi proposto por Rowe (2008) ao verificar que os conhecimentos que os pais apresentam em relação ao desenvolvimento da criança condicionam o modo como estes comunicam diariamente com os filhos. Por seu turno, o conhecimento que os interlocutores apresentam do desenvolvimento das crianças está relacionado com o nível de escolaridade que possuem (Rowe, 2008). As mães de níveis socioeconómicos mais baixos tendem a falar menos e a utilizar um vocabulário pouco diversificado. As mães de estatutos mais elevados produzem, em média, mais 20% de palavras do que as mães de estatuto socioeconómico inferior. Por conseguinte, as crianças de estatutos mais baixos também produzem menos palavras e menos variadas do que as de estatuto elevado. Verifica-se que as mães de estatutos baixos não estão sensibilizadas para as mudanças ocorridas nas capacidades e necessidades das crianças, pelo que a sua forma de ensinar não está orientada para o futuro e não consegue acompanhar o rápido desenvolvimento da linguagem para níveis mais complexos (Ninio, 1980).

Westerlund e Lagerberg (2008) indicam que as mães que apresentam uma boa capacidade de comunicar com a criança, através de atividades de descrição, esclarecimento,

incitação de conversação, promovem o desenvolvimento vocabular dos seus filhos. São também as mães que apresentam uma boa capacidade comunicativa que leem mais para os seus filhos. A idade da mãe parece estar também associada ao vocabulário das crianças, sendo que às mães mais novas correspondem filhos com maior vocabulário; contudo, são as mães mais velhas que leem mais para as suas crianças (Hoff-Ginsberg, 1991; Hoff, 2003; Sulzer-Azaroff, 1997).

A leitura conjunta, que consiste na leitura do livro pelo adulto com a participação e envolvimento da criança, durante ou após essa mesma leitura, pode influenciar positivamente as competências vocabulares das crianças quando implementadas técnicas específicas (Pollard-Durodola *et al.*, 2011). No estudo de Westerlund e Lagerberg (2008) concluiu-se que uma leitura conjunta entre os pais e a criança, durante pelo menos 6 vezes por semana, promove um aumento do vocabulário de, pelo menos, 0.3 desvios-padrão. O discurso da mãe para a criança durante a atividade de leitura conjunta de livros apresenta uma diversidade vocabular superior à utilizada no jogo com brinquedos (Hoff-Ginsberg, 1991).

Para além do contexto familiar, a produção de novas palavras pode ser aumentada em contexto escolar através de diversas atividades. Atividades que promovam uma interação entre educador/professor e a criança facilitam a integração e posterior produção das novas palavras (McKeown & Beck, 2014).

#### **I. 1. 4. Diversidade Vocabular: quantificação e avaliação**

##### **I. 1. 4. 1. Item lexical: o problema da definição de palavra**

Num estudo onde se pretenda estimar as dimensões do vocabulário dos falantes de uma língua, torna-se essencial definir o que é uma palavra, uma vez que a escolha da definição pode influenciar os resultados da pesquisa (Read, 2000). Todavia, a palavra não é um conceito de fácil definição, pelo que ainda não existe um consenso a este nível, na comunidade linguística (Azuaga, 1996; Read, 2000). Empiricamente, um comum falante de uma língua reconhece que a palavra é uma estrutura intermédia, menor do que uma frase e maior do que um som/fonema. Porém, esta será apenas uma das diferentes noções de palavra existentes, das quais se destacam e descrevem, seguidamente, quatro: palavra ortográfica, palavra fonológica, formas de palavras e lexema (Azuaga, 1996; Singleton, 2000).

A noção de palavra enquanto palavra ortográfica refere-se a qualquer unidade da escrita que se distingue de outra por espaços e por representações gráficas diferentes, tal como por exemplo *concerto*, *conserto* e *mãe*. Contudo, esta definição pode ser limitada, na medida em que se suporta na linguagem escrita, sendo esta competência posterior à linguagem oral e não utilizada em algumas comunidades. Para além disso, não clarifica a noção de palavra em relação àquelas que são representadas graficamente pelos mesmos sinais, mas que se distinguem no significado (palavras homógrafas). Sob outro ponto de vista, a palavra pode ser definida como uma sequência sonora, cujos constituintes estão unidos por um único acento principal, que possa ser delimitada por pausas. Assim, de acordo com esta noção, na frase que se segue podem ser identificadas três unidades com acento próprio, ou seja, três palavras fonológicas: “O menino | esqueceu-se | do chapéu”. Também esta definição levanta problemas, na medida em que, em determinadas línguas, a prosódia utilizada não simplifica esta divisão. De acordo com outra definição, a palavra corresponde a um lexema, ou seja, à unidade lexical. A um determinado lexema podem estar associadas várias formas de palavras que, quando formadas por flexão, não se distinguem enquanto itens lexicais autónomos, por serem apenas diferentes manifestações de um mesmo item lexical (Azuaga, 1996; Jackson & Amvela, 2000; Singleton, 2000). Gramaticalmente, o que se observa é a adição de terminações de flexão a uma forma base, sem alterar o seu significado ou a classe de palavras a que pertence. Tanto as terminações de flexão, como o radical da palavra são designados por morfemas. Uma outra definição considera que uma palavra é aquela à qual corresponde qualquer forma que represente uma ocorrência individual e que apresente uma forma física distinta – formas de palavra (Azuaga, 1996; Singleton, 2000).

#### **I. 1. 4. 2. A importância da definição de critérios no cálculo da DV**

A DV, tal como referido no capítulo introdutório deste estudo, relaciona-se com a capacidade de um indivíduo utilizar vocabulário diversificado, evitando a repetição (Malvern *et al.*, 2004) e reflete a complexidade do seu conhecimento vocabular.

Por razões práticas é impossível calcular todas as palavras que um falante nativo de uma língua conhece e é capaz de utilizar. Na seleção de palavras estão implicados aspetos como o contexto social, tema abordado, posição social dos interlocutores e meio de transmissão. Assim, os itens que um indivíduo utiliza numa conversa de café, com um amigo, sobre as férias, serão muito diferentes daquele que utilizará ao escrever um livro

sobre um determinado assunto específico. Deste modo, é importante que, ao determinar o tamanho do vocabulário de um indivíduo, se faça em relação a um contexto particular (Read, 2000).

Nos estudos que envolvem a análise da DV, a definição de *tokens* e *types* é fundamental (Youmans, 1990). O total de palavras produzidas por um falante é denominado por *token*, o que quer dizer que todas as ocorrências de uma palavra produzida num texto são contabilizadas. Por outro lado, o número de *types* corresponde ao número de diferentes formas de palavras produzidas, de modo a que uma determinada palavra seja contabilizada apenas uma vez, independentemente do número de vezes que ocorre (Read, 2000). No cálculo da DV está consequentemente implicada a decisão daquilo que constituirá um *type*.

Um primeiro conceito de *type* corresponde a palavras com um radical comum. Nalguns estudos de vocabulário, uma das primeiras etapas é proceder à identificação dos morfemas, de modo a que as terminações de flexão sejam contabilizadas como ocorrências do mesmo radical. Por exemplo, no inglês “*went*”, “*go*” e “*goes*” serão classificados como dois “*types*” diferentes, assim como “*é*”, “*és*” e “*fui*”, no português. Um problema na definição de palavras, que ocorre quando os *types* são contabilizados automaticamente em programas, é a existência de palavras homógrafas, uma vez que à mesma representação gráfica correspondem significados diferentes que são contabilizados como uma única palavra, quando na verdade deveriam ser duas (Read, 2000).

Owen e Leonard (2002) procuraram comparar a DV de crianças com Perturbações Específicas do Desenvolvimento da Linguagem (PEDL) com as de crianças com desenvolvimento típico da linguagem. Para o cálculo da DV consideraram que um *type* corresponde a um lema, podendo estar associado a diferentes radicais (regulares e irregulares), adotando assim o conceito de unidade lexical na definição de *types*. Assim, para estes autores, “*went*”, “*go*” e “*goes*” correspondem a um único *type*, bem como “*é*”, “*és*” e “*fui*”. Segundo os mesmos autores, o uso de formas de palavras conduzirá provavelmente a uma inflação dos valores de D.

No estudo de Olinghouse e Leaird (2009) considerou-se que cada estrutura morfológica diferente constituía uma palavra diferente, assumindo que cada uma tem um significado preciso e distinto que determina, não só o significado geral das ideias transmitidas, mas também estabelece relações diferentes dentro das próprias frases. Youmans (1990) seguiu a mesma definição de palavra, sugerindo, assim, que uma palavra corresponde a uma

cadeia distinta de caracteres alfanuméricos, incluindo hífens, precedidos e antecidos por um espaço. Seguindo esta definição, palavras como “chuva”, “guarda”, “guarda-chuva”, “bola” e “bolas”, constituem palavras diferentes. Também “*went*”, “*go*” e “*goes*” serão três *types*, bem como “é”, “és” e “fui”.

Malvern *et al.* (2004) investigaram o efeito dos critérios utilizados na definição de *types* no cálculo da DV. Obtiveram três versões diferentes da DV, mediante o critério utilizado. Estabeleceram, portanto, numa primeira abordagem que os *types* correspondiam a formas de palavras. Num segundo cálculo, as unidades com a mesma base morfológica e as formas irregulares constituíram palavras diferentes, incluindo casos de alomorfa no radical. A terceira versão resultou dos *types* enquanto unidades lexicais, ou seja, todas as formas regulares e irregulares foram reduzidas ao item lexical (lema) correspondente. Constataram que aos três critérios correspondem valores de DV fortemente correlacionados, contudo com algumas diferenças associadas ao desenvolvimento morfológico da criança.

No presente estudo, tal como se poderá constatar no capítulo II.5., o conceito de *type* utilizado vai ao encontro das formas de palavra, uma vez que desta forma o cálculo da DV considerará também o desenvolvimento morfossintático, em que quanto maior a diversidade de formas flexionadas, maior será o número de *types*. Para além disso, e retomando o estudo de Malvern *et al.* (2004), anteriormente exposto, em que se observou uma forte correlação entre os valores de DV dos três critérios descritos, a tomada de decisão fundamenta-se no critério que se apresenta mais vantajoso ao simplificar o processo de transcrição e análise dos dados. Também Costa (2010) terá seguido este critério para analisar o RTT, calculando-o automaticamente pelo CLAN, em produções escritas de alunos portugueses, prescindindo de uma análise morfológica prévia. Os dados de Costa (2010) mostram que há evolução de RTT nas produções escritas dos alunos.

No cálculo da DV, utilizando *software* específico, torna-se fundamental a definição de critérios que permitam ao sistema identificar itens válidos. Assim, após clarificar a noção de *type* a adotar num estudo, atendendo às definições de palavra acima descritas, no momento da transcrição o investigador deve selecionar normas de exclusão e inclusão de itens para uniformizar a contagem de *types* e *tokens*. Dos estudos realizados neste âmbito, destacam-se como mais importantes e frequentes os critérios que se listam de seguida (Malvern *et al.*, 2004; Owen & Leonard, 2002; Youmans, 1990):

- Exclusão de repetições que não tenham função de ênfase;

- Exclusão de todo o tipo de produções que não tenha valor referencial, tais como onomatopeias e interjeições;
- Inclusão de produções desviantes ou palavras incompletas resultantes de processos fonológicos ou de coarticulação, acompanhadas das palavras corretas/completas, de modo a não constituírem palavras diferentes;
- Nomes compostos por justaposição e aglutinação contabilizados como uma única palavra;
- Distinção entre nomes próprios e nomes comuns.

#### **I. 1. 4. 3. Cálculo da Diversidade Vocabular: medidas tradicionais**

Tem-se verificado uma necessidade permanente de obter uma medida fiável que permita determinar, com rigor, a diversidade de itens utilizados pelos falantes numa conversação. Diversos têm sido os estudos realizados neste sentido, não se verificando ainda, no entanto, consenso relativamente ao processo que confere resultados mais fiáveis, nem mesmo quanto ao termo mais adequado a utilizar: diversidade lexical, DV ou riqueza lexical (Malvern *et al.*, 2004). Uma vez que estes estudos focam medidas que contabilizam o número de palavras que as crianças são capazes de utilizar em conversação, e não refletem o número de palavras que estas compreendem, considera-se adequada a expressão Diversidade Vocabular, pelo que esta foi a designação adotada neste trabalho. A DV refere-se, portanto, à variedade de palavras que são selecionadas e utilizadas pelo falante num texto (oral ou escrito). Apesar de a comunidade científica reconhecer a importância da DV, a tarefa de identificação de um índice sólido, capaz de representá-la, tem-se revelado complexa (McCarthy & Jarvis, 2010). Ao longo dos anos, têm vindo a ser propostos e utilizados diversos índices ou medidas, no cálculo da DV, a maioria dos quais calculados através das relações estatísticas entre *types* e *tokens* (Jarvis, 2013). Estas medidas têm-se revelado extremamente importantes por permitirem retirar conclusões sobre o nível de competências linguísticas de um falante, para além de fornecerem informações sobre os estádios de aquisição vocabular (Malvern *et al.*, 2004).

Os primeiros estudos da DV, descritos por Jarvis (2013), remontam a 1935, ano em que George Zipf demonstrou a existência de uma taxa média de repetibilidade, ao verificar que o intervalo entre duas repetições de uma palavra pode ser determinado através de uma constante. Posteriormente, em 1938, John B. Carroll terá utilizado o termo DV para

descrever a existência de uma relação existente entre as palavras diferentes ocorridas num texto e a frequência com que estas se repetiam. Este autor propôs uma equação para verificar de que forma poderia a taxa de repetição de palavras variar em amostras de diferentes dimensões. Contudo, esta equação veio a revelar-se desajustada, pelo que houve necessidade de continuar os estudos neste âmbito. Em 1939, Wendell Johnson propôs o RTT, enquanto medida experimental, para determinar aquilo que o próprio designou de flexibilidade ou variabilidade vocabular. Este índice é um dos mais conhecidos neste âmbito e estabelece uma relação entre *types* e *tokens*. Será descrito detalhadamente no capítulo I.1.4.4.

No seguimento do RTT foi proposta outra medida de DV, designada por “Number of Different Words” (NDW). Esta medida, reconhecida como a mais simples, diz respeito, tal como a própria designação indica, ao número de palavras diferentes encontradas em amostras de linguagem com dimensões padronizadas (Malvern *et al.*, 2004). Nos estudos realizados por Klee, em 1992, e por Miller, em 1991, descritos por Malvern *et al.* (2004), foi possível concluir que o NDW é um bom indicador de desenvolvimento linguístico e apresenta uma forte correlação com a idade e com a EME. À semelhança do RTT, este índice está dependente do número total de palavras. Os valores de NDW, em amostras cuja dimensão é definida por um determinado período de tempo, vão depender do número de produções ocorridas. Assim, uma criança que fale mais terá certamente um valor de NDW superior ao de outra criança mais reservada. Também as amostras definidas pelo número de enunciados poderão induzir a uma confusão entre o que se pretende calcular, DV, e a EME, uma vez que crianças mais velhas produzem enunciados mais extensos que, por conseguinte, contêm mais palavras.

Para os estudos de DV o uso de medidas sólidas é essencial. A validade dessas medidas tem sido determinada não só pela capacidade de prever aspetos do desenvolvimento linguístico, mas também pela capacidade de não sofrer a influência das dimensões da amostra (Jarvis, 2013; Koizumi & In'ami, 2012). Outras medidas terão surgido, então, como forma de colmatar as limitações do RTT e NDW, no que diz respeito à influência exercida pelas dimensões da amostra. Destas, destaca-se o índice D, o índice de Diversidade Hipergeométrica de D (HD-D) e o “*Measure of Textual Lexical Diversity*” (MTLD). O índice D, mencionado por McKee, Malvern, e Richards (2000), é calculado através de *software* específico e tem como base técnicas de amostragem aplicadas ao índice RTT. A sua descrição será efetuada de modo mais pormenorizado no capítulo I.1.4.5. Em alternativa, foi desenvolvido por McCarthy e Jarvis, em 2007, o índice HD-D (McCarthy &

Jarvis, 2010). Este índice é determinado através do cálculo da probabilidade de um qualquer *type* surgir numa amostra aleatória de *tokens* (entre 35 e 50), tal como são recolhidos no cálculo do índice D. No cálculo desta probabilidade, conhecida como probabilidade de amostragens, está envolvida a distribuição hipergeométrica de diversidade, pela qual a cada amostra são atribuídas duas possibilidades, sucesso ou insucesso, e em que as amostras não são substituídas. O sucesso verifica-se quando uma palavra amostrada corresponde a um novo *type* e, por conseguinte, aumenta a contagem de *types* e de *tokens*; o insucesso ocorre quando uma palavra da amostra já foi anteriormente encontrada e, portanto, só há um aumento no número de *tokens*. As probabilidades obtidas para todos os *types* da amostra são somadas e o resultado corresponde ao valor obtido para HD-D. Os autores desta medida demonstraram que, enquanto o D faz um cálculo aproximado da probabilidade de encontrar um *type* numa amostra aleatória, pelo cálculo repetido de RTT, tal como será explicado no capítulo I.1.4.5., o HD-D é mais direto (deBoer, 2014; McCarthy & Jarvis, 2010).

O índice MTLT desenvolvido por McCarthy, em 2005, calcula o valor de RTT para cada palavra de forma sequencial, até atingir um valor de RTT constante, normalmente 0.72. À sequência de palavras obtidas até à fronteira de RTT, dá-se o nome de fator. Assim que é atingido o valor 0.72, o cálculo de RTT é recomeçado e as palavras anteriores são excluídas da contagem, até que o valor constante de RTT seja novamente atingido. O valor de MTLT é calculado dividindo o número total de palavras pelo número total de fatores.

#### I. 1. 4. 4. Rácio Type-Token

O RTT é uma medida de desenvolvimento linguístico amplamente utilizada para determinar a proporção entre *types* e *tokens* (Read, 2000) e foi proposto, tal como referido anteriormente, em 1939, por Wendell Johnson (Jarvis, 2013). É uma medida simples de calcular e reflete a DV de uma determinada amostra de linguagem de um falante. O RTT resulta da divisão efetuada entre o número de palavras diferentes (*types*) pelo número total de palavras produzidas na amostra (*tokens*) e é representado pela seguinte expressão (Shipley & McAfee, 2008):

$$RTT = \frac{\text{número de types}}{\text{número de tokens}}$$

Dado que corresponde a uma proporção de um total, o RTT varia entre um máximo de 1 e um mínimo teórico de 0 e é representado por um gráfico que inicia com uma linha reta, em que o número de *types* iguala o número de *tokens*, até que ocorra a primeira repetição

de uma palavra. Posteriormente, com o aumento das repetições, o número de *tokens* começa a ser superior ao dos *types* e há uma desaceleração da curva. O número de *types* atinge um máximo quando se esgota o vocabulário do falante. Consequentemente, o número de *tokens* aproxima-se de infinito, quando o número de *types* se aproxima do total do vocabulário do indivíduo (Youmans, 1990). Conclui-se, por isso, que um valor de RTT igual ou aproximado a 1 indica a existência de uma grande DV, enquanto valores próximos de 0 representam o oposto, ou seja, uma reduzida DV.

Dos estudos realizados destaca-se o de Templin, em 1957, por ter sido pioneira no estudo da DV recorrendo ao cálculo de *types* e *tokens* e, por conseguinte, o mais citado na literatura (Hess, Haug, & Landry, 1989; Hess, Sefton, & Landry, 1986; Malvern *et al.*, 2004; Miller, 1981; Scherer, Casarin, Zart, Ramos, & Assencio-Ferreira, 2002; Scherer & Souza, 2011; Watkins, Kelly, Harbers, & Hollis, 1995; Youmans, 1990). Esta autora apresentou uma metodologia de análise da DV em amostras de 50 enunciados de 480 crianças, provenientes de famílias com estatutos socioeconómicos diversificados, com desenvolvimento típico da linguagem e idades compreendidas entre os 3 e os 8 anos (distribuídas por 8 grupos etários – 3A, 3A6M, 4A, 4A6M, 5A, 6A, 7A e 8A - constituídos por 30 crianças do sexo feminino e 30 crianças do sexo masculino, cada um). Contudo, Templin na descrição do seu estudo nunca fez menção ao RTT, nem apresentou valores para este índice, limitando-se a indicar que, nas amostras analisadas, o rácio de palavras diferentes pelo número total de palavras era de 1:2, ou seja, a proporção seria de aproximadamente uma palavra diferente para pouco mais de duas palavras produzidas (Malvern *et al.*, 2004). Miller (1981) terá calculado os valores médios de RTT, seguindo a fórmula anteriormente representada, a partir dos resultados obtidos no estudo de Templin. Nos valores apresentados na Tabela 1, foi possível observar que os valores médios de RTT, para todos os grupos etários e estatutos socioeconómicos, se aproximam, de forma consistente, de 0.50. Esta consistência nos resultados tornou esta medida numa ferramenta extremamente valiosa e válida nos estudos linguísticos subsequentes, considerando-se o valor de 0.50 uma referência e, com valores muito inferiores a 0.50, a apontar para a existência de uma alteração linguística (Miller, 1981).

**Tabela 1** – Resultados médios de RTT obtidos por Miller, a partir do estudo de Templin. (Miller, 1981)

| Faixas etárias | Types |      | Tokens |      | Valores médios de RTT |
|----------------|-------|------|--------|------|-----------------------|
|                | Média | DP   | Média  | DP   |                       |
| 3;0            | 92.5  | 26.1 | 204.9  | 61.3 | 0.45                  |
| 3;6            | 104.8 | 20.4 | 232.9  | 50.8 | 0.45                  |
| 4;0            | 120.4 | 27.6 | 268.8  | 72.6 | 0.45                  |
| 4;6            | 127.0 | 23.9 | 270.7  | 65.3 | 0.47                  |
| 5;0            | 132.4 | 27.2 | 286.2  | 75.5 | 0.46                  |
| 6;0            | 147.0 | 27.6 | 328.0  | 65.9 | 0.45                  |
| 7;0            | 157.7 | 27.2 | 363.1  | 51.3 | 0.43                  |
| 8;0            | 166.5 | 29.5 | 378.8  | 80.9 | 0.44                  |

Contudo, a constância de valores de RTT ao longo das faixas etárias foi alvo de crítica, na medida em que a evidência demonstra que crianças com desenvolvimento típico da linguagem vão diversificando o seu vocabulário com o avançar da idade, não só pela aplicação de vocabulário mais específico, como também pelo uso de um maior número de sinónimos e antónimos, que conduz à redução de repetições (Malvern *et al.*, 2004).

O RTT de amostras, devidamente transcritas, é facilmente calculado recorrendo ao comando do programa *freq*, incluído, por sua vez, no programa CLAN. A descrição destes programas encontra-se no capítulo II.5. Representa-se, graficamente, por uma função cujos valores dependem fortemente do número total de palavras produzidas. Assim, em amostras reduzidas, o valor de RTT será elevado e a amostras maiores corresponderão valores de RTT menores. Este facto indica que, para o cálculo do RTT, é necessária padronização, relativamente às dimensões da amostra (Malvern *et al.*, 2004). Miller (1981) ressalva que a comparação com os dados obtidos pelo estudo de Templin, só é válida quando são seguidos os mesmos procedimentos, nomeadamente, no que concerne ao número de enunciados a utilizar (50 enunciados completos e inteligíveis). No cálculo do RTT deve ser estabelecido um

número mínimo de enunciados que se recomenda que sejam 50, de forma a refletir a diversidade adequada. Abaixo dos 50 enunciados obtêm-se valores reduzidos para a DV o que pode induzir a uma interpretação errada dos mesmos (Scherer & Souza, 2011).

Para além do número de enunciados, encontra-se na literatura referência a padronizações com base em unidades de tempo e número de *tokens*. Hess *et al.* (1989) e Hess *et al.* (1986), apoiados nos resultados do seu estudo, concluíram que valores de RTT não devem ser comparados quando obtidos a partir de amostras com diferente número de palavras. Recomendam a utilização de amostras constituídas por 350 palavras, no mínimo, para a obtenção de valores de RTT fiáveis. Malvern *et al.* (2004) referem que a amostra deve ser definida em função do número de *tokens*, em detrimento do número de enunciados e intervalo de tempo. De acordo com Hess *et al.* (1989), nenhum dado obtido para o RTT deve ser considerado normativo devido à fiabilidade questionável da medida. Já Retherford (1993), citado por Watkins *et al.* (1995), apresenta uma opinião favorável à utilização dos dados de Templin, defendendo que o valor médio de RTT de 0.50 deve ser utilizado como norma orientadora, em análises deste âmbito.

Dos estudos de DV realizados até ao momento, com recurso ao RTT, destacam-se, para o Português do Brasil, os de Scherer *et al.* (2002) e Scherer e Souza (2011). Não são conhecidos, até à data, estudos realizados para o português europeu.

Scherer *et al.* (2002) propuseram-se determinar valores médios de RTT para as faixas etárias de 3, 4 e 5 anos e analisar comparativamente os resultados obtidos pelas crianças do sexo masculino e feminino. No estudo participaram 60 crianças, falantes nativas do português brasileiro, com desenvolvimento típico da linguagem, 30 do sexo masculino e 30 do sexo feminino, distribuídas igualmente pelas três faixas etárias. Foram gravados 45 minutos de produções espontâneas das crianças, que depois de transcritas foram submetidas ao cálculo do RTT total e de segmentos de 50 enunciados. Dos resultados obtidos foi possível observar a inexistência de diferenças significativas entre géneros. No que diz respeito ao RTT total e ao RTT obtido nos 50 enunciados, verificou-se que, para ambos, o desempenho das crianças de 4 anos foi superior comparativamente com os de 3 e 5 anos, em virtude de o grupo apresentar uma maior DV.

Scherer e Souza (2011) procuraram analisar a evolução dos valores do RTT entre os 18 e 32 meses de idade, assim como estabelecer comparações do mesmo rácio entre géneros. Neste estudo, que seguiu a metodologia de Scherer *et al.* (2002), participaram 60 crianças do sexo feminino e masculino, com idades compreendidas entre os 18 e os 32 meses, falantes

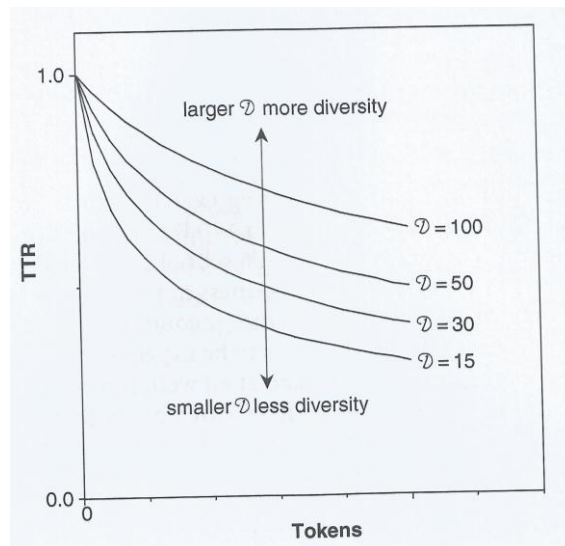
nativas do português brasileiro e com desenvolvimento típico da linguagem. Os autores concluíram que as diferenças verificadas entre sexos não foram estatisticamente significativas, à semelhança do que foi verificado por Scherer *et al.* (2002), e que entre os 18 e os 32 meses há uma evolução nos valores médios de *types* e *tokens*.

Em suma, o RTT tem sensibilidade para diferenciar faixas etárias, porém não deve ser utilizado isoladamente na avaliação da linguagem e no estabelecimento de diagnósticos, devendo ser complementado com testes padronizados. A medida deve ser utilizada apenas como referência, uma vez que descreve aspetos esperados numa determinada faixa etária, podendo variar com o contexto e o interlocutor (Scherer & Souza, 2011).

#### I. 1. 4. 5. D

O índice D surge da necessidade de criar uma medida de DV, válida e fiável, em alternativa às existentes, tal como o RTT. Este índice, proposto por McKee *et al.* (2000) utiliza cálculos repetidos de RTT, para estimar de que forma é que o RTT se comporta à medida que a dimensão de uma amostra é alterada. Matematicamente, é obtido pelo cálculo repetido do RTT numa série de *tokens* (de 35 a 50) seleccionados aleatoriamente e sem reposição (em cada cálculo). Os 16 pontos obtidos entre os 35 e 50 *tokens* são considerados suficientes para a execução do gráfico, porque, de acordo com Malvern *et al.* (2004), permitem subamostras satisfatórias das produções das crianças.

Este processo inicia, então, com o cálculo do RTT médio para 35 *tokens*, resultante de 100 repetições, de modo a garantir que o cálculo é realizado em toda a gama de *tokens*. Posteriormente é efetuado o mesmo cálculo para 36 *tokens* e assim sucessivamente até serem atingidos os 50. Os valores médios de RTT obtidos em cada amostra são assinalados num gráfico. O valor do índice D resulta da melhor aproximação conseguida entre a curva obtida no gráfico que contém os dados da amostra e aquela que é considerada a curva ideal, segundo a teoria da probabilidade, porque prediz os valores ideais esperados para o RTT, para aquela amostra. A aplicação de uma fórmula matemática, em diversos estudos realizados, permitiu a obtenção de um conjunto de curvas, representadas no Gráfico 1, e que são consideradas ideais uma vez que representam amostras reais de linguagem, nas quais é evidente a relação do RTT com o número de *tokens*, à medida que as dimensões de uma amostra aumentam.



**Gráfico 1** – Representação das curvas ideais de RTT e relação com o índice D. (Malvern *et al.*, 2004)

A forma como as curvas se encontram dispostas no Gráfico 1 mostra a distância que poderá existir entre duas amostras, demonstrando assim diferentes níveis de DV, representados pelos valores de D. Não é objeto deste trabalho descrever minuciosamente este processo, importa antes referir que o índice D resulta da diferença entre o traçado de uma curva ideal e a curva de uma dada amostra, correspondendo ao melhor ajuste encontrado. Os valores de D situam-se entre 10 e 100, considerando-se que valores elevados de D refletem um alto nível de DV e vice-versa (Malvern *et al.*, 2004; McCarthy & Jarvis, 2007; McKee *et al.*, 2000). De acordo com Malvern *et al.* (2004) e McKee *et al.* (2000), o índice D apresenta três vantagens: é independente do número de *tokens*; faz uso de todos os dados disponíveis; é mais informativo, na medida em que indica de que modo é que o RTT vai variando numa determinada amostra.

Todo este procedimento é bastante complexo, mas possível de executar através do comando do programa *vocd* que, à semelhança do programa *freq*, utilizado para o cálculo do RTT, é realizado no programa CLAN. Este *software* exige um mínimo de 50 *tokens* para calcular o índice D e verifica-se que, de cada vez que o comando é executado, obtêm-se valores diferentes resultantes da técnica de amostragem utilizada. Contudo, e apesar de essas diferenças serem irrelevantes, o *vocd* efetua um cálculo triplo do D, sendo que o valor final corresponde à média dessas três repetições, garantindo assim um nível elevado de consistência (Malvern *et al.*, 2004; McCarthy & Jarvis, 2007; McKee *et al.*, 2000).

Owen e Leonard (2002) propuseram-se calcular e analisar comparativamente a DV, através da aplicação do índice D, em 38 crianças com desenvolvimento típico da linguagem,

com idades compreendidas entre os 2;2 e os 6;1, e em 33 crianças com PEDL, com idades compreendidas entre os 3;7 e os 7;3. A DV foi calculada para amostras de 100 enunciados, 250 e 500 *tokens*, produzidos por cada criança, recolhidos de amostras de discurso espontâneo resultantes da interação entre os investigadores e a criança. Os autores do estudo concluíram que o índice D é capaz de refletir diferenças desenvolvimentais e clínicas, ao observarem que tanto as crianças mais velhas, como aquelas que apresentam desenvolvimento típico da linguagem, obtiveram valores mais elevados neste índice, comparativamente com as crianças mais novas e com PEDL, respetivamente. Verificaram também que o índice D sofre alguma influência das dimensões da amostra, no entanto não na mesma proporção que outros índices, tais como o RTT e o NDW. Malvern *et al.* (2004) demonstraram a existência de alguns fatores que influenciam os resultados de D. Observaram que o índice é afetado pelo critério que define os *types*. Assim, quando são utilizadas formas flexionadas o valor de D é superior ao obtido quando são utilizados os lexemas. Os mesmos autores consideraram que o uso de lexemas é o mais adequado no cálculo da DV, uma vez que os resultados estarão associados e dependerão fundamentalmente das questões vocabulares e não do desenvolvimento morfológico, tal como sucede quando são utilizadas formas flexionadas. Para línguas ricas em formas flexionadas, como é o caso do espanhol, Malvern *et al.* (2004) sugerem a utilização de uma outra medida, designada por Diversidade Flexional (DF), calculada a partir da subtração do valor de D, obtido para lexemas, ao valor de D, obtido para formas flexionadas. Também o contexto influencia os valores de D. Malvern *et al.* (2004) referem que contextos mais familiares, onde as situações são mais liberais, conduzem a valores de D mais elevados, comparativamente com os resultados obtidos em contextos laboratoriais ou clínicos, onde as atividades realizadas são controladas, reduzindo o número de tópicos abordados. Aqui se verifica, mais uma vez, que é muito importante adotar critérios padronizados de forma a poder efetuar comparações entre estudos de DV.

O índice D tem aplicações não só no estudo do desenvolvimento da linguagem na criança, mas também na análise das produções do adulto. Fergadiotis e Wright (2010) procuraram determinar a existência de diferenças entre adultos com e sem afasia, no que concerne à DV, recorrendo a técnicas de eliciação de discurso diferentes: descrição de sequências de imagens, descrição de uma única imagem e conto de uma história. Para o efeito, utilizaram o índice D como medida da DV. Concluíram que, em todas as tarefas solicitadas, os indivíduos com afasia apresentaram níveis inferiores de DV, comparativamente com os indivíduos sem afasia.

O índice D é uma medida válida e apresenta níveis aceitáveis de fiabilidade quando aplicado ao estudo linguístico de crianças com desenvolvimento típico e não é, ao contrário do índice RTT, afetado pelo tamanho da amostra de linguagem (Mckee et al., 2000).

#### **I. 1. 4. 6. Procedimentos de avaliação**

Os terapeutas da fala, enquanto profissionais responsáveis pela prevenção, avaliação, diagnóstico e tratamento de perturbações relacionadas com a linguagem, têm estado envolvidos em estudos científicos direcionados para o processo de desenvolvimento da linguagem na criança, desde 1950. As metodologias utilizadas nos estudos de avaliação da linguagem são diversificadas e a sua seleção depende do aspeto linguístico que se pretende analisar (Lund & Duchan, 1983).

Para descrever um determinado comportamento linguístico é muito importante optar por uma abordagem que permita a obtenção de níveis ótimos de performance, isto é, um desempenho que se aproxime da realidade, tanto quanto possível. As abordagens mais utilizadas na recolha de dados linguísticos são os testes standardizados, escalas de desenvolvimento e procedimentos informais/não standardizados, onde estão incluídas as observações naturalistas, podendo estas três abordagens ser utilizadas isoladamente ou de forma combinada (Miller, 1981).

A recolha de amostras de discurso espontâneo é um dos métodos mais recorrentes no estudo da linguagem oral da criança, uma vez que permite uma análise, segundo uma perspetiva mais naturalista, das competências linguísticas da criança (Heilmann, Nockerts, & Miller, 2010; Pan, Rowe, Spier, & Tamis-Lemonda, 2004). A análise de amostras de linguagem constitui um dos meios mais válidos para determinar a performance linguística de uma criança, porque decorre de uma recolha de discurso em variados contextos comunicativos, contribuindo, então, para a obtenção de medidas mais exatas. Para isso, é importante que no processo de recolha se tenha em linha de conta a obtenção de uma amostra representativa das produções espontâneas da criança (Leadholm & Miller, 1992). A dimensão da amostra pode ser definida pelo número de enunciados produzidos pela criança ou pela duração da interação. De acordo com Miller (1981), numa amostra dimensionada pelo fator tempo, a análise de frequências é mais significativa e a mais flexível. Este autor considera que a duração da amostra não deve ultrapassar os 30 minutos, evitando-se assim a fadiga da criança. Neste período de tempo de interação, uma criança com idade superior a 24 meses e com desenvolvimento típico da linguagem produz entre

100 e 200 enunciados. Leadholm e Miller (1992), Retherford (1993), assim como Cole, Mills e Dale (1989), citados por Heilmann *et al.* (2010) referem que uma amostra deve conter 100 enunciados completos e inteligíveis, sendo esta a quantidade necessária para que a criança exiba uma variedade de competências linguísticas. Heilmann *et al.* (2010) observaram no seu estudo que as medidas de DV e de extensão de enunciado são mais viáveis quando são utilizadas amostras pequenas. No que diz respeito à DV, estudos recentes indicam que esta pode ser influenciada pelo número de palavras que contém a amostra. Koizumi e In'ami (2012) concluíram que o índice MTLTD comparativamente com o índice D e o RTT é a medida que menos influência sofre das dimensões da amostra, quando esta contém entre 50 e 200 *tokens*. Porém, num estudo realizado também para determinar os efeitos das dimensões de amostras nos índices D, NDW e RTT, em que Silverman e Ratner (2002) compararam os valores dessas medidas em amostras completas e em metades das mesmas, o índice D não demonstrou sensibilidade a alterações das dimensões da amostra. Já o índice RTT, neste e noutros estudos, tem-se revelado, de facto, bastante sensível às dimensões das amostras, verificando-se que a amostras grandes correspondem valores baixos de RTT. Hess *et al.* (1986) referem que, no cálculo do índice RTT, para amostras de linguagem oral de crianças em idade pré-escolar, não devem ser comparadas amostras com diferentes números de palavras, devendo este total ser, no mínimo, de 350 palavras para que se obtenha um mínimo de viabilidade de 0.70.

O método de recolha de dados baseado nas amostras de discurso espontâneo pode sofrer uma grande influência de alguns fatores, tais como contexto situacional, temas abordados, materiais utilizados e familiaridade com o interlocutor (Yont, Snow, & Vernon-Feagans, 2003). Os motivos mais frequentes que justificam a não utilização da mesma, na prática clínica, prendem-se com o tempo que é necessário despendido na elicitación, transcrição e análise das amostras. O tempo necessário à transcrição depende do tamanho da amostra e da inteligibilidade das produções; em média, uma amostra linguística de 15 minutos pode demorar entre 1 hora e 15 minutos e 2 horas a ser transcrita. Uma das formas de contornar este aspeto é utilizar pequenas amostras de discurso. Porém, é importante ressaltar que a utilização de amostras mais reduzidas pode constituir um risco, na medida em que a criança não possui tantas oportunidades para demonstrar as suas competências (Heilmann *et al.*, 2010).

Tal como referido, o contexto é muitas vezes evocado nos estudos de aquisição e desenvolvimento da linguagem na criança para explicar determinados fenómenos. Segundo Yont *et al.* (2003), as crianças fazem um uso diferenciado da linguagem mediante o

contexto situacional. Assim, as técnicas de elicitación utilizadas pelo interlocutor devem assegurar um número suficiente de oportunidades para a criança demonstrar as suas competências linguísticas (Heilmann *et al.*, 2010). O desempenho de cada falante na seleção do vocabulário pode ser influenciado por fatores extralinguísticos, tais como o tipo de discurso que o falante utiliza para responder a uma técnica de elicitación, o tópico do discurso e o interlocutor. Deste modo, a DV, determinada pelos mais diversos índices disponíveis, pode variar com o tipo de discurso e a técnica de elicitación utilizada (Fergadiotis & Wright, 2010). Das técnicas utilizadas destacam-se como as mais frequentes a conversação e o uso de narrativas (Heilmann *et al.*, 2010). Estas podem ser utilizadas nos mais diversos contextos situacionais, sendo a leitura conjunta de livros e a conversação durante a realização de jogos com brinquedos os mais frequentes nos estudos da linguagem na criança. Contudo, estes contextos também apresentam as suas limitações. A utilização de um número reduzido ou abundante de brinquedos vai limitar o número de produções, verificando-se que é necessária alguma razoabilidade nessa escolha. Assim, é importante manter alguma consistência nos materiais utilizados de modo a assegurar que o nível de estimulação é idêntico nas diferentes recolhas (Heilmann *et al.*, 2010). Os brinquedos devem ser selecionados de acordo com a idade da criança e o seu interesse. Aqueles que possibilitam a realização de mais tarefas diversificadas, principalmente em jogo simbólico, são os que mais facilitam a elicitación de discurso. Miller (1981) sugere a utilização de bonecas, utensílios de cozinha, quintas com os respetivos animais e bombas de combustível com carros e pessoas. Refere ainda que para fomentar o envolvimento da criança na interação deve ser-lhe dada a possibilidade de escolha dos brinquedos a utilizar.

A conversação resulta do principal objetivo da comunicação, que é a partilha de ideias através da troca de mensagens. As crianças com desenvolvimento típico são capazes de se envolver numa conversa desde tenra idade, desde que reconheçam no seu interlocutor o interesse pela interação. Esta interação é essencial na recolha de dados linguísticos, considerando-se que os cinco primeiros instantes são determinantes no decorrer do processo. Se, eventualmente, ocorre uma falha no estabelecimento da relação entre o adulto e a criança, pode ficar de imediato comprometida a espontaneidade necessária, por parte da criança. Deste modo, para desbloquear algum tipo de constrangimento, nos momentos iniciais da interação é aconselhável que o adulto faça uma breve e simples explicação dos propósitos da interação, bem como, por exemplo, uma apresentação dos materiais a utilizar. Durante a conversação propriamente dita, é importante que o interlocutor seja paciente, esperando de forma estruturada pelas

intervenções da criança, reduza ao máximo as questões e seja capaz de ouvir e permitir que a criança lidere a conversação, seguindo os seus tópicos. (Leadholm & Miller, 1992; Miller, 1981) Miller (1981) e Leadholm e Miller (1992) sugerem um protocolo para recolha de amostras de linguagem no qual são definidos 15 minutos para a conversação, onde são abordados tópicos como as férias, atividades realizadas em família, atividades realizadas na escola, animais domésticos, entre outros. Após a conversação, é realizada a narrativa, também com a duração de 15 minutos. Dentro da narrativa, há diferentes meios para eliciar discurso, podendo recorrer-se à descrição de uma única imagem, a sequências de imagens, recontar uma história conhecida ou através da criação espontânea a partir de um livro sem texto escrito (Heilmann *et al.*, 2010).

As crianças envolvem-se desde cedo e com facilidade na conversação, contudo, a narrativa só ocorre por volta dos 3 anos, ainda que de forma primitiva e produzindo poucos enunciados. As produções ocorridas nas conversas têm enfoque principal nas tarefas que a criança está a desenvolver no momento, nos seus brinquedos, no aqui e agora. À medida que a idade vai avançando, as crianças tornam-se aptas a utilizar a linguagem de modo a transpor o espaço e tempo em que estão inseridas (Leadholm & Miller, 1992). No estudo realizado por Yont *et al.* (2003) constatou-se que as crianças produziram mais *tokens* por minuto numa interação durante o jogo, do que na leitura conjunta de livros, tendo a produção de *types* sido semelhante nos dois contextos.

Leadholm e Miller (1992) verificaram que as medidas de análise linguística são mais sensíveis às alterações desenvolvimentais, nas amostras de discurso narrativo, comparativamente com as amostras de discurso conversacional, em crianças em idade pré-escolar e escolar.

Todo o processo de recolha de amostras de discurso espontâneo deve ser gravado em formato áudio ou audiovisual, sendo esta última preferível, com equipamento capaz de efetuar um registo de boa qualidade. A gravação é útil, na medida em que liberta o interlocutor na interação e facilita a posterior transcrição e análise dos comportamentos linguísticos da criança, que muitas vezes não são imediatamente identificados no momento da recolha.

## CAPÍTULO II: METODOLOGIA

No presente capítulo, enunciam-se as questões de investigação que conduziram à realização deste estudo, assim como todos os aspetos que orientaram a mesma, nomeadamente as variáveis envolvidas e hipóteses. Posteriormente, efetua-se a caracterização da amostra e a descrição dos materiais e procedimentos utilizados na recolha, tratamento e análise dos dados.

Importa referir que toda a componente prática descrita adiante, que resultou na constituição da amostra, recolha e tratamento de dados, foi realizada em colaboração com a investigadora Ana Rita Pina Coimbra de Campos, o que proporcionou a utilização de uma amostra e *corpus* maior.

### II. 1. Questões de investigação e tipo de estudo

No processo de avaliação das competências lexicais de um indivíduo, é importante determinar a variedade e especificidade das palavras que este utiliza na abordagem de um tópico. Tal como foi exposto anteriormente, torna-se imprescindível que, no momento de avaliar a DV, o investigador/avaliador conheça as vantagens e limitações dos diferentes índices existentes, nomeadamente do RTT e D, descritos nos capítulos I.1.4.4. e I.1.4.5., respetivamente, de modo a selecionar o mais adequado. Deste modo, considera-se importante obter resposta para as seguintes questões:

**Questão 1:** A DV, medida pelos índices RTT e D, aumenta com a idade?

**Questão 2:** Existem diferenças nos valores médios dos índices RTT e D obtidos pelas crianças em idade pré-escolar e escolar?

**Questão 3:** Em qual dos contextos discursivos, narrativo e espontâneo, se verifica um melhor desempenho no RTT e D?

**Questão 4:** O desempenho das crianças nos índices RTT e D, em contexto narrativo, é diferente entre grupos etários?

**Questão 5:** O desempenho das crianças nos índices RTT e D, em contexto espontâneo, é diferente entre grupos etários?

**Questão 6:** Estarão os índices RTT e D relacionados com desenvolvimento linguístico, da mesma forma que a idade?

## II. 2. Hipóteses e Variáveis

Para cada uma das questões anteriormente colocadas, formularam-se hipóteses de modo a predizer os resultados esperados. Assim, para a primeira questão de investigação definiu-se a seguinte hipótese:

**H1:** As crianças mais velhas apresentam um desempenho de RTT e D superior ao verificado nas crianças mais novas.

Esta hipótese apoia-se nos estudos referidos na revisão bibliográfica (Malvern *et al.*, 2004) que mostram haver alterações nos valores médios de RTT e D, com o evoluir da idade, por oposição ao verificado por Miller (1981), em que o valor de RTT mantém alguma constância, tal como se pode observar na Tabela 1.

No que diz respeito à segunda questão enunciou-se a seguinte hipótese.

**H2:** O desempenho no RTT e D das crianças em idade pré-escolar é inferior ao verificado nas crianças em idade escolar.

Esta segunda hipótese parte da observação de Duarte *et al.* (2011), que indicam existir uma evolução na diversidade vocabular das crianças, decorrente do processo de escolarização.

Relativamente à terceira questão formulou-se a seguinte hipótese:

**H3:** As crianças apresentam uma DV maior em contexto narrativo, comparativamente com o contexto espontâneo.

Esta hipótese apoia-se nos resultados obtidos por Yont *et al.* (2003), que observaram uma maior produção de *tokens* por minuto numa interação durante o jogo, do que na leitura conjunta de livros, bem como uma produção de *types* semelhante nos dois contextos. Daqui se depreende que o rácio é superior no contexto narrativo.

Para as questões 4 e 5 definiram-se as seguintes hipóteses:

**H4:** O desempenho nos índices RTT e D, em contexto narrativo, apresenta variações de acordo com a idade da criança.

**H5:** O desempenho nos índices RTT e D, em contexto espontâneo, apresenta variações de acordo com a idade da criança.

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. Tendo como base as hipóteses 4 e 5, procurar-se-á verificar se existem efetivamente diferenças na DV das crianças, medida pelo RTT e D, nos diferentes contextos, de acordo com a idade.

Para a questão 6 estabeleceu-se como hipótese:

**H6:** Tanto a idade como a DV, medida pelos índices RTT e D, estão correlacionadas positivamente com o desenvolvimento linguístico, de tal modo que:

**H6.1:** crianças mais velhas apresentam um desenvolvimento linguístico maior.

**H6.2:** a um nível linguístico superior está associada uma maior DV.

Para as análises supracitadas consideram-se como variáveis dependentes o índice D - total, narrativo e espontâneo e o índice RTT – total, narrativo e espontâneo (variáveis rácio), e como variáveis independentes o desenvolvimento linguístico (variável rácio), a idade das crianças, em meses (variável rácio), faixas etárias (variável ordinal) e o nível de ensino em que se encontravam as crianças – pré-escolar e escolar (variável nominal).

### **II. 3. Caracterização da amostra**

A amostra é constituída por 50 crianças, 24 do sexo masculino e 26 do sexo feminino, falantes nativos do português europeu, com idades compreendidas entre os 3 anos e 0 meses e os 7 anos e 11 meses, com desenvolvimento típico da linguagem e residentes no concelho de Santa Comba Dão.

A seleção da amostra foi efetuada com a cooperação dos educadores e professores do 1º ciclo do Agrupamento de Escolas de Santa Comba Dão, de acordo com os critérios idade, língua materna e desenvolvimento linguístico. Para além destes critérios de seleção, definiu-se também que seriam incluídas na amostra apenas as crianças que, no momento da recolha de dados, produzissem um mínimo de 50 enunciados, por ser o número

comummente recomendado e indispensável à obtenção de medidas válidas e fiáveis, na análise de competências semânticas e sintáticas das crianças (Heilmann *et al.*, 2010; Miller, 1981), e pelo menos 50 *tokens*, por ser o número mínimo exigido no cálculo do índice D (McKee *et al.*, 2000).

As crianças estavam matriculadas no Agrupamento de Escolas referido, 35 frequentavam o ensino pré-escolar, 15 frequentavam o 1º ano de escolaridade e encontravam-se distribuídas pelos Centros Educativos do Norte, Centro e Sul.

A amostra foi subdividida em 5 grupos menores de 10 crianças de ambos os sexos, de acordo com as seguintes faixas etárias: [3;0-3;11], [4;0-4;11], [5;0-5;11], [6;0-6;11], [7;0-7;11] e a sua distribuição pode ser analisada na Tabela 2.

**Tabela 2** – Distribuição da amostra por sexo e faixa etária.

|                       |                  | Faixa etária |            |            |            |            |
|-----------------------|------------------|--------------|------------|------------|------------|------------|
|                       |                  | [3;0-3;11]   | [4;0-4;11] | [5;0-5;11] | [6;0-6;11] | [7;0-7;11] |
| <b>Média de Idade</b> |                  | <b>3;8</b>   | <b>4;5</b> | <b>5;6</b> | <b>6;5</b> | <b>7;3</b> |
| <b>Sexo</b>           | <b>Feminino</b>  | 9            | 4          | 5          | 4          | 4          |
|                       | <b>Masculino</b> | 1            | 6          | 5          | 6          | 6          |
| <b>Total</b>          |                  | 10           | 10         | 10         | 10         | 10         |

#### II. 4. Material e Procedimentos da Recolha da Dados

Efetuiu-se um primeiro contacto pessoal e, posteriormente, por escrito, com a Diretora do Agrupamento de Escolas de Santa Comba Dão, informando-a sobre os propósitos da investigação e procedimentos da recolha de dados. A autorização foi formalmente concedida (Apêndice A) após a obtenção da aprovação da Direção-Geral da Educação do Ministério da Educação e Ciência, entidade máxima que aprecia os pedidos para realização de estudos de investigação, em contexto escolar. Seguidamente, foi solicitada a colaboração de todos os educadores e professores do 1º ciclo (a lecionar o 1º ano de escolaridade) deste agrupamento de escolas, na constituição da amostra, através da seleção de crianças que se enquadrassem nos critérios pré-definidos para o estudo em questão. Identificados potenciais participantes, seguiu-se uma breve apresentação, por

escrito, do projeto de dissertação (Apêndice B) aos respectivos encarregados de educação. De forma a assegurar os princípios éticos da beneficência, respeito pela dignidade humana e justiça, inerentes aos estudos científicos, elaborou-se um termo de consentimento livre (Apêndice C) e esclarecido, que foi preenchido e assinado pelos mesmos, aprovando a participação de cada criança na recolha de dados.

De um total de 94 documentos de consentimento informados entregues aos encarregados de educação das crianças identificadas pelos educadores e professores, receberam-se 84 preenchidos. Destes, excluíram-se cinco devido ao preenchimento incorreto, por parte dos encarregados de educação. De um total de 79 crianças, efetuou-se uma pré-seleção de 50 crianças elegíveis para o estudo, procurando, sempre que possível, um número aproximadamente igual de ambos os sexos e a maior heterogeneidade possível, dentro de cada faixa etária, no que diz respeito à idade em meses. Durante o processo de recolha de dados houve necessidade de substituir 6 crianças, das quais cinco foram substituídas por incumprimento dos critérios de inclusão no estudo e uma por falta de comparência nos dias em que foram agendadas as recolhas.

As 50 crianças selecionadas para participar na recolha de dados foram distribuídas equitativamente por dois subgrupos, cada um constituído por 25 crianças de todos os grupos etários, de modo a facilitar o processo de recolha pelas duas investigadoras. A cada investigadora foi atribuído, também aleatoriamente, um subgrupo.

A recolha de dados foi realizada nos Centros Educativos do Agrupamento de Escolas de Santa Comba Dão, em dois momentos que ocorreram com um intervalo de tempo inferior ou igual a 7 dias. A literatura não é clara relativamente ao tempo a decorrer entre recolhas de amostras de linguagem. A escolha deste intervalo de tempo prendeu-se com a necessidade de evitar alterações nos parâmetros a analisar, decorrentes do processo normal de desenvolvimento da linguagem.

Cada momento consistiu numa interação entre a investigadora e a criança que foi gravada utilizando equipamento audiovisual (gravador áudio digital portátil - Sony ICD-P 520, camcorder digital, vídeo HDD - Sony DCR-SR 58 e tripé para camcorder – Vivitar V2400 GX). Este tipo de gravação possibilita a recolha de produções verbais orais e assume uma grande importância no momento da transcrição dos dados (Demuth, 1998). Cada gravação teve a duração aproximada de 30 minutos, cronometrados com telemóvel, de modo a garantir o número de enunciados e *tokens* pretendidos (Miller, 1981). Uma gravação com esta duração possibilita a realização de diversas tarefas e aumenta o leque de

análises a efetuar, posteriormente (Lund & Duchan, 1983; Miller, 1981). Quando possível, as investigadoras realizaram as recolhas em simultâneo, contudo, sempre em salas diferentes.

Nos momentos de recolha de dados recorreu-se a tarefas diferenciadas, tendo em vista a produção de discurso, por parte da criança. Todas as crianças foram previamente esclarecidas relativamente às atividades a desenvolver e tomaram conhecimento da gravação que seria realizada. Miller (1981) refere que os primeiros momentos de interação são extremamente importantes e definem a colaboração da criança ao longo da recolha. Uma abordagem inicial agradável para a criança por parte do investigador favorece a espontaneidade da mesma e, conseqüentemente, a sua performance linguística aproximar-se-á da realidade. Por isso, sugere que, no início de cada sessão de gravação, a criança seja informada do propósito da mesma, assim como dos materiais de gravação a utilizar.

Ambos os momentos seguiram uma seqüência de atividades que serão seguidamente descritas.

O primeiro momento de cada criança iniciou com conversação de caráter informal, estabelecida entre a investigadora e a criança, com tópicos antecipadamente definidos e que se considerava ir ao encontro do interesse da mesma, nomeadamente, férias, atividades curriculares, atividades extra-curriculares e família (Miller, 1981). Seguidamente, e sem ordem definida, criaram-se situações de jogo simbólico entre a criança e a investigadora e solicitou-se à criança a produção de discurso narrativo a partir da descrição de duas seqüências de imagens.

No jogo simbólico, de forma a maximizar o interesse da criança, foi-lhe pedido que seleccionasse até três conjuntos de brinquedos de um total de oito conjuntos, de acordo com a sua preferência (Miller, 1981). Perante os objetos escolhidos a investigadora estabelecia, conjuntamente com a criança, uma ordem para a utilização dos mesmos. Os conjuntos de brinquedos apresentados apropriavam-se às faixas etárias em questão e iam ao encontro das eventuais preferências de ambos os géneros; continham miniaturas diversificadas e aludiam às seguintes temáticas ou profissões: polícia, bombeiro, construtor civil, médico, oficina, casa, parque e quinta. Os conjuntos de brinquedos, que podem ser observados no Apêndice D, eram compostos pelas seguintes miniaturas:

- polícia: transportes, polícias, sinalética e árvores, cenário de cidade;
- bombeiros: transportes, bombeiros, sinalética, tendas, postos de incêndio, candeeiros e árvores, cenário de cidade;

- construção civil: transportes, trabalhadores de construção civil, sinalética, barreiras de proteção, montes de areia, postos de incêndio e candeeiros, cenário de cidade;

- médico: miniaturas de estetoscópio, espelho laríngeo, lupa, óculos, espátula, xarope, colher, caderno de registo, otoscópio, tesoura, lanterna de exploração, espelho de otorrinolaringologista, candeeiro de cirurgia;

- oficina: parafusos, porcas, chave de fendas, óculos de proteção, paquímetro, grampo de serralheiro, roquete, extensão para roquete, serrote de metal;

- casa: família, cama, almofadas, colchão, lençol, guarda-roupa, mesas-de-cabeceira, candeeiros, mesa, cadeiras, pratos, copos, garrafas, sanita, lavatório, chuveiro, fogão, forno, frigorífico, armário, máquina de lavar roupa, lava-loiça, aspirador;

- parque: menina, gato, rato, baloiço, cama de rede, carrinho de bebé, balde, pá, garrafa, garfo, faca, máquina fotográfica e saco;

- quinta: agricultor, trator, estábulo, capoeira, árvores, porcos, ovelhas, cavalos, coelhos, galo, patos, cão, cercas e cenário do solo.

Na tarefa de descrição de imagens solicitou-se à criança a criação de uma pequena história, para cada sequência. As sequências foram previamente ordenadas pela investigadora e apresentadas separadamente. De um total de seis sequências de imagens, a investigadora selecionou, aleatoriamente, duas sequências e efetuou o registo dessa escolha para cada criança, a fim de não ocorrer repetição no segundo momento da recolha, em que também foram utilizadas sequências de imagens. Na seleção das seis sequências de imagens a utilizar pretendia-se que os respetivos desenhos fossem de fácil interpretação para todas as crianças e representativos de situações familiares para as mesmas. Assim, da coleção “*LDA Language Cards - Sequential Thinking*” foram selecionadas as sequências em que personagens diferentes representam uma consulta médica, ida à pesca, acidente de viação e execução de um castelo de areia na praia. Da coleção “Obrigado, Kim!” utilizaram-se duas sequências, uma em que a personagem, Kim, confeciona biscoitos com a mãe e outra em que bebe sumo com uma menina. As sequências eram constituídas por cinco, seis ou oito imagens e podem ser observadas no Apêndice E e F.

O segundo momento de gravação de cada criança foi antecedido pela aplicação de um teste formal de linguagem, com o objetivo de se assegurar uma amostra com desenvolvimento típico. Uma vez que não é conhecido um teste de avaliação da linguagem oral, aferido para o português europeu, que se aplique simultaneamente às crianças com

idades compreendidas entre os 3 e os 7 anos, utilizaram-se o Teste de Avaliação da Linguagem na Criança (TALC) (Kay & Tavares, 2006) e a Grelha de Observação da Linguagem – Nível Escolar (GOL-E) (Kay & Santos, 2003).

O TALC foi desenvolvido para colmatar a lacuna existente na avaliação da linguagem da criança em idade pré-escolar. Deste modo, apresenta dados normativos para as crianças com idades compreendidas entre os 2 anos e 6 meses e 5 anos e 11 meses, avaliando a linguagem oral, nas suas componentes compreensiva (Parte I) e expressiva (Parte II). A primeira parte que o constitui avalia o vocabulário (identificação de objetos e imagens), relações semânticas (de duas e três palavras de conteúdo) e frases complexas. A segunda parte do teste contém os subtestes vocabulário (nomeação de objetos e imagens), frases absurdas, constituintes morfossintáticos e funções comunicativas.

A GOL-E avalia o domínio da linguagem oral das crianças que se encontram na fase final do nível pré-escolar e/ou a iniciar o nível escolar, mais concretamente aquelas que apresentam idade compreendida entre os 5 anos e 7 meses e os 10 anos. Possibilita a obtenção de dados relativamente à semântica, morfossintaxe e fonologia, através da aplicação de tarefas que avaliam a capacidade recetiva (apenas no âmbito da fonologia) e a expressiva. Este teste é composto por 3 subtestes de estrutura semântica (definição de palavras, nomeação de classes e opostos), 4 subtestes de estrutura morfossintática (reconhecimento de frases agramaticais, coordenação e subordinação de frases, ordem de palavras na frase e derivação de palavras), 4 subtestes de estrutura fonológica (discriminação de palavras, discriminação de pseudo-palavras, identificação de palavras que rimam e segmentação silábica).

Finda a avaliação das competências linguísticas das crianças, seguiu-se o segundo momento de recolha de dados. À semelhança do primeiro momento, e tal como referido anteriormente, a investigadora solicitou a cada criança a construção de narrativas para as quatro restantes sequências de imagens que não haviam sido selecionadas, aquando do primeiro momento de recolha. Para além desta, realizou-se outra atividade que possibilitou a recolha de amostras de discurso narrativo. A investigadora pediu à criança para folhear um livro, observando-o atentamente e em silêncio para, depois, construir oralmente uma história. Às crianças com idades compreendidas entre os 3 e os 5 anos foi apresentado o livro “A Festa de Anos do Tomás”. Neste livro, Tomás faz 5 anos e festeja o seu aniversário, em casa, com os amigos. Perante uma falha de energia elétrica, as crianças que até aí brincavam cada uma para o seu lado com diferentes equipamentos tecnológicos, necessitam de recorrer à imaginação para continuarem a divertir-se. As crianças de 6 e 7

anos visualizaram o livro “É Tudo Verdade ou mentiras do pequeno corvo para todas as ocasiões ”. Nesta história, o Pequeno Corvo, o Javali, o Urso e o Texugo viajam de comboio até à quinta da Avó Javali, onde vão passar férias. Contudo, decidem levar, em segredo, todos os seus amigos. Durante a estadia na quinta vão contando diversas mentiras de modo a conseguir alimentar e esconder os amigos, até que a verdade é descoberta. Ambos os livros fazem parte das recomendações do Plano Nacional de Leitura e foram selecionados após uma cuidada verificação dos temas e ilustrações. Toda a componente escrita dos livros foi ocultada com autocolante opaco para evitar que as crianças leitoras fossem influenciadas pela história original.

Nas atividades que pressupunham a construção de narrativas reduziram-se ao máximo as interrupções, por parte da investigadora. Esta interveio recorrendo a comentários e questões apenas quando a criança demonstrou dificuldades em iniciar e/ou desenvolver a narrativa.

O segundo momento da recolha de dados foi complementado com jogo simbólico, apenas nas situações em que a criança concluiu as tarefas supramencionadas num tempo inferior ao estabelecido para cada momento (30 minutos).

Para o processo de recolha, foram definidas estratégias a adotar pelas investigadoras, de forma a favorecer a interação com a criança e, assim, potenciar o número de produções verbais. Miller (1981) e Leadholm e Miller (1992) sugerem que, para iniciar e manter a interação comunicativa, o investigador deve:

- demonstrar entusiasmo através do uso de sorrisos, inflexão vocal e contacto ocular;
- saber esperar pelas intervenções da criança, aceitando as pausas e evitando interpelar constantemente a criança;
- ouvir e seguir o interesse da criança utilizando comentários e questões abertas, tais como: “Conta-me mais.”, “O que aconteceu a seguir?”, “E depois?”;
- reduzir ao mínimo o número de ordens e de questões fechadas;
- adaptar a sua linguagem ao nível de desenvolvimento linguístico de cada criança.
- interagir com a criança durante o jogo simbólico, procurando intervir verbalmente apenas nos primeiros minutos da brincadeira.

## II. 5. Tratamento e Análise dos dados

O processo de transcrição é utilizado para preservar e tornar o discurso da criança acessível a qualquer pessoa que o pretenda analisar. Esta deve representar fielmente o que a criança disse e, por conseguinte, quanto mais detalhada for, mais análises serão depois permitidas (Leadholm & Miller, 1992; Miller, 1981).

Todas as produções da criança e da investigadora foram transcritas com recurso a um programa de transcrição ortográfica de amostras de discurso, designado por CLAN. Este programa foi desenvolvido e é utilizado num sistema de arquivo internacional de amostras de fala espontânea de crianças, o Child Language Data Exchange System (CHILDES) e está disponível para download gratuito na seguinte página da internet: <http://childes.psy.cmu.edu>. A transcrição de dados realizada neste programa segue o formato CHAT e requer a utilização de critérios e códigos específicos. Tanto as instruções de instalação e uso do programa CLAN, como as convenções e princípios da transcrição, em formato CHAT, encontram-se descritos nos manuais originais (MacWhinney & University, 2012) ou em manuais adaptados, como é o caso da adaptação realizada para o português europeu por Pulido, Simões, Moreira, e Branco (2011).

Para além da edição de texto, o programa CLAN possibilita, simultaneamente, a análise linguística das amostras transcritas, nomeadamente no que concerne à DV, extensão média de enunciado e frequência de palavras.

Neste estudo, cada investigadora executou, individualmente, a transcrição ortográfica das 50 recolhas que efetuou. No total, obtiveram-se 100 ficheiros de transcrição (compilados e anexados num CD), dois por criança, correspondentes aos dois momentos de gravação, com a extensão *.cha*. A designação atribuída a cada ficheiro contém a abreviatura de transcrição - *trans* -, seguida do número atribuído à criança (1 a 50) e a letra A, quando a transcrição corresponde ao primeiro momento de gravação, ou letra B se corresponde ao segundo momento. Exemplificando, o ficheiro de transcrição do primeiro momento de gravação da criança número 25 contém a seguinte designação: *trans25A*. Todos os ficheiros apresentam a mesma estrutura de formatação: cabeçalho e transcrição (Apêndice G). O cabeçalho inclui informação relativa à criança, nomeadamente o nome (em iniciais), data de nascimento e idade cronológica. Contém também dados da recolha, tal como a data, local e atividades realizadas.

A transcrição propriamente dita é um processo muito moroso, requerendo um grande investimento de tempo. Demuth (1998) refere que no seu estudo uma hora de gravação correspondeu a aproximadamente sete horas de transcrição.

As transcrições foram efetuadas em português europeu, à luz do novo acordo ortográfico, e contêm, para além das linhas ortográficas, as linhas dependentes e independentes onde se descreveram todos os comportamentos ocorridos e considerados importantes na contextualização do que foi verbalizado.

No momento da transcrição, foram ignorados os primeiros 5 minutos de cada gravação, uma vez que o início da interação é considerado o momento mais difícil na recolha, pela natural inibição da criança perante uma pessoa que lhe é desconhecida, bem como pela própria situação de gravação, com a qual muitas crianças não estão familiarizadas (Araújo, 2007; Miller, 1981). Os 25 minutos restantes contêm, portanto, as produções da criança e da investigadora. A inclusão das produções do adulto que interage com a criança é importante, na medida em que estas fornecem informações relativamente ao contexto linguístico gravado (Demuth, 1998; Miller, 1981).

Tanto os critérios de segmentação como os de transcrição foram pré-definidos pelas investigadoras, com o objetivo de se garantir a maior consistência possível. Para o efeito, utilizou-se como referência a metodologia descrita e utilizada em alguns estudos realizados no âmbito das temáticas em questão. Foi, também, imprescindível a consulta dos manuais, originais e adaptado, do CLAN e formato CHAT. Neste processo tornou-se fundamental uma seleção cuidada e particularizada de critérios, atendendo às análises a concretizar. Importa referir que estes critérios foram utilizados apenas na transcrição das produções da criança.

Na segmentação de enunciados seguiram-se, entre outros, alguns dos critérios convencionados para o *software* Systematic Analysis of Language Transcripts (SALT) que são, contudo, compatíveis com a formatação do tipo CHAT. Este *software*, à semelhança do CLAN, é utilizado para conduzir os processos de elicitação, transcrição e análise de amostras de discurso.

Assim, na segmentação do discurso da criança, considerou-se que cada enunciado correspondia a uma unidade comunicacional, designada por *C-unit*. Uma *C-unit* consiste numa oração independente e respetivos modificadores, ou seja, inclui uma oração principal e todas as subordinadas que dela dependem (Miller, Andriacchi, & Nockerts, 2012). Para além das orações subordinadas (1), as orações coordenadas complexas (2), ligadas por uma

conjunção, corresponderam a um enunciado. Porém, sempre que ocorreram várias orações coordenadas ligadas pela conjunção “e” efetuou-se a segmentação, de modo a que a terceira oração fosse contabilizada como outro enunciado, tal como verificado no estudo de Rice *et al.* (2010). As orações simples (3) constituíram também um único enunciado. Apresentam-se como exemplos os seguintes enunciados:

(1) \*CHI: e o menino, quando acorda, repara que está no hospital.

(2) \*CHI: e começou a escrever num papel e deu ao pai.

(3) \*CHI: estes já estão bons.

Sempre que foi introduzido discurso direto pela criança, sem a utilização de uma oração subordinada completiva, optou-se pela segmentação do enunciado, tal como nos exemplos (4) e (5).

(4) \*CHI: e o coelho disse.

\*CHI: podes!

(5) \*CHI: o menino disse que o pai foi trabalhar.

Considerando que, por vezes, poderiam surgir incertezas na tarefa de determinar as fronteiras dos enunciados, recorrendo aos critérios anteriores, em caso de indecisão determinou-se que seriam considerados como terminações de enunciados os contornos entoacionais do discurso, assim como as pausas, quando superiores a 2 segundos (Leadholm & Miller, 1992; Miller, 1981). De acordo com Miller (1981), 80% das pausas superiores a dois segundos ocorrem entre enunciados.

As respostas elíticas e de sim/não (6) foram considerados enunciados dependentes, tal como as interrogativas-tag (*eg.*: “..., pois não?”, “..., não foi?”, “..., não é?”), as estruturas parentéticas (*eg.*: “penso eu”, “acho eu”) e os marcadores de discurso (*eg.*: “olha,...”) (7).

(6) \*CHI: não, eu não preciso.

(7) \*CHI: olha, este também já podia sair da prisão.

Contudo, mantiveram-se como enunciados independentes todas as repetições exatas dos mesmos, apenas quando a criança os utilizou de forma intencional, para enfatizar (8).

(8) \*CHI: está aqui um!

\*CHI: está aqui um!

As contagens, sequências ou outras enumerações (9) foram registadas como parte integrante dos enunciados e consideradas para a análise do RTT e do índice D. Porém, foram assinaladas de modo a poderem ser excluídas noutras análises, tal como a realizada no estudo da investigadora Ana Rita Pina Coimbra de Campos, onde se pretendia determinar a extensão média de enunciado (EME) (Leadholm & Miller, 1992; Lund & Duchan, 1983; Miller, 1981).

(9) \*CHI: era uma vez o porco, a toupeira, o cão e o pássaro e o porco.

Cada enunciado produzido pela criança foi transcrito numa linha ortográfica distinta, iniciado com letra minúscula e finalizado com um sinal de pontuação, de acordo com o tipo de frase e/ou entoação melódica. A codificação utilizada foi criteriosamente selecionada tendo sempre como referência o que se pretendia incluir ou excluir posteriormente em cada análise linguística e pode ser consultada no Apêndice H. Deste modo, e retomando o exposto no capítulo I.1.4.2., para o estudo em questão definiu-se que todas as palavras seriam consideradas diferentes (*types*) desde que assumissem formas distintas e, por isso, transmitissem diferentes valores gramaticais, fossem eles de número, género, pessoa ou tempo.

O nome da criança transcreveu-se utilizando sempre as iniciais em maiúsculas, separadas por um espaço; para siglas utilizou-se apenas a letra inicial em maiúscula (10).

(10) \*CHI: jogar na Nba.

Os pronomes clíticos foram separados do verbo por um espaço ao invés do hífen, uma vez que alguns dos comandos do programa CLAN estão configurados para excluir o que segue imediatamente esse sinal gráfico (11). O programa CLAN está também formatado para excluir a representação gráfica “um”, que no caso do português corresponde a um numeral, pelo que esta palavra foi representada pela forma “umm”(12) e os elementos das palavras compostas por justaposição foram ligados com o sinal gráfico *underscore* (“\_”), de forma a serem consideradas como uma única palavra (13) (Basílio, 2004; Miller, 1981).

(11) \*CHI: e esse chama se João.

(12) \*CHI: tenho, mas agora vou fazer lá umm palco.

(13) \*CHI: precisamos de fazer a casa\_de\_banho aqui.

As disfluências, interjeições, onomatopeias que não têm funcionamento de palavra, palavras incompletas, palavras impercetíveis (14) e enunciados considerados supérfluos, tais como canto, leitura e repetições exatas de enunciados proferidos previamente pela investigadora, foram codificados de modo a serem excluídos da análise (Miller, 1981).

(14) \*CHI: são xxx coisas de ferro.

As repetições exatas ou parciais e as reformulações foram assinaladas de modo a serem contabilizadas apenas as palavras ou expressões contidas na repetição/reformulação e excluídas as que foram repetidas/reformuladas (15). Tal como efetuado no estudo de Demuth (1998) e sugerido por Miller (1981), apenas nas situações em que as palavras ou expressões foram utilizadas repetidamente para enfatizar se contabilizaram, então, todas as ocorrências.

(15) \*CHI: daqui [/] daqui [/ -] na próxima segunda-feira se eu me portar bem vou estar no quadro de honra.

Nas produções desviantes, utilizou-se codificação específica que possibilitasse a recuperação da produção-alvo, nomeadamente quando se verificou a omissão de palavras (16) ou apenas de segmentos fónicos e quando foi possível identificar a palavra distorcida. As contrações de palavras resultantes de processos de coarticulação e as omissões resultantes de processos fonológicos foram assinaladas de forma a serem contabilizadas as palavras-alvo (17).

(16) \*CHI: por isso 0é que elas não a deixavam sair da casa.

(17) \*CHI: prá[s] [: para as] fagulhas não irem pró[s] [: para os] olhos.

Todos os enunciados incompletos foram sinalizados considerando situações específicas. A codificação utilizada permitiu diferenciar enunciados incompletos retomados (18) de enunciados incompletos não retomados (19), assim como identificar o sujeito responsável pela interrupção. Todos os enunciados incompletos, exceto os retomados, foram codificados com “[+esc]” para serem excluídos da análise da EME. Os enunciados com desvios sintáticos e morfofonológicos foram também assinalados com [+ bch], de modo a possibilitar, caso se justifique, uma análise posterior (20).

(18) \*CHI: <rebuçados na &me> [//] +/.

\*INV: convites.

\*CHI: +, rebuçados na mesa, convites, batatas fritas. [+ esc]

(19) \*CHI: e depois o pai quando chegou +... [+ esc]

(20) \*CHI: e depois a soldar para pôr os ferros não caem. [+ bch]

Atendendo ao que se pretendia analisar posteriormente, o corpo da transcrição foi ainda segmentado de acordo com o tipo de discurso utilizado. Assim, demarcou-se o início dos contextos narrativo e espontâneo (21).

(21) @G: narrativo.

\*CHI: era uma vez uns animais que foram se embora para ir para a escola.

Definidos todos os critérios, testou-se a consistência das transcrições entre investigadoras. Para o efeito, ambas as investigadoras efetuaram, em simultâneo, a transcrição das mesmas gravações selecionadas aleatoriamente, em excertos de 10 minutos; estes segmentos não devem ser inferiores a 10% do total da amostra (Demuth, 1998). Após a transcrição de cada segmento, foram comparadas as duas transcrições e discutidas todas as disparidades, tal como sugerido por Miller (1981). De modo a atingir uma percentagem de acordo igual ou superior a 90%, determinada pelo cálculo dos índices RTT, D e EME, foi necessária a realização de 5 testes. De acordo com Riffe, Lacy e Fico (1998), citados por Rourke, Anderson, Garrison, e Archer (2001), nos estudos realizados no âmbito da comunicação humana deve ter-se como referência uma percentagem mínima de acordo inter-investigadores de 80%.

Seguiu-se, então, a transcrição de todas as gravações efetuadas. Após a conclusão da transcrição de cada momento de gravação, competia a cada investigadora a verificação do ficheiro, executando o comando CHECK no programa CLAN. Este comando é extremamente importante, na medida em que permite identificar erros de formatação e de codificação e assim garantir uma transcrição mais rigorosa (MacWhinney & University, 2012).

Finalizado todo o processo de segmentação e transcrição, procedeu-se ao cálculo dos índices D e RTT, através dos comandos dos programas *vocd* e *freq*, respetivamente, do programa CLAN. As linhas de comandos utilizadas para estes programas podem variar no que diz respeito à sua composição e, assim, influenciar a inclusão/exclusão de determinadas palavras e a contabilização de palavras diferentes. As linhas de comando utilizadas neste estudo encontram-se representadas nos seguintes exemplos:

*freq +t\*CHI trans50A.cha*

*freq* – programa que executa a análise do RTT.

+*t\*CHI* – limita a análise às linhas ortográficas correspondentes à criança.

*trans50A.cha* – nome do ficheiro a analisar.

*vocd +t\*CHI +r6 +k trans13B.cha*

*vocd* – programa que executa a análise do D.

+*t\*CHI* – limita a análise às linhas ortográficas correspondentes à criança.

+*r6* – exclui todo o material repetido.

+*k* – contabiliza como palavras diferentes as palavras homónimas que se distinguem graficamente apenas na letra inicial (maiúscula ou minúscula). *Eg.*: campos (nome comum) e Campos (nome próprio).

Na janela de output do programa *freq* foi possível visualizar, para cada ficheiro, o total de *types* e *tokens*, assim como o resultado do RTT. Para além disso, foi aí consultada a lista de todas as formas transcritas que foram consideradas palavras e palavras diferentes. Estes dados revelaram-se extremamente vantajosos, na medida em que permitiram a identificação de erros impossíveis de detetar através do comando CHECK e assegurar o cumprimento do critério de inclusão respeitante ao número de *tokens* (50, no mínimo). Importa referir que o número de enunciados, considerado neste estudo também como fator de inclusão, foi verificado nesta fase, contudo, pela investigadora Ana Rita Pina Coimbra de Campos através do comando do programa *mlu*, utilizado no seu estudo para o cálculo da EME.

Na janela de output do programa *vocd*, obtida para cada ficheiro, constou a lista de todos os enunciados produzidos pela criança e três tabelas que continham os valores necessários ao cálculo do D e que correspondiam às três amostras aleatórias de *tokens* analisadas. Cada tabela era acompanhada pelo valor médio do índice e respetivo desvio-padrão, assim como o valor a que correspondia o melhor ajuste designado por *D\_optimum*. Da média dos três valores ótimos resultou o valor final do D.

Os procedimentos supracitados conduziram à obtenção de dois resultados de RTT e D (respeitantes aos dois momentos de gravação) por criança e foram registados numa tabela, para posterior análise estatística.

## CAPÍTULO III: APRESENTAÇÃO DOS RESULTADOS

### III. 1. Descrição dos resultados

Neste capítulo apresentam-se os resultados obtidos no tratamento informático dos dados, em função das hipóteses definidas anteriormente, com suporte em tabelas que evidenciam os dados mais relevantes para uma melhor interpretação.

A base de dados foi tratada estatisticamente com recurso ao *software* Statistical Package for the Social Sciences (SPSS), versão 20 para Windows. Numa primeira fase procurou-se descrever a amostra através de uma análise estatística descritiva. Assim, a primeira abordagem aos dados teve como objetivo a definição dos valores médios e respetivos desvios-padrão dos índices RTT e D totais e obtidos em cada contexto discursivo (espontâneo e narrativo), para cada faixa etária.

Seguidamente apresentam-se os resultados obtidos nos processos estatísticos inferenciais realizados, de modo a testar as hipóteses definidas. A estatística inferencial possibilita, através de técnicas de estimação e de testes de hipóteses, a obtenção de conclusões para uma determinada amostra (Laureano, 2011). Assim, numa segunda fase, foi investigada a existência de diferenças estatisticamente significativas<sup>2</sup> nos valores médios totais de RTT e D entre as diferentes faixas etárias, níveis de escolaridade e contextos discursivos, bem como a existência de correlações entre os índices, idade em meses e desenvolvimento linguístico.

De modo a determinar a existência de diferenças nos valores médios dos índices D e RTT, entre os grupos etários, tal como definido na hipótese 1, procedeu-se à realização do teste paramétrico Análise de Variância a um fator (“one-way ANOVA”) e ao teste não paramétrico Kruskal-Wallis, respetivamente.

Para verificar se a DV difere entre os níveis de escolaridade pré-escolar e escolar (hipótese 2), realizou-se o teste paramétrico para duas amostras independentes, teste T, no que concerne ao D. Já em relação ao RTT optou-se pelo teste não-paramétrico para duas amostras independentes, Mann-Whitney.

---

<sup>2</sup>Sempre que as variáveis fossem rácio e as amostras apresentassem  $N \geq 30$ , pelo Teorema do Limite Central, assumiu-se a normalidade das distribuições. Quando  $N < 30$ , testou-se a normalidade através do teste de adesão à normalidade Shapiro-Wilk. Podendo assumir-se a normalidade da distribuição da amostra, recorreu-se a testes paramétricos (e.g., ANOVA). Verificando-se o oposto, optou-se pelos equivalentes não-paramétrico (e.g., Kruskal-Wallis).

Na análise que permitiu testar a hipótese 3, foi comparada a DV nos contextos discursivos narrativo e espontâneo. Esta hipótese foi testada em concomitância com as hipóteses 4 e 5, onde se pretendia verificar a existência de diferenças na DV das crianças, nos contextos discursivos mencionados, segundo a faixa etária. Para esse fim, recorreu-se ao método estatístico ANOVA para medidas mistas e depois ao teste *post-hoc* de Tukey com o objetivo de se identificar as diferenças entre os subgrupos.

A última análise estatística realizada, relativa à hipótese 6, consistiu na identificação de correlações entre índices, idade, em meses, e desenvolvimento linguístico, representado, neste estudo, pelos resultados obtidos pelas crianças em idade pré-escolar no TALC, nas tarefas expressivas, e pelos resultados obtidos pelas crianças em idade escolar na GOL-E, nas provas da componente semântica. Procedeu-se inicialmente à verificação de correlação apenas entre índices e idade, em meses, através do coeficiente de correlação linear de Pearson, que pôde ser calculado dado todas as variáveis serem quantitativas. Seguidamente, explorou-se a existência de correlações entre as quatro variáveis mencionadas anteriormente- RTT, D, idade e desenvolvimento linguístico, também pela realização do teste ao coeficiente de correlação linear de Pearson.

Importa referir que, neste estudo, as inferências realizadas a partir dos testes de hipótese seguiram a teoria da decisão, admitindo um erro máximo de 5%. Assim, o nível de significância utilizado foi de 0.05<sup>3</sup>, uma vez que é considerado o mais usual (Laureano, 2011).<sup>3</sup>

### **III. 2. Valores médios e desvios-padrão de RTT e D, segundo faixa etária**

Definiu-se como análise inicial a identificação da média, desvios-padrão, máximos e mínimos de todas as medidas da amostra. Os valores encontram-se representados na Tabela 3.

---

<sup>3</sup>Foram ainda considerados valores de significância a variar entre 0.05 e 0.1, considerados “marginalmente significativos”.

**Tabela 3** – Média, desvio-padrão, valores mínimos e máximos dos índices RTT e D total, espontâneo e narrativo, da amostra.

|                | <b>n</b> | <b>Média</b> | <b>Desvio-padrão</b> | <b>Mínimo</b> | <b>Máximo</b> |
|----------------|----------|--------------|----------------------|---------------|---------------|
| RTT total      | 50       | 0.260        | 0.304                | 0.200         | 0.369         |
| RTT narrativo  | 50       | 0.340        | 0.446                | 0.263         | 0.447         |
| RTT espontâneo | 50       | 0.304        | 0.451                | 0.226         | 0.443         |
| D total        | 50       | 83.080       | 14.184               | 60.300        | 120.060       |
| D narrativo    | 50       | 58.947       | 15.931               | 35.570        | 111.490       |
| D espontâneo   | 50       | 82.688       | 14.756               | 111.490       | 114.760       |

As mesmas medidas foram depois calculadas para cada faixa etária e apresentam-se no decorrer das análises que se seguem.

### III. 3. Comparação dos desempenhos no D e RTT, segundo as faixas etárias

Posteriormente, procedeu-se à realização de inferências estatísticas a fim de se determinar a presença de diferenças estatisticamente significativas entre faixas etárias (variável independente), no que concerne aos dois índices (variáveis dependentes), tal como definido em **H1**.

Tendo-se observado uma violação ao pressuposto da normalidade, relativamente ao RTT<sup>4</sup>, recorreu-se, então, ao teste não paramétrico Kruskal-Wallis que indicou que não se verificam diferenças estatisticamente significativas, nos valores de RTT, entre os grupos etários,  $H(4) = 3.539$ ,  $p = 0.472$ , representados na Tabela 4. Assim, considera-se que, nesta amostra, os desempenhos no RTT são equivalentes entre os diversos grupos etários.

Para testar o D, e verificada a normalidade das distribuições das amostras<sup>5</sup>, aplicou-se o teste estatístico ANOVA a um fator que indicou a existência de diferenças significativas entre faixas etárias,  $F(4,45)=3.916$ ,  $p=.008$ .

Compararam-se os grupos etários através do teste *post-hoc* de Tukey para identificar quais as que apresentavam diferenças entre si. Este teste indicou que as únicas diferenças estatisticamente significativas ocorrem entre a faixa dos [3;00-3;11] ( $M=72.994$ ,  $DP=11.339$ ) e dos [6;00-6;11] ( $M=91.464$ ,  $DP=14.178$ ),  $p=.018$  e entre a faixa dos [3;00-

<sup>4</sup> $SW(10) \leq 0.963$ ,  $p' s \geq .026$

<sup>5</sup> $SW(10) \leq .951$ ,  $p' s \geq 0.161$

3;11] e dos [7;00-7;11] ( $M=91.014$ ,  $DP=16.364$ ),  $p=.02$  (para todas as outras comparações,  $p's \geq .05$ ), tal como se pode observar na Tabela 4.

**Tabela 4** – Médias e desvios-padrão dos valores dos índices RTT e D, para cada faixa etária.

| Faixa Etária | n  | Rácio Type-Token Total |               | D Total                |               |
|--------------|----|------------------------|---------------|------------------------|---------------|
|              |    | Média                  | Desvio-padrão | Média                  | Desvio-padrão |
| [3;00-3;11]  | 10 | 0.244                  | 0.009         | 72.994 <sup>a</sup>    | 3.586         |
| [4;00-4;11]  | 10 | 0.261                  | 0.007         | 80.273 <sup>a, b</sup> | 2.300         |
| [5;00-5;11]  | 10 | 0.267                  | 0.013         | 79.656 <sup>a, b</sup> | 4.029         |
| [6;00-6;11]  | 10 | 0.260                  | 0.010         | 91.464 <sup>b</sup>    | 4.483         |
| [7;00-7;11]  | 10 | 0.266                  | 0.007         | 91.014 <sup>b</sup>    | 5.175         |

*Nota.* Valores com letra distinta, em coluna, diferem a  $p < .05$ .

Assim, no que diz respeito ao D, as crianças de 3 anos mostraram desempenhos estatisticamente inferiores às crianças de 6 e de 7 anos, sendo os desempenhos equivalentes nos restantes grupos etários.

### III. 4. Comparação dos desempenhos no D e RTT, segundo os níveis de escolaridade

Pretendia-se, também, identificar diferenças nos valores médios dos índices RTT e D entre os níveis de escolaridade pré-escolar e escolar, representados na Tabela 5, de acordo com **H2**.

Relativamente ao índice RTT, verificou-se a violação do pressuposto da normalidade<sup>6</sup>, por isso, recorreu-se ao teste não paramétrico equivalente ao teste  $T$ , o teste de Mann-Whitney, sendo o RTT a variável dependente e o nível de escolaridade o fator. Este teste permitiu observar a existência de diferenças estatisticamente significativas entre o pré-escolar e escolar ( $U = 169.000$ ,  $p=.048$ ). Verificou-se que o desempenho nas crianças do pré-escolar foi inferior ao desempenho das crianças do escolar, concluindo-se, portanto, que o nível de escolaridade influencia os valores médios de DV, para o RTT.

<sup>6</sup> Nível pré-escolar:  $SW(35) \leq 0.921$ ,  $p' \geq .015$ ; nível escolar:  $SW(15) \leq .956$ ,  $p' \geq 0.629$

Verificada a normalidade da distribuição para o D, em ambos os níveis<sup>7</sup>, seguiu-se a realização do teste T, considerando o D como variável dependente e o nível de escolaridade como fator. O resultado deste teste revelou que a média dos valores de D no nível pré-escolar é diferente da média dos valores de D no nível escolar  $t(48) = -4.248, p < .001$ . Conclui-se, assim, que, em média, as crianças do pré-escolar apresentaram um desempenho inferior no D ( $M=78.275, DP=10.699$ ) comparativamente com as crianças do escolar ( $M=94.292, DP=15.291$ ).

**Tabela 5** – Médias e desvios-padrão dos índices RTT e D obtidos para cada nível de escolaridade, dimensões de cada amostra e significância obtida nos testes de hipótese.

|                       |             | RTT                                |               | D                        |               |
|-----------------------|-------------|------------------------------------|---------------|--------------------------|---------------|
|                       |             | Média                              | Desvio-padrão | Média                    | Desvio-padrão |
| Nível de Escolaridade | Pré-escolar | 0.255                              | 0.317         | 78.275                   | 10.699        |
|                       | Escolar     | 0.270                              | 0.025         | 94.292                   | 15.291        |
|                       |             | $p = .048$<br>(teste Mann-Whitney) |               | $p < 0.001$<br>(teste T) |               |

### III. 5. Comparação dos desempenhos no D e RTT, segundo os contextos discursivos

Para averiguar a existência de diferenças nos índices, consoante o contexto discursivo, narrativo e espontâneo (fatores repetidos, ou intra-sujeitos), tal como estabelecido em **H3**, mas também segundo a faixa etária (fator independente ou inter-sujeitos), de acordo com **H4** e **H5**, realizou-se o teste paramétrico ANOVA para medidas mistas.

Relativamente ao RTT, os resultados obtidos indicam que são estatisticamente significativas as diferenças observadas entre o RTT narrativo e o RTT espontâneo ( $M_{RTT \text{ espontâneo}} = .304, DP=.045; M_{RTT \text{ narrativo}} = .340, DP=.045$ ),  $F(1,45)=20.376, p < 0.001$ . Conclui-se, por conseguinte, que para este índice, o desempenho das crianças em contexto espontâneo é, no seu todo, inferior ao desempenho das mesmas crianças em contexto narrativo. Para além disso, verificou-se que o desempenho nos dois contextos discursivos (RTT narrativo e RTT espontâneo), pelo teste *post-hoc* Tuckey, é diferente consoante a faixa etária

<sup>7</sup> Nível pré-escolar:  $SW(35) \leq 0.956, p \leq 0.169$ ; nível escolar:  $SW(15) \leq .938, p \leq 0.360$ .

$F(4,45)=2.716, p=.041$  (tabela 6). A realização deste teste indicou que, para o índice RTT, o desempenho das crianças com 6 anos é superior no contexto narrativo ( $M = .360, DP = .013$ ), comparativamente com o contexto espontâneo ( $M=.288, DP=.012$ ),  $p=.008$ .

**Tabela 6** – Média e desvio-padrão dos valores do índice RTT em contexto espontâneo e narrativo, para cada faixa etária.

| Faixa Etária | N  | Rácio Type-Token Espontâneo |               | Rácio Type-Token Narrativo |               |
|--------------|----|-----------------------------|---------------|----------------------------|---------------|
|              |    | Média                       | Desvio-padrão | Média                      | Desvio-padrão |
| [3;00-3;11]  | 10 | 0.286a                      | .014          | 0.333a                     | .013          |
| [4;00-4;11]  | 10 | 0.299a                      | .011          | 0.348a                     | .015          |
| [5;00-5;11]  | 10 | 0.325a                      | .017          | 0.323a                     | .013          |
| [6;00-6;11]  | 10 | 0.288a                      | .012          | 0.360b                     | .013          |
| [7;00-7;11]  | 10 | 0.321a                      | .015          | 0.335a                     | .015          |

*Nota.* Valores com letra distinta, em linha, diferem a  $p < .05$ .

Quanto ao D verificaram-se diferenças significativas entre o desempenho no contexto narrativo e o contexto espontâneo ( $M_{D_{espontâneo}}=82.689, DP=14.756; M_{D_{narrativo}}=58.947, DP=15.931$ ),  $F(1,45)=101.214, p<0.001$ , pelo que se conclui que, ao contrário do RTT, o desempenho das crianças em contexto espontâneo é, no seu todo, superior ao desempenho das mesmas crianças em contexto narrativo. Verificou-se, ainda, um efeito de interação marginalmente significativo entre os contextos espontâneo e narrativo e a faixa etária,  $F(4,45)=2.295, p=0.074$ . Com a realização de testes *post-hoc* e analisando a Tabela 7, conclui-se que, na faixa dos 3 anos, o desempenho das crianças em contexto espontâneo ( $M = 57.301, DP = 4.138$ ) é inferior ao desempenho das crianças de 6 e 7 anos, no mesmo contexto, ( $M = 63.249, DP = 6.598$ ),  $p = .005$  e ( $M = 66.891, DP = 4.281$ ),  $p = .003$ . As crianças que se encontram na faixa etária dos 6 anos apresentam um desempenho superior em contexto espontâneo comparativamente com o desempenho em contexto narrativo ( $M=63.249, DP=6.598$ ),  $p<.001$ . Constata-se, também, que na faixa dos 7 anos, o desempenho das crianças em contexto espontâneo é superior ao desempenho das mesmas crianças em contexto narrativo ( $M=92.729, DP=4.094$ ),  $p<.001$ .

**Tabela 7** – Média e desvio-padrão dos valores do índice D obtidos em contexto espontâneo e narrativo, para cada faixa etária.

| Faixa Etária | n  | D Espontâneo |               | D Narrativo |               |
|--------------|----|--------------|---------------|-------------|---------------|
|              |    | Média        | Desvio-padrão | Média       | Desvio-padrão |
| [3;00-3;11]  | 10 | 57.301a      | 4.138         | 66.891a     | 4.280         |
| [4;00-4;11]  | 10 | 54.115a      | 2.689         | 79.822a     | 2.725         |
| [5;00-5;11]  | 10 | 54.039a      | 4.749         | 82.124a     | 4.235         |
| [6;00-6;11]  | 10 | 63.249a      | 6.598         | 91.878b     | 2.973         |
| [7;00-7;11]  | 10 | 66.891a      | 4.281         | 92.729b     | 4.094         |

*Nota.* Valores com letra diferente, em linha, diferem a  $p < .05$ .

### III. 6. Correlação entre RTT, D, idade e desenvolvimento linguístico

Finalmente procurou-se verificar qual das variáveis, RTT, D (H6.2) e idade (H6.1), em meses, seria o melhor indicador de desenvolvimento linguístico, representado, neste estudo, pelos resultados obtidos pelas crianças em idade pré-escolar nas tarefas expressivas do TALC, e pelos resultados obtidos pelas crianças em idade escolar nas provas da componente semântica da GOL-E. Uma vez que o teste de desenvolvimento linguístico aplicado às crianças em idade pré-escolar se distinguiu do teste aplicado às crianças em idade escolar, constituíram-se duas amostras, uma que compreendia as crianças dos 3 aos 5 anos ( $n = 30$ ) e a segunda as crianças de 6 e 7 anos ( $n = 20$ ). Aplicou-se o teste de normalidade e verificou-se que esse pressuposto foi cumprido em ambas<sup>8</sup>, pelo que, para determinar a existência de correlação entre as variáveis, se utilizou o teste ao coeficiente de correlação linear de Pearson, cujos coeficientes se encontram na Tabela 8.

Este teste indicou que nas crianças mais novas os índices não estão correlacionados significativamente ( $r = -.012$ ,  $p = .952$ ) e que apenas o índice D se correlaciona positiva e significativamente com o desenvolvimento linguístico ( $r = 0.425$ ,  $p = .019$ ). Também nesta amostra se verifica que a idade se relaciona positivamente com o desenvolvimento linguístico,  $r = 0.750$ ,  $p < .001$ . Nas crianças mais velhas os índices estão correlacionados positivamente ( $r = 0.669$ ,  $p = 0.01$ ) e verifica-se, também, que o índice RTT se relaciona positivamente com o desenvolvimento linguístico ( $r = 0.465$ ,  $p = .039$ ), estando também esta última variável correlacionada com o índice D ( $r = .721$ ,  $p < .001$ ). Também a idade e o

<sup>8</sup> Amostra das crianças mais novas:  $SW(30) \leq 0.952$ ,  $p$ 's  $\leq .194$ ; Amostra das crianças mais velhas:  $SW(20) \leq .968$ ,  $p$ 's  $\leq 0.705$ .

desenvolvimento linguístico se encontram relacionados positivamente:  $r = 0.521$ ,  $p = .018$ . Conclui-se, portanto, que nas faixas dos 6 e 7 anos um bom desempenho no RTT indica um bom desempenho no D.

Em suma, constata-se que tanto o D como a idade são as variáveis que mais se correlacionam com o desenvolvimento linguístico.

**Tabela 8** – Coeficientes de correlação linear de Pearson entre os índices RTT e D, idade e desenvolvimento linguístico.

|                             | RTT | D         | Idade em meses     | Desenvolvimento Linguístico |
|-----------------------------|-----|-----------|--------------------|-----------------------------|
| RTT                         | -   | -.012a)   | 0.318a)            | .088a)                      |
| D                           |     | 0.669b)** | 0.231b)<br>0.261a) | 0.465b)*<br>0.425a)*        |
| Idade em meses              |     |           | 0.321b)<br>-       | 0.721b)***<br>0.750a)***    |
| Desenvolvimento Linguístico |     |           |                    | 0.521b)*<br>-               |

\*  $p < .05$  \*\*  $p < .01$  \*\*\*  $p < .001$

*Nota.* a) Coeficiente de correlação linear de Pearson para a amostra das crianças com 3, 4 e 5 anos ( $n = 30$ ); b) Coeficiente de correlação linear de Pearson para a amostra das crianças com 6 e 7 anos ( $n = 20$ )

## CAPÍTULO IV: DISCUSSÃO DOS RESULTADOS

### IV. 1. Discussão dos Resultados

Os estudos realizados até à data, no âmbito da DV, são consensuais quanto à importância deste parâmetro, não só na compreensão do desenvolvimento linguístico, mas também na identificação de variações da performance vocabular em indivíduos com alterações linguísticas específicas. As investigações realizadas têm progredido no sentido da descoberta de medidas válidas e fiáveis que permitam calcular a DV e, assim, determinar valores que possam ser utilizados como indicadores dos diferentes níveis de desenvolvimento linguístico. Contudo, tal como sublinhado nos capítulos anteriores, nos estudos realizados até ao momento, que não incluem o português europeu, ainda não existe unanimidade relativamente ao processo que confere resultados mais credíveis.

No presente capítulo, pretende-se analisar detalhadamente os resultados obtidos confrontando-os com as hipóteses definidas para as questões de investigação, indo, deste modo, ao encontro do principal propósito deste trabalho que se prende com a identificação de uma medida capaz de representar a DV, bem como com a obtenção de valores de referência para o português europeu.

**H1: As crianças mais velhas apresentam um desempenho de RTT e de D superior ao verificado nas crianças mais novas.**

Seguindo o exposto no enquadramento teórico deste trabalho, segundo Malvern *et al.* (2004) as crianças vão aumentando a sua DV, com o avançar da idade. No que concerne aos valores de RTT, existe aparentemente um aumento nos valores médios entre as faixas etárias dos 3 e 5 anos e entre os 6 e 7 anos. Contudo, através da análise estatística observou-se que as diferenças encontradas não são estatisticamente significativas, pelo que se constata que os desempenhos no RTT são equivalentes entre os diversos grupos etários. Estes resultados não confirmam as conclusões de Scherer *et al.* (2002) que, no seu estudo, referem um aumento nos valores de RTT até aos 4 anos e uma diminuição na passagem dos 4 para os 5 anos de idade.

Admite-se que os resultados obtidos podem evidenciar a influência que as dimensões da amostra exercem nos valores de RTT que, neste estudo, foram definidas pelo

intervalo de tempo, ao contrário do proposto por diversos autores que sugerem a padronização das dimensões da amostra pelo número de *tokens*.

Por sua vez, no que diz respeito ao D, verificaram-se diferenças nos valores desse índice entre alguns grupos etários. As crianças com 3 anos de idade obtiveram valores de D inferiores aos obtidos pelas crianças de 6 e de 7 anos. Nos restantes grupos etários não foram observadas diferenças significativas. Assim se conclui que a DV, medida pelo índice D, é superior nas crianças mais velhas (com 6 e 7 anos), comparativamente com as crianças mais novas (3 anos). Estes resultados vão, em parte, ao encontro daquilo que foi constatado por Owen e Leonard (2002). Segundo estes autores, o índice D é capaz de refletir diferenças desenvolvimentais, pelo que a crianças mais velhas correspondem valores mais elevados de D e a crianças mais novas valores mais reduzidos.

À semelhança do que se constatou para o RTT, também os resultados obtidos para o índice D podem ser explicados pela influência das dimensões da amostra, tal como referido por Owen e Leonard (2002). Mais uma vez se constata que a escolha do intervalo de tempo para definir a amostra não terá sido a melhor opção, uma vez que se tem demonstrado noutros estudos, em que a amostra foi limitada pelo número de *tokens*, que o índice D se relaciona com a idade. À luz do que foi observado por Malvern *et al.* (2004) (Capítulo I. 1. 4. 5.), também o critério utilizado neste estudo, para a classificação de *types*, poderá ter influenciado os valores de D. Com o uso de formas flexionadas, o cálculo da DV dependerá manifestamente do desenvolvimento morfológico, estando, portanto, envolvido um maior número de *types*. A classificação de *types* de acordo com os lexemas será, porventura, a mais apropriada, uma vez que estará fundamentalmente associada ao desenvolvimento vocabular, ignorando a influência das questões morfológicas

Desta forma, os valores médios obtidos para ambos os índices não deverão ser utilizados como referência para o português europeu.

**H2: O desempenho no RTT e D das crianças em idade pré-escolar é inferior ao verificado nas crianças em idade escolar.**

A fim de se verificar a existência de diferenças vocabulares entre os níveis escolar e pré-escolar, analisaram-se os desempenhos nos índices RTT e D. Ambos os índices foram reveladores dessa diferença.

Na comparação dos dois níveis, verificou-se que as 35 crianças que frequentavam o ensino pré-escolar obtiveram um desempenho inferior ao verificado pelas 15 crianças que frequentavam o 1.º ano de escolaridade, em ambos os índices. Em média, as crianças do pré-escolar apresentaram um desempenho no RTT' de 0.255 ( $DP=0.317$ ) e de D de 78.275 ( $DP = 10.699$ ) e, por sua vez, as crianças do escolar obtiveram um valor médio de RTT' de 0.270 ( $DP=0.025$ ) e de D de 94.292 ( $DP = 15.291$ ).

Os presentes resultados suportam o parecer de Duarte *et al.* (2011), uma vez que indicam que tanto o índice RTT', como o D são sensíveis ao processo de escolarização, assinalando uma evolução na DV nas crianças, na passagem do ensino pré-escolar para o escolar. Esta evolução na DV estará, certamente, associada à aprendizagem de conceitos mais específicos abordados nas unidades curriculares inerentes ao nível escolar.

**H3: As crianças apresentam uma DV maior em contexto narrativo, comparativamente com o contexto espontâneo.**

A principal motivação que conduziu à análise da DV em diferentes situações discursivas (conversação, descrição de imagens e manipulação de objetos) foi verificar se as crianças fazem um uso diferenciado da linguagem, do vocabulário em concreto, tal como defendem Yont *et al.* (2003). Também Fergadiotis, Wright, e Capilouto (2011) referem que a DV, medida pelo índice D, é influenciada pelo tipo de discurso.

Para tal, procedeu-se ao cálculo e comparação dos valores médios de RTT' e D, em contexto narrativo e contexto espontâneo. O contexto narrativo correspondeu, tal como referido na metodologia deste trabalho, a tarefas de descrição de sequências de imagens e à construção de narrativa a partir das imagens de livros de histórias infantis. Em contexto espontâneo promoveram-se momentos de conversação informal, com tópicos antecipadamente definidos e do interesse da criança, assim como de interação em situação de jogo simbólico.

Perante a análise estatística efetuada, observou-se que a DV é distinta nos contextos referidos. Contudo, verificou-se que o comportamento dos dois índices é antagónico. As crianças obtiveram valores médios de RTT' superiores, em contexto narrativo, 0.340, quando comparados com os valores médios obtidos em contexto espontâneo, 0.304. Por seu turno, no índice D os resultados conduzem a uma constatação oposta. Existe uma diferença estatisticamente significativa entre os valores médios de D obtidos em discurso espontâneo e em discurso narrativo. Porém, ao contrário do RTT', o índice D indica que o

desempenho das crianças em contexto espontâneo é superior ao seu desempenho, em contexto narrativo.

Daqui se conclui que apenas o índice RTT vai ao encontro dos resultados obtidos no estudo realizado por Yont et al. (2003). Estes autores constataram que as crianças produziram mais *tokens*, por minuto, numa interação durante o jogo - contexto espontâneo -, do que na leitura conjunta de livros – contexto narrativo -, tendo a produção de *types* sido semelhante nos dois contextos. Aplicando a fórmula de cálculo de RTT aos dados de Yont et al., em que o número de *types* é semelhante em ambos os contextos e o número de *tokens* é superior em contexto espontâneo, obtém-se um rácio inferior neste contexto (0.117) comparativamente com o obtido no contexto narrativo (0.476).

Apesar de os resultados obtidos não serem esclarecedores, aceita-se que a DV possa ser superior em contexto espontâneo, uma vez que a utilização de brinquedos permite a construção de uma interação mais variada, elicitando uma maior quantidade e variedade de discurso espontâneo. Por sua vez, no discurso narrativo, as atividades de descrição de imagens podem tornar-se muito redutoras, uma vez que o vocabulário é como que imposto pelas imagens apresentadas, e incitam à utilização de enunciados aprendidos e não espontâneos, tais como “era uma vez”, “vitória, vitória, acabou-se a história”.

**H4: O desempenho nos índices RTT e D, em contexto narrativo, apresenta variações de acordo com a idade da criança.**

**H5: O desempenho nos índices RTT e D, em contexto espontâneo, apresenta variações de acordo com a idade da criança.**

A análise baseada nas hipóteses supracitadas possibilitou a identificação de diferenças na DV, nos dois contextos discursivos, espontâneo e narrativo, em função da idade.

Relativamente ao RTT, foi possível observar que as crianças com 6 anos apresentaram um valor médio superior no contexto narrativo comparativamente com o valor obtido em contexto espontâneo.

Também no que concerne ao índice D relata-se um efeito de interação entre os contextos espontâneo e narrativo e a faixa etária. Observou-se que, no contexto espontâneo, as produções das crianças de 3 anos são menos diversificadas do que as das crianças de 6 e 7 anos; as crianças de 6 anos apresentaram valores médios de D superiores

em contexto espontâneo, quando comparados com os valores obtidos em contexto narrativo.

Miller, em 1991 (Leadholm & Miller, 1992), terá referido que as medidas de análise linguística utilizadas no seu estudo (NDW, TNW e EME) se encontram fortemente correlacionadas com a idade nas amostras de discurso narrativo, sugerindo que este contexto é o que melhor reflete o desenvolvimento linguístico nas suas diversas vertentes. Segundo o autor, isto justifica-se, provavelmente, porque é exigido, em maior número, a utilização da linguagem oral em tarefas de narrativa do que na interação durante um jogo.

Os resultados obtidos para este trabalho indicam, portanto, que ambos os índices de análise de DV utilizados sofrem a influência da idade, tanto no contexto espontâneo como no contexto conversacional, contudo com comportamentos diferentes.

Em conversação, as crianças mais novas tendem a utilizar consistentemente a estratégia da repetição, quer seja para manter o tópico do diálogo, insistir num pedido ou para chamar a atenção (Sim-Sim, 1998). Nestas faixas etárias, os tópicos estão principalmente centrados nas tarefas que a criança está a desenvolver no momento. Com a progressão da idade, as crianças tornam-se aptas a utilizar e a diversificar a linguagem nas conversações, pelo que se considera válida a existência de variações na DV, em função da idade, em contexto espontâneo.

A narrativa constituirá, provavelmente, a primeira experiência de produção espontânea, sem apoio de um interlocutor, vivida por uma criança. As crianças começam a produzir narrativas de forma primitiva aos 3 anos e, aos 4 anos, aproximadamente, já se verifica a utilização de enunciados aprendidos, tais como “era uma vez” ou “e foram felizes para sempre.” Com o avançar da idade, a criança vai enriquecendo as suas narrativas, selecionando de forma mais cuidada o vocabulário (Sim-Sim, 1998). Considerando que há um desenvolvimento desta competência, é igualmente expectável que também a DV aumente à medida que a criança cresce.

Apesar dos resultados obtidos, considera-se plausível que a DV seja superior em discurso espontâneo, comparativamente com o discurso narrativo, indo ao encontro do comportamento observado para o índice D. Em discurso espontâneo, a criança estará em condições de explorar mais tópicos e assim aplicar um vocabulário mais diversificado, desde que devidamente envolvida na interação. Por seu turno, no discurso narrativo, principalmente em tarefas de descrição de sequências de imagens, verificou-se uma

tendência das crianças para relatar as imagens de forma muito simplificada, onde os conceitos utilizados estavam mais limitados à ação e às personagens.

**H6: Tanto a idade como a DV, medida pelos índices RTT e D, estão correlacionadas positivamente com o desenvolvimento linguístico, de tal modo que:**

**H6.1: crianças mais velhas apresentam um desenvolvimento linguístico maior.**

**H6.2: a um nível linguístico superior está associada uma maior DV.**

O índice de DV, NDW, proposto por Miller, em 1991 (Leadholm & Miller, 1992), está, segundo este autor, fortemente correlacionado com a idade, constituindo, por isso, um bom indicador de desenvolvimento linguístico.

Com as hipóteses aqui definidas em H6, H6.1 e H6.2, pretende-se atestar se também estas medidas de DV são bons indicadores de desenvolvimento linguístico, mais do que a idade, aqui ponderada em meses.

O estudo realizado e aqui exposto confirmou a existência de correlação entre a idade e o desenvolvimento linguístico, aqui representado pelos resultados obtidos pelas crianças em idade pré-escolar e escolar, nos testes TALC e GOL-E, respetivamente.

Tal como seria de esperar, sendo estes dois testes de avaliação linguística aferidos para o português europeu, observou-se a ocorrência de uma progressão no desempenho nos testes TALC e GOL-E ao longo dos grupos etários afetos, a que correspondiam respetivamente as crianças mais novas e mais velhas.

Constatou-se que nas crianças mais novas apenas o índice D se correlaciona positiva e significativamente com o desenvolvimento linguístico. Já nas crianças mais velhas verificou-se que ambos os índices se relacionam positivamente com o desenvolvimento linguístico. Desta forma, em consonância com o proposto por Miller para o índice NDW, confirma-se que crianças mais velhas apresentam um desenvolvimento linguístico maior e a um nível linguístico superior está associada uma maior DV, quando medida pelo índice D.

Tal como referido anteriormente, os valores obtidos para estes índices não deverão, de todo, ser utilizados como valores normativos para o português europeu. Todavia, e considerando o facto de o índice D se correlacionar com o desenvolvimento linguístico, a sua utilização na prática profissional, nomeadamente dos Terapeutas da Fala, poderá

fornecer dados importantes nos momentos de avaliação e monitorização da intervenção em crianças com perturbações da linguagem.

O tempo necessário para a recolha, transcrição e análise de dados poderia constituir uma desvantagem na determinação da DV, na prática comum destes profissionais, contudo, e tal como foi referido no capítulo descritivo do D (I. 1. 4. 5.), o cálculo deste índice exige apenas um mínimo de 50 *tokens*, pelo que a sua utilização será exequível.

## IV. 2. Limitações do Estudo

No decorrer deste estudo reconheceram-se alguns aspetos que podem, eventualmente, estar na base dos resultados inconclusivos.

Tal como referido por vários autores, a recolha de amostras de discurso espontâneo constitui um dos meios mais válidos para determinar a performance linguística de uma criança, contudo apresenta as suas limitações. Porém, este método de recolha de dados sofre uma grande influência do contexto situacional e da familiaridade com o interlocutor. Apesar da preocupação em realizar as recolhas no contexto escolar das crianças, este foi condicionado pelos espaços disponibilizados para o efeito e que corresponderam, na sua maioria, a salas desconhecidas pelas crianças. Para além desse fator, considera-se que os 5 minutos iniciais de interação entre a criança e o investigador, utilizados para o estabelecimento da relação, por vezes, não foram suficientes, comprometendo, assim, a espontaneidade necessária, por parte da criança. Também a utilização de material de gravação induziu, em algumas crianças, a alguma inibição inicial, apesar da explicação prévia sobre o propósito da utilização do mesmo.

Durante as recolhas, observou-se que as crianças mais novas manifestam dificuldades em estabelecer diálogos no jogo simbólico, porque tendem a manipular apenas os brinquedos e a ignorar a presença do interlocutor. Também nas atividades de descrição de imagens ou de construção de histórias optam, muitas vezes, apenas pela enunciação da ação que visualizam, enquanto as crianças mais velhas correspondem ao pretendido através da criação de narrativas.

Considera-se também como limitação o tempo estipulado para as recolhas, bem como o número de tarefas propostas para eliciar discurso narrativo. As crianças mais novas revelaram alguma saturação no decorrer das recolhas.

Outra limitação a atender tanto nas recolhas, como nas transcrições, prende-se com o facto de ambas terem sido efetuadas por duas investigadoras. Apesar de se ter procurado seguir um conjunto de normas, verifica-se que existem sempre variações nas interações e uma forte componente de subjetividade nas transcrições. Aquando das interações, ambas as investigadoras relataram dificuldades em controlar o número de perguntas de resposta fechada, principalmente com as crianças mais novas que não se envolveram tão bem na conversação. Constatou-se que um dos aspetos que dificultou as transcrições, por parte das investigadoras, foi a presença de alterações articulatórias, uma vez que comprometem a inteligibilidade do discurso.

## CONCLUSÃO

O presente estudo surgiu da necessidade de se verificar a utilidade dos índices RTT e D, enquanto medidas representativas da DV, no estudo do desenvolvimento linguístico para o português europeu. Uma vez que não existem estudos neste âmbito para a língua em questão, pretendia-se também traçar um perfil evolutivo de ambos os índices entre os 3 e os 7, de forma a obter valores de referência para proveito de estudos e/ou análises futuras.

A análise realizada aos resultados obtidos conduziu a algumas conclusões que serão aqui apresentadas à luz das hipóteses formuladas:

- O RTT não é capaz de refletir a evolução da DV, entre os 3 e os 7 anos de idade.
- Os valores médios de D diferenciam a DV apenas nas crianças de 3, 6 e 7 anos, sendo que aos 3 anos correspondem produções menos diversificadas.
- Existe um incremento na DV na transição do ensino pré-escolar para o escolar, testemunhado por ambos os índices.
- A DV verificada nos contextos narrativos distingue-se da DV nos contextos espontâneos, em função da idade. Não obstante, da análise efetuada não foi possível concluir em que contexto há uma superioridade na DV, uma vez que as medidas apresentaram comportamentos antagónicos.
- Há uma evolução na DV, medida pelo índice D, em contexto espontâneo. Nesse contexto as crianças de 3 anos utilizam vocabulário menos diversificado do que as crianças de 6 e 7 anos.
- O índice D e a idade são indicadores de desenvolvimento linguístico.

A investigação que esteve na base desta dissertação não permitiu cumprir o seu principal pressuposto, que consistia na identificação de valores de referência para a DV, em crianças com idades compreendidas entre os 3 e 7 anos. Todavia, apesar de todas as limitações identificadas, considera-se que este trabalho assume importância no âmbito da Terapia da Fala, uma vez que evidencia o índice D enquanto indicador de desenvolvimento linguístico.

Para estudos futuros, a realizar no âmbito da DV, sugere-se uma redefinição do critério utilizado, para a classificação de *types*, e julga-se importante a adoção de um número limitado de *tokens*, na constituição da amostra.

Considerando que as dificuldades vocabulares afetam negativamente as produções das crianças, podendo, por conseguinte, limitar a sua participação nas atividades comunicativas, os índices que avaliam a sua performance ao nível das competências discursivas revelam-se bastante informativos e importantes. O índice D, mais do que o RTT, pode ser uma ferramenta muito útil para os terapeutas da fala que procuram implementar a prática baseada na evidência, na medida em que permitirá uma monitorização dos progressos linguísticos das crianças que estão a acompanhar, determinando assim a eficácia da intervenção, para além da sua utilidade na avaliação inicial das competências linguísticas da criança (Heilmann *et al.*, 2010).

Para além disso, o *corpus* constituído para este estudo constitui um importante ponto de partida para investigações futuras, a realizar para o português europeu no âmbito do desenvolvimento linguístico, em geral, e da DV, em particular.

## BIBLIOGRAFIA

Albuquerque, F. (2000). *À Descoberta da Palavra Redondinha - A Linguagem na Primeira Infância* Porto: Porto Editora.

Anónimo. The Brain From Top to Bottom.

Araújo, K. d. (2007). *Desempenho gramatical de crianças em desenvolvimento normal e com Distúrbio Específico de Linguagem*. (Doutorado), 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.

Basílio, M. (2004). O conceito de vocábulo na obra de Mattoso Câmara. *Documentação de Estudos em Linguística Teórica e Aplicada, Vol. 20*.

Bates, E., Dale, P. S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. F. B. MacWhinney (Ed.), *Handbook of Child Language*. Oxford: Basil Blackwell.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., Reilly, J., & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal Child of Language, Vol. 21*, 85-123.

Bornstein, M. H., Leach, D. B., & Haynes, O. M. (2004). Vocabulary competence in first- and secondborn siblings of the same chronological age. *Journal of Child Language, Vol. 31*, 855-873.

Carey, S. (1978). The child as a word learner. In M. Halle, G. A. Miller & J. Bresnan (Eds.), *Linguistic Theory and Psychological Reality*. Cambridge: MIT Press.

Chiappe, P., Chiappe, D. L., & Gottardo, A. (2004). Vocabulary, Context, and Speech Perception Among Good and Poor Readers. *Educational Psychology: An International Journal of Experimental Educational Psychology, Vol. 24(6)*, 825-843.

Clark, E. (1993). *The Lexicon in Acquisition*. Cambridge: Cambridge University Press.

Costa, A. L. d. P. M. B. G. (2010). *Estruturas Contrastivas: Desenvolvimento do Conhecimento Explícito e da Competência de Escrita*. (Doutoramento em Linguística), Universidade de Lisboa - Faculdade de Letras, Lisboa.

Costa, J., & Santos, A. L. (2003). *A Falar como os Bebés*. Lisboa: Editorial Caminho.

D'Odorico, L., Carubbi, S., Salerni, N., & Calvo, V. (2001). Vocabulary development in Italian children: a longitudinal evaluation of quantitative and qualitative aspects. *Journal of Child Language*, *Vol. 28*, 351-372.

deBoer, F. (2014). Evaluating the comparability of two measures of lexical diversity. *System*, *Vol. 47*, 139-145.

Demuth, K. (1998). Collecting Spontaneous Production Data. In D. McDaniel, C. McKee & H. S. Cairns (Eds.), *Methods for Assessing Children's Syntax*. Cambridge: Massachusetts Institute of Technology.

Denes, P. B., & Pinson, E. N. (1993). *The Speech Chain: The Physics and Biology of Spoken Language*. Oxford: W. H. Freeman and Company.

Duarte, I., Colaço, M., Freitas, M. J., & Gonçalves, A. (2011). *O Conhecimento da Língua: Desenvolver a Consciência Lexical*. Lisboa: Ministério da Educação - Direcção-Geral de Inovação e de Desenvolvimento Curricular.

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, *Vol. 4*(2), 139-155.

Fergadiotis, G., & Wright, H. H. (2010). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology*, *Vol. 25*(11), 1414-1430.

Fergadiotis, G., Wright, H. H., & Capilouto, G. J. (2011). Productive vocabulary across discourse types. *Aphasiology*, *Vol. 25*(10), 1261-1278.

Foy, J. G., & Mann, V. A. (2012). Speech production deficits in early readers: predictors of risk. *Reading And Writing*, *Vol. 25*(4), 799-830.

Gershkoff-Stowe, L., & Hahn, E. R. (2013). Word comprehension and production asymmetries in children and adults. *Journal of Experimental Child Psychology*, *Vol. 114*, 489-509.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*, 515-531.

Guasti, M. T. (2002). *Language Acquisition: The Growth of Grammar*. Cambridge: MIT Press.

Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics*, *Vol. 72*, 42-53.

Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, *Vol. 25*, 408-422.

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language*, *Vol. 27*, 689-705.

Heilmann, J., Nockerts, A., & Miller, J. F. (2010). Language sampling: does the length of the transcript matter? *Language, Speech, and Hearing Services in Schools*, *Vol. 41*, 393-404.

Hess, C. W., Haug, H. T., & Landry, R. G. (1989). The Reliability of Type-Token Ratios for the Oral Language of School Age Children. *Journal of Speech, Language, and Hearing Research*, *Vol. 32*, 536-540.

Hess, C. W., Sefton, K. M., & Landry, R. G. (1986). Sample size and type-token ratios for oral language of preschool children. *Journal of Speech, Language, and Hearing Research*, *Vol. 29*, 129-134.

Hickok, G. (2012). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders*, *Vol. 45*, 393-402.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, *Vol. 4*, 131-138.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, *Vol. 92*, 67-99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *Vol. 8*, 393-402.

Hoff-Ginsberg, E. (1991). Mother-child conversation in different social classes and communicative settings. *Child Development*, *Vol. 62*, 782-796.

Hoff, E. (2003). The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *Vol. 74*(5), 1368-1378.

Hoff, E., Laursen, B., & Tardif, T. (2002). Socioeconomic status and parenting - Biology and Ecology of Parenting. In M. H. Bornstein (Ed.), *Handbook of parenting* (Vol. Vol. 2). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Huttenlocher, J. (1998). Language input and language growth. *Preventive Medicine*, *Vol. 27*, 195-199.

Jackson-Maldonado, D., Thal, D., Marchman, V., Bates, E., & Gutierrez-Clellen, V. (1993). Early lexical development in Spanish-speaking infants and toddlers. *Journal of Child Language*, Vol. 20, 523-549.

Jackson, H., & Amvela, E. Z. (2000). *Words, Meaning And Vocabulary*. London: Continuum.

Jarvis, S. (2013). Capturing the Diversity in Lexical Diversity. *Language Learning*, Vol. 63(1), 87-106.

Jr., D. W. Gow. (2012). The cortical organization of lexical knowledge: A dual lexicon model of spoken language processing. *Brain and Language*, Vol. 8, 273-288.

Kay, E. S., & Santos, M. E. (2003). *Grelha de Observação da Língua - Nível Escolar*. Escola Superior de Saúde de Alcoitão.

Kay, E. S., & Tavares, M. D. (2006). *Teste de Avaliação da Linguagem na Criança*. Oficina Didática.

Kim, M., McGregor, K. K., & Thompson, C. K. (2000). Early lexical development in English and Korean-speaking children: language-general and language-specific patterns. *Journal of Child Language*, 225-254.

Koizumi, R., & In'ami, Y. (2012). Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens. *System*, Vol. 40, 522-532.

Laureano, R. M. S. (2011). *Testes de Hipóteses com o SPSS - O Meu Manual de Consulta Rápida*. Lisboa: Edições Sílabo.

LDA. LDA Language Cards - Sequential Thinking Cards: LDA.

Leadholm, B. J., & Miller, J. F. (1992). *Language Sample Analysis: The Wisconsin Guide*. Wisconsin Department of Public Instruction.

Lund, N. J., & Duchan, J. F. (1983). *Assessing Children's Language in Naturalistic Contexts*. New Jersey: Prentice-Hall, Inc.

Ma, Q. (2009). *Second Language Vocabulary Acquisition* Germany: Peter Lang AG, International Academic Publishers

MacWhinney, B., & University, C. M. (2012). *The CHILDES Project: Tools for Analyzing Talk* (3<sup>a</sup> ed.). New Jersey: Lawrence Erlbaum Associates.

Malvern, D. D., Richards, B. J., Chipere, N., & Durán, P. (2004). *Lexical Diversity and Language Development: Quantification and Assessment*. England: Palgrave Macmillan.

Mata, F. S. (2003). *Como Prevenir as Dificuldades na Expressão Escrita*. Porto Alegre: ArtMed Editora.

Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Salamon, G., Dehaene, S., Cohen, L., & Mehler, J. (1993). The Cortical Representation of Speech. *Journal of Cognitive Neuroscience, Vol. 5*, 467-479.

McCarthy, P. M., & Jarvis, S. (2007). vocd: A Theoretical and empirical evaluation. *Language Testing, Vol. 24*(4), 459-488.

McCarthy, P. M., & Jarvis, S. (2010). MTLTD, voc-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, Vol. 42*(2), 381-392.

McKee, G., Malvern, D., & Richards, B. (2000). Measuring Vocabulary Diversity Using Dedicated Software. *Literary and Linguistic Computing, Vol. 15*, 323-337.

McKeown, M. G., & Beck, I. L. (2014). Effects of vocabulary instruction on measures of language processing: Comparing two approaches. *Early Childhood Research Quarterly, Vol. 29*(4), 520–530.

Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary - Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press.

Miller, J. F. (1981). *Assessing Language Production in Children: Experimental Procedures* (Vol. Vol. 1). Texas: Pro-Ed.

Miller, J. F., Andriacchi, K., & Nockerts, A. (2012). *Assessing Language Production Using SALT Software*: SALT Software, LLD.

Ninio, A. (1980). Picture-book reading in mother-infant dyads belonging to two subgroups in Israel. *Child Development, Vol. 51*, 587-590.

Olinghouse, N. G., & Leaird, J. T. (2009). The relationship between measures of vocabulary and narrative writing quality in second- and fourth-grade students. *Reading And Writing, Vol. 22*, 545-565.

Owen, A. J., & Leonard, L. B. (2002). Lexical Diversity in the Spontaneous Speech of Children With Specific Language Impairment: Application of D. *Journal of Speech, Language, and Hearing Research, Vol. 45*, 927-937.

Pan, B. A., Rowe, M. L., Spier, E., & Tamis-LeMonda, C. (2004). Measuring productive vocabulary of toddlers in low-income families: concurrent and predictive validity of three sources of data. *Journal of Child Language, Vol. 31*, 587-608.

Papaeliou, C. F., & Rescorla, L. A. (2011). Vocabulary development in Greek children: a cross-linguistic comparison using the Language Development Survey. *Journal of Child Language, Vol. 38*, 861-887.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, Vol. 8*, 976-988.

Pikulski, J. J., & Templeton, S. (2004). Teaching and developing vocabulary: key to long-term reading success. *Houghton Mifflin Company, 15*, 1-12.

Pollard-Durodola, S. D., Gonzalez, J. E., Simmon, D. C., Kwork, O., Taylor, A. B., Davis, M. J., Kim, M., & Simmons, L. (2011). The effects of an intensive shared book-reading intervention for preschool children at risk for vocabulary delay. *Council for Exceptional Children, Vol. 77*, 161-183.

Pulido, M. T., Simões, M., Moreira, M., & Branco, T. (2011). *Manual de Transcrição do Programa Clan*. Centro de Linguística da Universidade Nova de Lisboa.

Puyuelo, M., & Rondal, J.-A. (2007). *Manual de Desenvolvimento e Alterações da Linguagem na Criança*. Porto Alegre: Artmed Editora.

Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Rebelo, J. A. d. S. (1993). *Dificuldades da Leitura e da Escrita em Alunos do Ensino Básico* (1ª ed.). Rio Tinto: Edições ASA.

Rescorla, L., Mirak, J., & Singh, L. (2000). Vocabulary growth in late talkers: lexical development from 2;0 to 3;0. *Journal of Child Language, Vol. 27*, 293-311.

Rescorla, L. A. (1980). Overextension in early language development. *Journal of Child Language, Vol. 7*, 321-335.

Retherford, K. S. (1993). *Guide to analysis of language transcripts* (2nd ed.). USA: Thinking Publications.

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, Vol. 53*, 333-349.

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2001). Methodological Issues in Content Analysis of Computer Conference Transcripts. *Journal of Artificial Intelligence in Education, Vol. 12*, 8-22.

Rowe, M. L. (2008). Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language, Vol. 35*, 185-205.

Scherer, S., Casarin, F. S., Zart, P., Ramos, A. P. F., & Assencio-Ferreira, V. J. (2002). Perfil evolutivo da relação type/token de crianças de 3 a 5 anos de idade. *Revista CEFAC, Vol. 4*, 223-228.

Scherer, S., & Souza, A. P. R. d. (2011). Types e Tokens na aquisição típica de linguagem por sujeitos de 18 a 32 meses falantes do Português Brasileiro. *Revista CEFAC, Vol. 13(5)*, 838-845.

Schubi. Obrigado, Kim! Alemanha: Schubi Lernmedien AG.

Setola, P., & Reilly, R. G. (2005). Words in the brain's language: An experimental investigation. *Brain and Language, Vol. 94*, 251-259.

Shipley, K. G., & McAfee, J. G. (2008). *Assessment in Speech-Language Pathology: A Resource Manual*. USA: Delmar Cengage Learning.

Silverman, S., & Ratner, N. B. (2002). Measuring lexical diversity in children who stutter: application of vocd. *Journal of Fluency Disorders, Vol. 27*, 289-304.

Sim-Sim, I. (1998). *Desenvolvimento da Linguagem*. Lisboa: Universidade Aberta.

Singleton, D. (2000). *Language and Lexicon - An Introduction*. Great Britain: Arnold.

Sulzer-Azaroff, B. (1997). Why should I talk to my baby? A review of meaningful differences in the everyday experience of young American children by Hart and Risley. *Journal of Applied Behaviour Analysis, Vol. 30*, 599-600.

Tardif, T., Fletcher, P., Liang, W., & Kaciroti, N. (2009). Early vocabulary in Mandarin (Putonghua) and Cantonese. *Journal of Child Language, Vol. 36*, 1115-1144.

Watkins, R. V., Kelly, D. J., Harbers, H. M., & Hollis, W. (1995). Measuring Children's Lexical Diversity: Differentiating Typical and Impaired Language Learners. *Journal of Speech and Hearing Research, Vol. 38*, 1349-1355.

Webb, W. G., & Adler, R. K. (2008). *Neurology for the Speech-Language Pathologist* (5 ed.). Missouri: Mosby.

Westerlund, M., & Lagerberg, D. (2008). Expressive vocabulary in 18-month-old children in relation to demographic factors, mother and child characteristics, communication style and shared reading. *Child: care, health and development, Vol. 34*(2), 257-266.

Wise, J. C., Sevcik, R. A., Morris, R. D., Lovett, M. W., & Wolf, M. (2007). The relationship among receptive and expressive vocabulary, listening comprehension, pre-reading skills, word identification skills, and reading comprehension by children with reading disabilities. *Journal of Speech, Language, and Hearing Research, Vol. 50*, 1093-1109.

Yeung, W. M. a. A. (2013). Multimodal speech chain with feedback loops. *Articulatory Phonetics*: Wiley-Blackwell.

Yont, K. M., Snow, C. E., & Vernon-Feagans, L. (2003). The role of context in mother-child interactions: an analysis of communicative intents expressed during toy play and book reading with 12-month-olds. *Journal of Pragmatics, Vol. 35*, 435-454.

Youmans, G. (1990). Measuring lexical style and competence: the type-token vocabulary curve. *Academic Journal, Vol. 24*(4), 584-599.

## LISTA DE FIGURAS

|   |      |
|---|------|
| <b>Figura 1</b> – Representação dos mecanismos envolvidos na produção e percepção de fala (Yeung, 2013) .....   | 5    |
| <b>Figura 2</b> – Esquema do hemisfério esquerdo do cérebro humano (In: <a href="http://thebrain.mcgill.ca/flash/a/a_10/a_10_cr/a_10_cr_lan/a_10_cr_lan.html">http://thebrain.mcgill.ca/flash/a/a_10/a_10_cr/a_10_cr_lan/a_10_cr_lan.html</a> ) ..... | 7    |
| <b>Figura 3</b> – Representação do modelo da dupla via (adaptado de Hickok (2012)). .....   | 8    |
| <b>Figura 4</b> – Representação esquemática do modelo DIVA (Guenther, 1994; Guenther & Vladusich, 2012). .....  | 9    |
| <b>Figura 5</b> – Conjunto de miniaturas utilizadas em situação de jogo simbólico, referentes à profissão de polícia. ....  | viii |
| <b>Figura 6</b> – Conjunto de miniaturas utilizadas em situação de jogo simbólico, referentes à profissão de bombeiro .....   | viii |
| <b>Figura 7</b> – Conjunto de miniaturas utilizadas em situação de jogo simbólico, referentes à profissão de construtor civil .....   | viii |
| <b>Figura 8</b> – Cenário utilizado no jogo simbólico com as miniaturas das profissões de bombeiro, polícia e construtor civil .....  | ix   |
| <b>Figura 9</b> – Conjunto de miniaturas utilizadas em situação de jogo simbólico, referentes à profissão de médico .....   | ix   |
| <b>Figura 10</b> – Conjunto de miniaturas utilizadas em situação de jogo simbólico, referentes à oficina .....  | ix   |
| <b>Figura 11</b> – Conjunto de miniaturas utilizadas em situação de jogo simbólico, referentes à casa .....   | x    |
| <b>Figura 12</b> – Conjunto de miniaturas utilizadas em situação de jogo simbólico, referentes ao parque .....  | x    |
| <b>Figura 13</b> – Conjunto de miniaturas utilizadas em situação de jogo simbólico, referentes à quinta .....   | x    |

## LISTA DE GRÁFICOS

**Gráfico 1** – Representação das curvas ideias de RTT e relação com o índice.....29

## LISTA DE TABELAS

|   |    |
|---|----|
| <b>Tabela 1</b> – Resultados médios de RTT obtidos por Miller, a partir do estudo de Templin.(Miller, 1981) .....   | 26 |
| <b>Tabela 2</b> – Distribuição da amostra por sexo e faixa etária .....   | 38 |
| <b>Tabela 3</b> – Média, desvio-padrão, valores mínimos e máximos dos índices RTT e D total, espontâneo e narrativo, da amostra. ....   | 54 |
| <b>Tabela 4</b> – Médias e desvios-padrão dos valores dos índices RTT e D, para cada faixa etária. ....   | 55 |
| <b>Tabela 5</b> –Médias e desvios-padrão dos índices RTT e D obtidos para cada nível de escolaridade, dimensões de cada amostra e significância obtida nos testes de hipótese. .... | 56 |
| <b>Tabela 6</b> – Média e desvio-padrão dos valores do índice RTT em contexto espontâneo e narrativo, para cada faixa etária. ....  | 57 |
| <b>Tabela 7</b> – Média e desvio-padrão dos valores do índice D obtidos em contexto espontâneo e narrativo, para cada faixa etária. ....  | 58 |
| <b>Tabela 8</b> – Coeficientes de correlação linear de Pearson entre os índices RTT e D, idade e desenvolvimento linguístico. ....  | 59 |

## **APÊNDICE A**

Autorização da Diretora do Agrupamento de Escolas de Santa Comba Dão

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 “Diversidade Vocabular como medida de desenvolvimento linguístico no português europeu” e “Extensão Média de Enunciado como medida de desenvolvimento linguístico no português europeu”, sob orientação da professora Doutora Maria Lobo e coorientação da professora Doutora 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 do 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 3 anos e 0 meses e os 7 anos e 11 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**

Explicação do Projeto de Mestrado entregue aos Encarregados de Educação

## EXPLICAÇÃO DOS PROJETOS DE MESTRADO

*“Diversidade Vocabular como medida de desenvolvimento linguístico no português europeu” e “Extensão Média de Enunciado como medida 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 Doutora 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 do 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 3 anos e 0 meses e os 7 anos e 11 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 (ex.: 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**

Declaração de Consentimento Informado entregue aos Encarregados de Educação

## DECLARAÇÃO DE CONSENTIMENTO INFORMADO

### Participação nos Projetos de Mestrado

*“Diversidade Vocabular como medida de desenvolvimento linguístico no português europeu” e “Extensão Média de Enunciado como medida 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)

## **APÊNDICE D**

Brinquedos utilizados na recolha de amostras de discurso espontâneo



**Figura 5** – Conjunto de miniaturas utilizadas em situação de jogo simbólico, referentes à profissão de polícia.



**Figura 6** – Conjunto de miniaturas utilizadas em situação de jogo simbólico, referentes à profissão de bombeiro.



**Figura 7** – Conjunto de miniaturas utilizadas em situação de jogo simbólico, referentes à profissão de construtor civil.





**Figura 11** – Conjunto de miniaturas utilizadas em situação de jogo simbólico, referentes à casa.



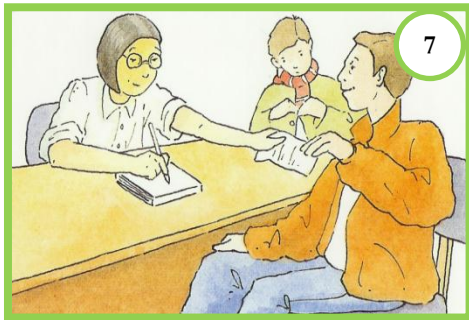
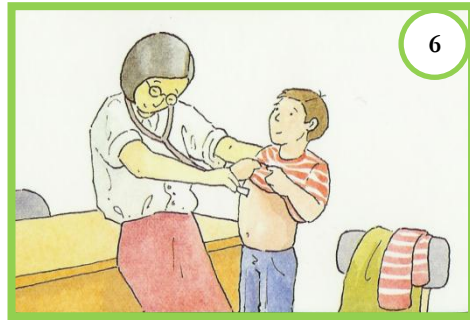
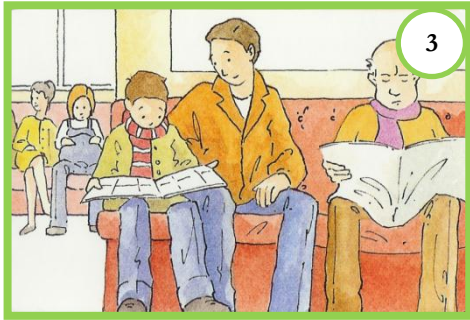
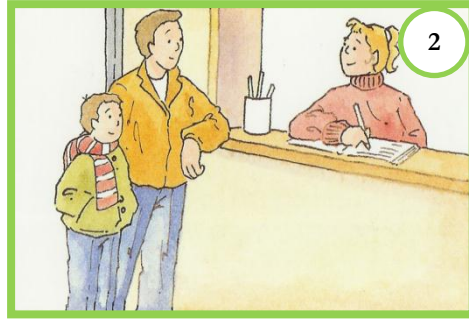
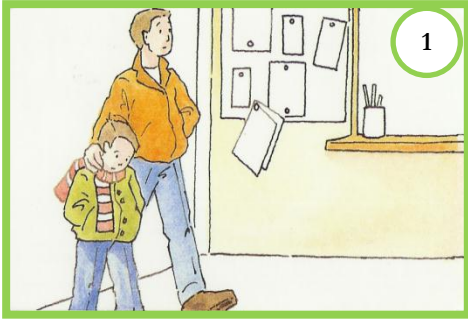
**Figura 12** – Conjunto de miniaturas utilizadas em situação de jogo simbólico, referentes ao parque.

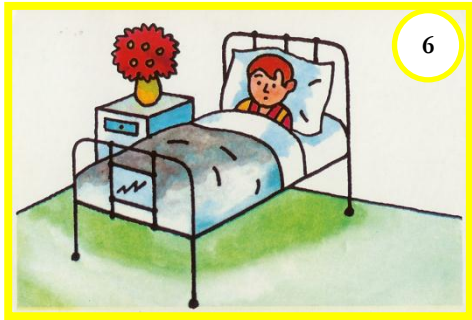
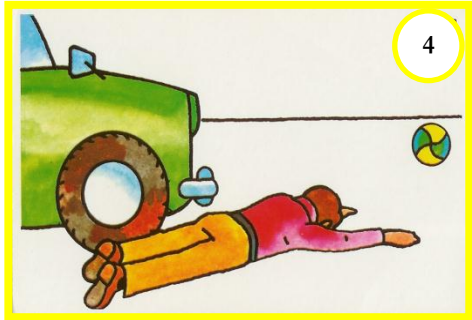
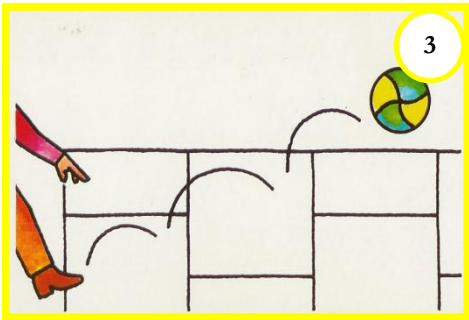
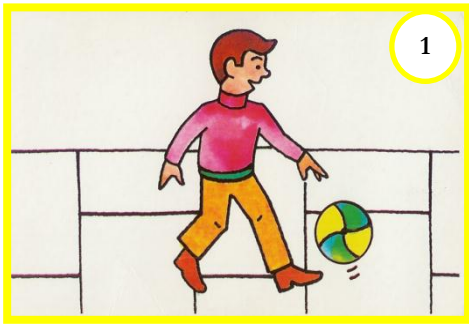


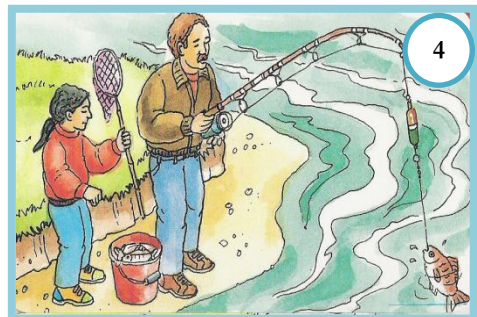
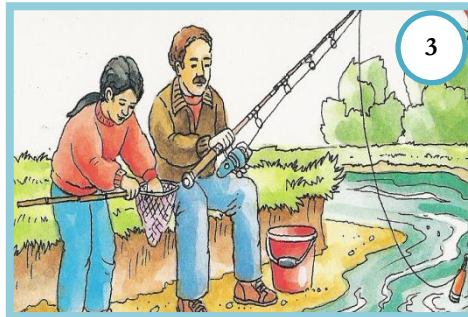
**Figura 13** – Conjunto de miniaturas utilizadas em situação de jogo simbólico, referentes à quinta.

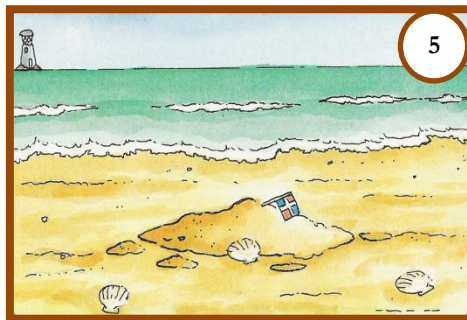
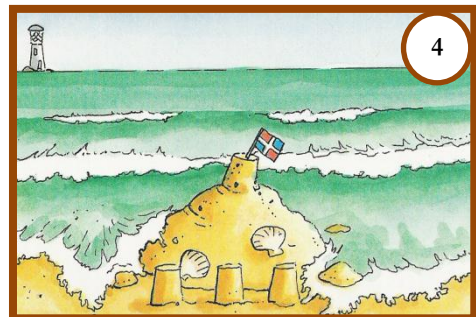
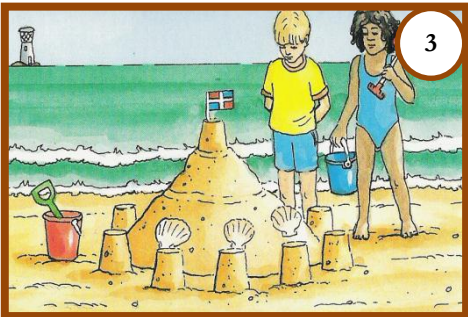
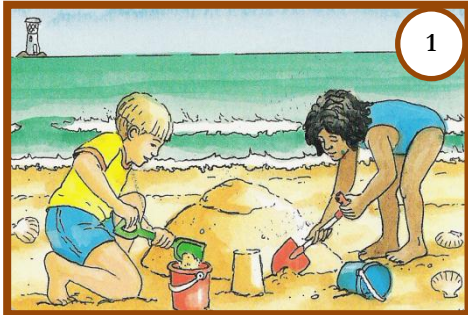
## APÊNDICE E

Sequências de imagens utilizadas na recolha de amostras de discurso narrativo, da coleção “*LDA Language Cards - Sequential Thinking*” (LDA)



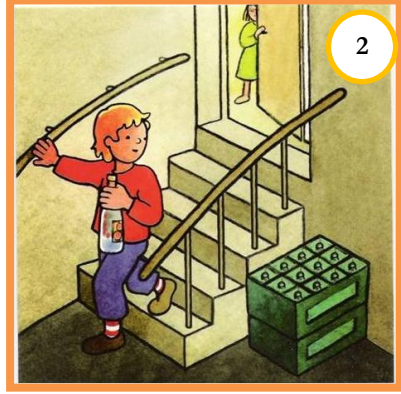


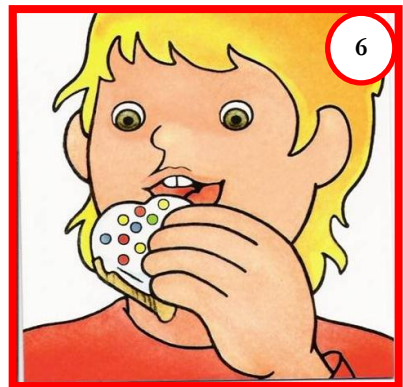




## **APÊNDICE F**

Sequências de imagens utilizadas na recolha de amostras de discurso narrativo, da coleção “Obrigado, Kim!” (Schubi)





## **APÊNDICE G**

Formato do ficheiro CHAT

@Begin

**(Cabeçalho)**

@Languages: por

@Participants: CHI (iniciais do nome e apelido da criança) Target\_Child, INV (nome da investigadora) Investigator

@ID: por|ritaraquel|CHI|(idade da criança a;m.d)|(sexo da criança)||Target\_Chil||

@ID: por|ritaraquel|INV||||Investigator||

@Birth of CHI:

@Date:

@Situation:

**(Transcrição)**

@End

## **APÊNDICE H**

Cr terios de Transcri o do Programa CLAN

| Estrutura do Ficheiro | Códigos utilizados | Descrição  | Exemplos  |
|-----------------------|--------------------|--|---|
| <b>A. Cabeçalho</b>   | @Begin             | Primeira linha obrigatória pela qual começa o ficheiro.  | <p><b>@Begin</b><br/> <b>@Languages:</b> por<br/> <b>@Participants:</b> CHI LD Target_Child, INV Rita Investigator<br/> <b>@ID:</b><br/> por   ritaraquel   CHI   3;9.03   female       Target_Child    <br/> <b>@ID:</b><br/> por   ritaraquel   INV         Investigator    <br/> <b>@Birth of CHI:</b> 15-SEP-2008<br/> <b>@Date:</b> 18-JUN-2012<br/> <b>@Transcriber:</b> Rita Campos<br/> <b>@Location:</b> Santa Comba Dão, Portugal<br/> <b>@Situation:</b> Narração de histórias a partir de sequências e de livros de imagens. Diálogo entre a investigadora e a criança no contexto dos tópicos abordados.<br/> <b>@G:</b> Tipo de discurso<br/> <b>@End</b></p> |
|                       | @Languages:        | Língua(s) utilizada(s) representadas de acordo com o 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<br>B.1. Linha ortográfica | CHI:<br>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.   |
|  | ·<br>!<br>?  | Sinais de pontuação utilizada para marcar o fim de uma linha ortográfica, de acordo com a entoação melódica.              | *INV: pronto!<br>*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 +...<br>*CHI: +, estavam lá a ler umm jornal.  |
|  | +//.   | Marcador final de enunciado incompleto resultante de uma auto-interrupção.  | *CHI: eu queria ver uma &gru +//. [+ esc]<br>*CHI: aqui dois homens iguais.  |
|  | +/.  | Marcador final de enunciado incompleto, interrompido pelo interlocutor.   | *CHI: eu não trouxe assim +/. [+ esc]<br>*INV: espera, que eu não percebi.<br>*INV: do Ben Ten?<br>*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.<br>*CHI: www.<br>*CHI: e depois ab(r)iu a boca pá [: para a] médica ver.                         |  |

| Estrutura do Ficheiro                    | Códigos utilizados         | Descrição  | Exemplos   |
|--|----------------------------|--|--|
| B. Transcrição<br>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 <b>0jogar</b> 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 <b>umm</b> lado pro  |
|  | [+ esc]                    | Marcador final de enunciados incompletos, exceto os que são retomados posteriormente; de contagens de elementos; de enumerações, quando excedem um sintagma nominal.   | [: para o] outro.<br>*CHI: dó ré mi fá só lá si dó ré mi dó ré mi fá sol lá.<br>[+ esc]                                    |
|  | [+ bch]                    | Marcador final de enunciados que contêm desvios sintáticos e morfológicos.   | *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.<br>As siglas são representadas apenas com a letra inicial maiúscula.<br>Os pronomes clíticos separam-se do verbo com um espaço, excluindo-se o hífen. | *INV: estiveste a almoçar <b>M A</b> ?<br>*CHI: &ah escolhia jogador de <b>Nba</b> .<br>*CHI: <b>acabou se</b> 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><b>%act:</b> ruído vindo do exterior.</p> <p><b>@G: espontâneo.</b></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.<br><br>@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><b>@G: narrativo.</b></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> |