

**Instituto de Ciências da Saúde
Universidade Católica Portuguesa**

**Escola Superior de Saúde do Alcoitão
Santa Casa da Misericórdia de Lisboa**

Palavras Complexas nas Primeiras Produções Infantis (estudo de caso)

Dissertação apresentada à Universidade Católica Portuguesa para obtenção do grau de
mestre em Terapia da Fala

Por

Paula Cristina Campino Garcia

Prof^ª. Dr^ª. Marina Vigário

*“O ambiente fornece os estímulos -
neste caso, estímulos linguísticos - e
a criança fornece as respostas –
tanto pela compreensão como pela
produção linguística”
(Vicente Martins)*

Agradecimentos

Em primeiro lugar à minha orientadora, Professora Doutora Marina Vigário, por todas as oportunidades que me deu, por me ter incentivado em todos os momentos, principalmente nos piores, não só por toda a literatura que me indicou como por aquela que me disponibilizou. Por todas as sugestões que fez, pelo olhar atento com que leu repetidas vezes os mesmos textos, por toda a disponibilidade, esclarecimentos, paciência e estímulos que prestou na elaboração desta tese.

À Professora Maria Amália Mendes por me ter facultado os dados respeitantes às classes gramaticais no *corpus* do adulto.

Para a extracção de frequências e número de sílabas foi determinante o uso da ferramenta FreP, desenvolvida pelo Laboratório de Fonética do Centro de Linguística da Universidade de Lisboa, nomeadamente pelos investigadores Fernando Martins, Sónia Frota e Marina Vigário, aos quais agradeço.

Às minhas colegas de mestrado (especialmente a Ana, Marta, Ana Raquel, Lili, Ângela, Bete) pelo companheirismo e pelas gargalhadas durante este longo caminho que iniciou em 2008.

À minha coordenadora de serviço, Dra. Cecília Vaz Pinto, Médica Fisiatra do Serviço de Medicina Física e Reabilitação do SAMS, instituição onde trabalho, pelos incentivos constantes e pelo seu apoio incondicional. O meu sincero obrigado com um carinho especial.

Ao meu colega Fisioterapeuta Luís Santos, pela amizade e companheirismo permanentes. Pela sua disponibilidade sempre presente, principalmente nesta fase final. Pela partilha de angústias e preocupações.

A todas as minhas amigas (especialmente Carmen, Sandra, Gina, Anna, Kamany e a ti Célia) pelo apoio constante, por terem acreditado em mim e principalmente pela Amizade com que me congratulam todos os dias.

Em último, mas em primeiro lugar, às pessoas mais importantes da minha vida, à minha família (pais e irmã), pelas minhas ausências consecutivas, pelo apoio e preocupação constantes, pelo apoio incondicional e pelo suporte emocional e familiar.

Ao Hugo por tornar a minha vida tão complicada o mais fácil possível, por ter aturado todos os meus momentos de impaciência e mau humor, pela ausência, pelo amor e amizade e por me ter ajudado em tantas coisas, sem as quais seria impossível realizar esta dissertação. A ele devo este trabalho, apresentando a minha profunda gratidão.

Agradeço a todos sem exceção, o apoio que me deram. A todos os outros que por lapso possa ter esquecido, o meu muito obrigado.

Resumo

Com esta investigação pretende-se verificar a emergência de palavras complexas no Português Europeu. Com base num diário linguístico que contém as produções de uma criança, falante do Português Europeu, entre 1;01 e os 3;03, procedeu-se à análise das palavras produzidas pela criança quanto à sua frequência, classe gramatical a que pertencem, estrutura interna e tamanho (em número de sílabas), bem como aos morfemas derivacionais concretos nelas contidos, quando existentes.

Com vista à avaliação do papel da frequência no processo de aquisição dos aspectos em investigação, foi feita uma comparação dos resultados descritos para a criança com os dados do *input*. Usou-se para este efeito o *corpus* contendo as produções do adulto. Sobre este *corpus* fez-se um levantamento da frequência das diferentes classes de palavras, das palavras simples e complexas, dos diferentes processos de formação de palavras disponíveis no Português, dos diferentes morfemas derivacionais, da relação entre o formato (tamanho) das palavras e a complexidade morfológica, na fala adulta.

São objectivos do estudo determinar: a) em que fase do desenvolvimento lexical esta criança começa a produzir palavras complexas; b) qual a ordem relativa de emergência dos diferentes tipos de palavras complexas c) nas palavras derivadas, quais os afixos que emergem primeiramente; d) qual o padrão de evolução das categorias das palavras usadas pela criança; e) como se correlacionam os dados de frequência na fala adulta, com a frequência e emergência dos diversos tipos categorias, na fala da criança; f) qual o papel do tamanho das palavras na emergência de palavras complexas.

Os resultados mostram que os dados encontrados no adulto (valores *token*) predizem os resultados que iremos encontrar na criança. Verificámos que o discurso da criança é marcado por palavras simples, que as palavras complexas se encontram em número diminuto e que a primeira palavra complexa surge aos 1;09 anos.

Na linha do desenvolvimento da aquisição da morfologia, contribui-se assim para um novo conhecimento dos padrões de desenvolvimento linguístico normal, para o Português Europeu.

Palavras-chave: palavras complexas, frequência e aquisição da morfologia.

Abstract

The purpose of this research is to verify the emergence of complex words in European Portuguese. Based on a linguistic daily data base which contains the productions of an European Portuguese child between 1;01 and 3;03 of age.

We analysed the words produced by this child, based on their frequency, grammatical class to which they belong, internal structure and size (number of syllables), and derivational morphemes.

To determine the role of frequency, we compared the results described for a child with the *input* data, using the *corpus* of productions by the adult. By means of the *corpus*, an analysis of different word frequency groups, simple and complex words, different derivational morphemes, the relationship between the size of words and their morphology complexity, in adult speech.

The objectives of this research are to determine: a) at what stage of the lexical development, does this child begin to produce complex words, b) in what order do different types of complex words appear, c) in derivatives words, which affixes emerge firstly, d) what is the pattern of evolution of word categories used by the child, e) and the correlation of adult speech frequency data with the emergence and frequency of different types of speech categories in the child, f) how does the size of the words influence the appearance of complex words.

The results found in the adult (*token data*) predict the results that we will find in the child. We found that the child's speech is marked by simple words, complex words are smaller quantity and the first complex word appears at 1;09 of age.

Following the course of the morphology of acquisition, this study contributes to a new understanding of the patterns of normal language development for European Portuguese

Keywords: Complex words, frequency, acquisition of morphology.

Índice

Agradecimentos	V
Resumo	VII
Abstract.....	IX
Índice	XI
Índice Figuras	XIII
Índice de Tabelas	XV
Lista Abreviaturas e Siglas	XVII
I. Introdução	19
I. ENQUADRAMENTO	23
1.1 Noções Morfológicas.....	23
1.2 Aquisição da Morfologia	31
1.3 Importância da informação de frequência no input	39
1.3.1.Factores Externos	44
II. METODOLOGIA	47
2.1 Tipo de Estudo.....	47
2.2 Objectivos	47
2.3 Os dados da Criança (Base de Dados LumaLiDa)	48
2.4 Os dados do Adulto – O Corpus do Português Falado	53
2.5 Critérios para a classificação das palavras	56
III. ANÁLISE DE RESULTADOS	61
III.I. Análise de Resultados na Criança	61
I) 3.1 Palavras simples <i>versus</i> palavras complexas	62
I) 3.2 Tipos de palavras em função da sua estrutura	67
I) 3.3 Classes de Palavras	82
I) 3.4 O tamanho das palavras simples e complexas	91
3.4.1 O tamanho das palavras na criança - produções <i>alvo</i>	91
3.4.2 O tamanho das palavras na criança – realização	95
3.5 Afixos utilizados.....	99
III.II. Análise de Resultados no Adulto	103
II) 3.1 Palavras simples <i>versus</i> palavras complexas.....	104
II) 3.2 Tipos de palavras em função da sua estrutura	106
II) 3.3 Classes de Palavras	111
II) 3.4 O tamanho de palavras simples e complexas no adulto	113
II) 3.5 Afixos utilizados.....	116
IV. DISCUSSÃO DOS RESULTADOS	119
4.1. Palavras simples versus palavras complexas.....	119
4.2. Tipos de Palavras em função da sua estrutura interna.....	122
4.3. Classes de Palavras	126
4.4. Tamanho de Palavras Simples e Complexas	128
4.5. Afixos mais utilizados	130
V. Conclusão	133
VI Referências	137
VII Anexos	143

Índice Figuras

Figura II.1. Extracto da Base de Dados LumaLiDaOn.	50
Figura II.2. Extracto da Base de Dados Léxico_Mensal_Target_LumaLiDa.	52
Figura II.3. Extracto da Base de Dados LEXICOD1Portugalanos90Adulto.....	54
Figura III.1. Palavras simples, complexas e com reduplicação no 1º ano de idade – valores absolutos (type e token).	63
Figura III.2. Tipos de palavras no 2º ano de idade – valores absolutos (type/token).....	64
Figura III.3. Tipos de Palavras no 3º ano de idade – valores absolutos (type/token).....	66
Figura III.4. Classes de Palavras (síntese por ano) – 1º ano (valores percentuais) – Criança (type).	85
Figura III.5. Classes de Palavras (síntese por ano) – 2º ano (valores percentuais) – Criança (type).	86
Figura III.6. Classes de Palavras (síntese por ano) – 3º ano (valores percentuais) – Criança (type).	86
Figura III.7. Classes de Palavras (síntese por ano) – 1º ano (valores percentuais) – Criança (token).	89
Figura III.8. Classes de Palavras (síntese por ano) – 2º ano (valores percentuais) – Criança (token).	90
Figura III.9. Classes de Palavras (síntese por ano) – 3º ano (valores percentuais) – Criança (token).	90
Figura III.10. Número de sílabas - alvo (valores percentuais) – 1º ano - Criança.....	93
Figura III.11. Número de sílabas - alvo (valores percentuais) – 2º ano - Criança.....	93
Figura III.12. Número de sílabas - alvo (valores percentuais) – 3º ano - Criança.....	94
Figura III.13. Número de sílabas - realização (valores percentuais) – 1º ano - Criança.	97
Figura III.14. Número de sílabas - realização (valores percentuais) – 2º ano - Criança.	97
Figura III.15. Número de sílabas - realização (valores percentuais) – 3º ano - Criança.	98
Figura III.16. Afixos utilizados durante o 1º ano (valores absolutos) - Criança (type).	100
Figura III.17. Afixos utilizados durante o 1º ano - Criança (token).	100
Figura III.18. Cinco afixos mais utilizados em cada mês, durante o 2º ano (valores absolutos) - Criança (type).	101
Figura III.19. Cinco afixos mais utilizados em cada mês, durante o 2º ano (valores absolutos) - Criança (token).	101
Figura III.20. Cinco afixos mais utilizados em cada mês, durante o 3º ano (valores absolutos) - Criança (type).	102
Figura III.21. Cinco afixos mais utilizados em cada mês, durante o 3º ano (valores absolutos) - Criança (token).	103
Figura III.22. Palavras simples, complexas e com reduplicação no adulto – valores absolutos (type e token).....	104
Figura III.23. Palavras simples, complexas e com reduplicação no adulto – valores percentuais (type e token).	105
Figura III.24. Tipo de palavras em função da estrutura interna (valores absolutos) - Adulto (type).	106

Figura III.25. Tipo de palavras em função da estrutura interna (valores percentuais) - Adulto (type).	107
Figura III.26. Tipo de palavras em função da estrutura interna (valores absolutos) - Adulto (token).	108
Figura III.27. Tipo de palavras em função da estrutura interna (valores percentuais) - Adulto (token).	108
Figura III.28. Tipo de palavras em função da estrutura interna (valores absolutos) - Adulto (type/token).	109
Figura III.29. Tipo de palavras em função da estrutura interna (valores percentuais) - Adulto (type/token).	110
Figura III.30. Classes de Palavras (valores absolutos) - Adulto (token)	112
Figura III.31. Classes de Palavras (valores percentuais) - Adulto (token).	112
Figura III.32. Número de sílabas (valores absolutos) - Adulto (palavras simples)....	113
Figura III.33. Número de sílabas (valores percentuais) - Adulto (palavras simples)..	114
Figura III.34. Número de sílabas (valores absolutos) - Adulto (palavras complexas).	115
Figura III.35. Número de sílabas (valores percentuais) - Adulto (palavras complexas).	115
Figura III.36. Os 20 afixos mais frequentes na fala adulta e sua frequência (valores absolutos) (type/token).	116
Figura III.37. Os 20 afixos mais frequentes na fala adulta e sua frequência (valores percentuais) (type/token).	117

Índice de Tabelas

Tabela III.1. Palavras simples, complexas e com reduplicação no 1º ano de idade – valores percentuais (<i>type e token</i>).....	63
Tabela III.2. Tipos de palavras no 2º ano de idade – valores percentuais (<i>type/token</i>)..	65
Tabela III.3. Tipos de Palavras no 3º ano de idade – valores percentuais (<i>type/token</i>)..	66
Tabela III.4. Tipo de palavras em função da estrutura interna (valores absolutos) - Criança (<i>type</i>).	68
Tabela III.5. Tipo de palavras em função da estrutura interna (valores percentuais) - Criança (<i>type</i>).	69
Tabela III.6. Ocorrência de palavras cuja estrutura interna revela mais de um processo morfológico (valores absolutos) - Criança (<i>type</i>).	71
Tabela III.7. Ocorrência de palavras cuja estrutura interna revela mais de um processo morfológico (valores percentuais) - Criança (<i>type</i>).	71
Tabela III.8. Distribuição das palavras resultantes de processos morfológicos (derivação, composição e sufixação (z-)avaliativa) (valores absolutos) - Criança (<i>type</i>).	72
Tabela III.9. Distribuição das palavras resultantes de processos morfológicos (derivação, composição e sufixação (z-)avaliativa) (valores percentuais) - Criança (<i>type</i>).	73
Tabela III.10. Tipo de palavras em função da estrutura interna (valores absolutos) - Criança (<i>token</i>).	74
Tabela III.11. Tipo de palavras em função da estrutura interna (valores percentuais) - Criança (<i>token</i>).	75
Tabela III.12. Ocorrência de palavras cuja estrutura interna revela mais de um processo morfológicos (valores absolutos) - Criança (<i>token</i>).	76
Tabela III.13. Ocorrência de palavras cuja estrutura interna revela mais de um processo morfológicos (valores percentuais) - Criança (<i>token</i>).	76
Tabela III.14. Distribuição das palavras resultantes de processos morfológicos de formação de palavras (derivação, composição e sufixação (z-)avaliativa) (valores absolutos) - Criança (<i>token</i>).	77
Figura III.15. Distribuição das palavras resultantes de processos morfológicos de formação de palavras (derivação, composição e sufixação (z-)avaliativa) (valores percentuais) - Criança (<i>token</i>).	78
Tabela III.16. Distribuição das palavras resultantes de processos morfológicos de formação de palavras (derivação, composição e sufixação (z-)avaliativa) (valores absolutos) - Criança (<i>type/token</i>).	79
Tabela III.17. Distribuição das palavras resultantes de processos morfológicos de formação de palavras (derivação, composição e sufixação (z-)avaliativa) (valores percentuais) - Criança (<i>type/token</i>).	80
Tabela III.18. Diferenças entre palavras <i>alvo</i> e realização - Criança (<i>type/token</i>).	81
Tabela III.19. Classes de Palavras (valores absolutos) – Criança (<i>type</i>).	83
Tabela III.20. Classes de Palavras (valores percentuais) – Criança (<i>type</i>).	84
Tabela III.21. Classes de Palavras (valores absolutos) – Criança (<i>token</i>).	87
Tabela III.22. Classes de Palavras (valores percentuais) – Criança (<i>token</i>).	88
Tabela III.23. Número de sílabas - <i>alvo</i> (valores absolutos) - Criança.	92
Tabela III.24. Número de sílabas - <i>realização</i> (valores absolutos) - Criança.	96

Lista Abreviaturas e Siglas

PE – Português Europeu

CV – Consoante Vogal

V – Vogal

CVC – Consoante Vogal Consoante

CVG – Consoante Vogal Glide

CCV – Consoante Consoante Vogal

CS – Child Speech

CDS – Child Directed Speech

ADS – Adult Direct Speech

AS – Adult Speech

SN – Sintagma Nominal

I. Introdução

A presente dissertação tem como foco a aquisição e desenvolvimento de (processos de formação de) palavras complexas no Português Europeu. Até onde sabemos, trata-se de um tópico ainda por trabalhar nesta variedade do Português, e também um tema relativamente pouco explorado nos estudos de aquisição sobre outras línguas.

Sendo os dados descritos para a criança sistematicamente avaliados relativamente às características da fala adulta, que constitui o *input* linguístico da criança, a dissertação traz também conhecimento novo sobre aspectos diversos da morfologia e do léxico no discurso adulto.

Esta tese é um estudo de caso, onde se descreve as primeiras produções de uma criança relativamente a vários aspectos morfológicos e lexicais, incluindo a emergência e desenvolvimento de palavras complexas e das diferentes classes gramaticais; também a relação que pode existir entre a evolução do uso de palavras complexas e a evolução na produção (ou tentativa de produção) de palavras maiores. Para além desta análise, faz-se uma avaliação sistemática do papel da frequência na fala adulta, no comportamento linguístico da criança.

O estudo contribui ainda para a ampliação do conhecimento sobre o desenvolvimento linguístico da criança, cujos dados são aqui explorados. Efectivamente, dados de produção da mesma criança, correspondentes às fases iniciais de produção, foram já analisados do ponto de vista da emergência e/ou evolução de diversos aspectos da fonologia, em articulação com o desenvolvimento de aspectos pertencentes a outras áreas da gramática, como a sintaxe e o léxico (entre os tópicos investigados encontra-se o acento de palavra, a associação tonal, os padrões temporais, os formatos de palavras, a dimensão do léxico, a extensão dos enunciados – veja-se a revisão recente de Frota 2010). Este ponto parece-nos significativo, uma vez que não é habitual, no quadro dos estudos do Português (e também das outras línguas), as primeiras produções linguísticas de um mesmo indivíduo serem analisadas de vários pontos de vista e considerando dados envolvendo mais de uma componente da gramática, permitindo assim uma visão integrada do processo de aquisição da linguagem.

Mais concretamente, o presente estudo incidirá sobre os dados constituídos na base de dados *LumaLiDa*, disponibilizada pelo Laboratório de Fonética da Faculdade de Letras de Lisboa, elaborada a partir de um diário linguístico de uma criança, falante do Português Europeu, compreendendo o período de idade entre os 1;01 e os 3;03 anos. A partir desta base de dados foram analisadas todas as palavras quanto à sua frequência, classe gramatical a que pertencem, estrutura interna e tamanho (em número de sílabas), bem como aos morfemas derivacionais concretos nelas contidos, quando existentes.

Com vista à avaliação do papel da frequência no processo de aquisição dos aspectos em investigação, foi feita uma comparação dos resultados descritos para a criança com os dados do *input*, isto é, os dados da língua adulta, a que a criança está exposta. Usou-se para este efeito o *corpus* do Corpus do Português Falado (Português de Portugal, Anos 90), CR-ROM disponibilizado pelo Centro de Linguística da Universidade de Lisboa. Sobre este *corpus* fez-se um levantamento da frequência das diferentes classes de palavras, das palavras simples e complexas (exibindo os diferentes processos de formação de palavras disponíveis no Português), e dos diferentes morfemas derivacionais, bem como da relação entre o formato (tamanho) das palavras e a complexidade morfológica, na fala adulta.

São objectivos do estudo determinar: a) em que fase do desenvolvimento lexical esta criança começa a produzir palavras complexas; b) qual a ordem relativa de emergência dos diferentes tipos de palavras complexas (i.e. qual a ordem de emergência de palavras derivadas, compostas e formadas com sufixação (z-) avaliativa; dentro das palavras derivadas, qual a ordem de emergência das palavras formadas por prefixação, sufixação e circunfixação); c) nas palavras derivadas, quais os afixos que emergem primeiramente; d) qual o padrão de evolução das categorias das palavras usadas pela criança; e) como se correlacionam os dados de frequência na fala adulta com a frequência e emergência dos diversos tipos categorias na fala da criança; f) qual o papel do tamanho das palavras na emergência de palavras complexas (estará a produção de palavras complexas dependente do tamanho de palavra que a criança tenta ou consegue produzir?).

Em resumo, o estudo pretende contribuir para o conhecimento dos padrões de desenvolvimento linguístico normal para o Português Europeu. Como ficou dito, até onde sabemos, existem muito poucos estudos sobre a aquisição da morfologia no

Português, desconhecendo-se por isso como se processa a aquisição dos processos de formação de palavras e como se caracteriza a expansão do léxico infantil do ponto de vista da complexidade morfológica.

Entre outras aplicações, os resultados desta investigação são relevantes para a área da Terapia da Fala. Nesta área, no trabalho desenvolvido com crianças, é indispensável o domínio aprofundado do conhecimento linguístico infantil, em todas as suas componentes, incluindo a Morfologia e o Léxico. O conhecimento sobre a emergência e uso de palavras complexas nas primeiras produções infantis, sobre os (eventuais) processos envolvidos, bem como sobre o padrão de emergência de novas palavras complexas é, certamente, importante para o trabalho a fazer no âmbito da avaliação do desenvolvimento linguístico da criança, no diagnóstico de possíveis problemas ou atrasos, e na identificação de procedimentos de intervenção visando corrigir problemas encontrados.

Esta dissertação está organizada em capítulos que a seguir sintetizamos:

No Capítulo I (Enquadramento), é feita uma revisão bibliográfica e introdução teórica de informação existente para o Português e para outras línguas sobre três áreas que sustentam este trabalho: noções morfológicas, aquisição da morfologia e importância da informação de frequência no *input*.

No Capítulo II (Metodologia), descreve-se o tipo de estudo desenvolvido, os objectivos e hipóteses colocadas e os *corpora* utilizados neste trabalho. É também realizada uma descrição dos critérios utilizados para a análise das bases de dados envolvidas.

O Capítulo III (Análise dos Resultados) constitui o capítulo central da dissertação. Nele se apresentam os resultados e respectiva análise. O Capítulo encontra-se dividido em duas partes: a análise de resultados na criança e a análise de resultados no adulto.

No Capítulo IV (Discussão), procede-se à discussão dos resultados obtidos quer da criança quer do adulto. É estabelecida uma correlação entre estes dois tipos de resultados e feito um confronto com os elementos referidos na literatura.

Por fim, o Capítulo V (Conclusão) sintetiza o essencial dos resultados obtidos e levanta um conjunto de questões a desenvolver em investigação futura.

I. ENQUADRAMENTO

Neste capítulo faz-se primeiramente uma breve revisão de conceitos básicos relativos à morfologia; de seguida, compila-se informação sobre o que se conhece presentemente acerca da aquisição e desenvolvimento da morfologia; são ainda revistos estudos que revelam o papel da frequência no *input* no comportamento linguístico infantil; finalmente, são considerados os factores externos para a aquisição da língua.

1.1 Noções Morfológicas

A morfologia tem como objecto de estudo central a análise da estrutura interna das palavras existentes e a descrição dos processos morfológicos de formação de novas palavras.

As palavras simples são palavras em que a relação entre a forma, o significado e as propriedades gramaticais que lhes estão associadas é uma relação arbitrária e imprevisível. Estas palavras possuem um único radical, que é uma unidade lexical não analisável, podendo ainda apresentar sufixos que o especificam morfológica e morfo-sintacticamente, variando nas suas categorias de flexão – por exemplo, em *claro*, *sistema* ou *ligo* temos os radicais *clar-*, *sistem-* e *lig-*, que são especificados pelos marcadores de classe ou flexão *-o*, *-a* e *-o* (Vilalva 2008).

Segundo Booij (2007), existem duas perspectivas quando se fala em morfologia: a) o estudo sistemático da correspondência entre forma e significado nas palavras da língua (desde que tomemos como ponto de partida as propriedades das classes de palavras para a análise morfológica), sendo esta considerada uma perspectiva pragmática; b) o estudo da estrutura interna das palavras, sendo esta classificada como uma perspectiva sintáctica.

Do lado da visão sintáctica, as palavras são analisadas em função dos seus morfemas constituintes, sendo os morfemas, isto é, os blocos de construção das palavras, definidos como as unidades linguísticas mínimas, com um significado lexical ou gramatical.

Sob o ponto de vista da aquisição de linguagem, o surgimento de palavras complexas inicia-se pela análise morfológica. Na aquisição da língua materna, é

necessário antes de mais que a criança detecte, nas palavras que integram o seu léxico, a existência de padrões morfológicos. Contudo, para este efeito é necessário que o léxico atinja uma certa dimensão, uma vez que é apenas quando a criança adquire um número suficiente de palavras do tipo relevante que ela pode identificar, com um padrão sistemático e abstracto, conjuntos de palavras relacionadas, que permitam isolar morfemas e inferir processos morfológicos, os quais podem, a partir daí, ser usados para produzir novas palavras (Booij, 2007).

Para além da flexão, que não gera novas palavras, mas antes as modifica ou especifica gramaticalmente, existem dois processos muito generalizados nas línguas que permitem a ampliação lexical: a criação de novas palavras pode ser obtida através da composição – que ocorre quando dois radicais ou palavras se juntam para formar uma nova palavra, ou através da derivação - quando são realizadas operações morfológicas sobre um radical ou palavra, consistindo, em geral, na adição de um ou mais afixos ao radical ou palavra.

Uma das razões para a formação de novas palavras, é, sem dúvida, a necessidade que os falantes têm de atribuir novas expressões para designar objectos ou conceitos novos. Para além disso, existe também uma razão pragmática, pois criando uma nova palavra para um novo conceito estamos a dar relevo a esse mesmo conceito.

Para além dos mecanismos regulares de formação de novas palavras, podem também emergir novas formas através da evolução morfológica. Um exemplo deste tipo de nova formação é a reanálise dos morfemas diminutivos, em muita línguas. Por exemplo, uma palavra com o sufixo *-zinho*, no Português, como *avozinho*, não significa avô pequeno, mas sim, tipicamente, avô querido.

Em diferentes línguas, a morfologia é extremamente importante para a dimensão do léxico. De acordo com Booij (2007), em todas as línguas europeias o número de palavras complexas é muito mais elevado que o número de palavras simples. São as regras morfológicas da língua que permitem a criação de novos lexemas.

A morfologia permite expandir o léxico - o conjunto de palavras estabelecido para a língua; mas esta não é a única fonte de criação de unidades lexicais e nem mesmo de todas as palavras complexas. Estas, podem também surgir de empréstimos ou de criação de palavras contendo novos radicais (Booij, 2007).

É extremamente importante que os dicionários mentais tenham um conhecimento das relações entre palavras. O conhecimento da língua inclui o

conhecimento sistemático nas relações entre a forma e o significado das palavras (Booij, 2007).

O aspecto criativo do conhecimento da morfologia permite-nos não apenas construir, mas também compreender palavras novas. O conhecimento morfológico pode, assim, levar à criatividade regulada no uso da linguagem.

Importa, de todo o modo, referir que existe uma diversidade considerável na morfologia das diferentes línguas do mundo.

Apesar do termo morfologia ter surgido há muito, presentemente, este termo refere-se ao estudo da estrutura interna das palavras e à correspondência sistemática entre forma e significado, nas diferentes palavras (Booij, 2007).

As palavras pertencentes a classes principais e que admitem formação morfológica de neologismos, podem ser analisadas morfológicamente; com base nesta análise morfológica é possível estabelecer uma distinção entre palavras simples e complexas. No Português, as palavras complexas podem pertencer à classe dos adjectivos, nomes, verbos e advérbios (Mateus *et al.*, 2003).

Nesta língua, as palavras complexas possuem um radical e um afixo derivacional, ou dois radicais ou palavras – por exemplo, em *clarificar* e *desligar* temos os radicais *clar-* e *lig-* que formam bases morfológicas às quais se juntam o sufixo *-ific-* (mais a vogal temática e o sufixo marcador da forma infinitiva) e o prefixo *des-*, respectivamente; e em *ecossistemas* e *surdo-mudo* temos os radicais concatenados *eco-* e *sistem-*, aos quais se junta a vogal de classe e o sufixo flexional, e as palavras, que integram radicais e morfemas de classe, *surdo* e *mudo*.

Nas palavras complexas, o radical é formado por mais de um constituinte morfológico, um radical e um afixo (prefixo ou sufixo), quando se trata de uma estrutura de afixação (derivação ou modificação), ou dois ou mais radicais, se for uma estrutura de composição (morfológica ou morfossintáticas) (Villalva, 2008).

Alterar a classe gramatical de uma palavra, como pode acontecer na criação de um verbo a partir de um nome, é designada por *conversão* e este processo pode ser incluído na derivação (Booij, 2007). Este processo de formação de novas palavras está também disponível no Português.

A derivação é um processo de formação de lexemas por via da afixação, conversão e reduplicação. Estas diferentes operações morfológicas, correspondem a diferentes funções sintáticas e semânticas.

A função básica do processo de derivação é permitir aos falantes fazer novos lexemas. Os lexemas pertencem a uma categoria lexical, como a nominal, verbal e adjectival e os lexemas derivados podem pertencer a categorias diferentes das suas bases.

Segundo Fabb (1998), a reduplicação de palavras complexas é por vezes descrita como um processo de composição, porque cada parte da palavra corresponde a uma palavra independente e correcta. Estas situações ocorrem em línguas como o Tamil (língua dravídica falada no sul da Índia) e o Inglês. Também nestas línguas existem casos em que, após a reduplicação, a segunda parte é ligeiramente modificada, formando uma nova palavra composta (Spencer and Zwicky, 2001).

A reduplicação constitui um processo morfológico regular nalgumas línguas do mundo, contudo, no Português tal não se verifica, pelo menos na língua adulta. É frequente, em fases precoces de aquisição das línguas haver reduplicações e existem palavras lexicalizadas na fala adulta cuja forma envolve a repetição de sílabas.

Geralmente não é feita uma distinção entre afixos derivacionais e afixos modificadores, na gramática portuguesa, mas ao tratar todos de forma igual, perde-se a especificidade de ambos os tipos de afixos (Villalva, 2008).

Os afixos modificadores restringem o conjunto de bases a que se podem associar, mas têm menos restrições de selecção do que os derivacionais. São frequentemente avaliativos, sendo a sua função principal veicular um juízo de valor do locutor acerca da base a que o afixo se associa. Os sufixos modificadores são exclusivamente chamados de sufixos diminutivos, aumentativos ou sufixos de grau. Todos eles exprimem um juízo de valor do locutor relativamente ao conteúdo semântico da forma de base, designando-se por avaliativos.

Uma questão interessante relacionada com a morfologia dos avaliativos, diz respeito à existência de duas séries de sufixos, que se distinguem pela ausência ou presença da consoante /z/ no início da sequência. Esta consoante é considerada como um elemento de ligação que se interpõe entre a forma de base e o sufixo. Na verdade, trata-se de diferentes sufixos; as formas de base a que se associam exibem diferentes propriedades. Efectivamente, a distribuição destes afixos parece sugerir que os sufixos avaliativos sem consoante inicial, como *-inho*, são preferidos para a modificação diminutiva/valorativa/afectiva de bases com menor número de sílabas e pertencem a um léxico mais frequente e informal (Villalva, 2008). Estes sufixos tomam o nome de sufixos avaliativo e de sufixos z-avaliativos (Mateus *et al.*, 2003).

A produtividade destes afixos está ainda por determinar, suspeitando-se que a distribuição dependerá de factores de natureza dialectal ou sociolectal.

A principal diferença entre estes dois tipos de sufixos está relacionada com a categoria morfológica da forma de base, pois os sufixos avaliativos associam-se a radicais, enquanto os sufixos *z*-avaliativos se associam a palavras.

É de realçar que *-inho* e *-zinho* não são os únicos sufixos deste tipo disponíveis no Português, sendo contudo possivelmente os mais frequentes em alguns dialectos, nomeadamente no de Lisboa. Consideram-se, também, *-aço* (*filmaço, febraço*), *-ão* (*carrão, pernã*), *-ote* (*caixote, rapazote*) (Mateus *et al.*, 2003).

Existem outros processos que permitem o alargamento do léxico no Português, mas que não têm um carácter regular. Por exemplo, a formação de acrónimos, abreviaturas e formas amalgamadas, os empréstimos e as alterações de significado de palavras existentes, resultantes de extensão metafórica, são também formas possíveis de criar novas palavras. As palavras originadas deste modo apresentam uma componente aleatória, pois não é possível prever as condições em que surgem, nem a forma que tomam, nem o significado que adquirem (Mateus *et al.*, 1990).

A *amálgama*, em particular, é a classificação que tem sido usada para designar o processo de combinação aleatória de segmentos de palavras que, de forma geral, consiste na justaposição da primeira parte da primeira palavra à última parte da segunda palavra, mas outras combinações também são possíveis. Algumas destas formas, estão perfeitamente integradas no léxico do Português, sendo outras sentidas como neologismos mais ou menos aceites, familiares ou integrados (Villalva, 2008).

A *amálgama* pode resultar da combinação de palavras de diferentes classes gramaticais e de diversas formas de sequências, originando-se a partir da sobreposição de sílabas homófonas em fronteira de palavra ou do truncamento numa ou em ambas as palavras. Pode aparecer truncada uma sequência mais ou menos longa no final da primeira palavra ou em ambas. De um modo geral, os segmentos truncados não são sufixos, nem prefixos, nem radicais, isto é, não são unidades morfológicamente reconhecíveis (Nunes & Coimbra, 2007).

Uma outra acepção de *amálgama* é a que envolve a lexicalização de palavras ou morfemas pertencentes a diferentes categorias, e que não resultam de um processo regular de composição – e.g. a combinação lexicalizada da preposição *em* com o artigo definido *a* (>*na*), ou o sufixo de flexão marcando o tempo-modo-aspecto e a pessoa-número nas formas verbais do pretérito perfeito do indicativo (e.g. Mateus *et al.*, 2003).

A flexão é o processo morfológico de formação de palavras que se caracteriza pela sua obrigatoriedade e sistemacidade, isto é, se uma dada categoria de palavras é flexionável numa dada categoria morfo-sintáctica, então todas as palavras pertencentes a essa categoria sintáctica são flexionáveis na referida categoria morfo-sintáctica (por exemplo, os adjectivos flexionam em número, então todos os adjectivos flexionam em número). Por outro lado, a realização da flexão só pode variar em função de subclasses morfológicas identificadas no domínio da categoria de palavras em questão (Mateus *et al.*, 2003).

No Português Europeu, é necessário distinguir entre flexão nominal e flexão verbal; os adjectivos e os nomes flexionam em número, através do sufixo -s e os verbos flexionam em tempo-modo-aspecto e pessoa-número, marcados pelos sufixos de flexão verbal. As preposições e os advérbios são palavras invariáveis.

As palavras estão divididas em duas grandes classes lexicais: classes abertas e classes fechadas. Na maioria das línguas, incluindo o Português, nomes, adjectivos e verbos formam classes abertas. Estas classes podem ser alargadas por formação de palavras. As palavras de função, tal como os determinantes, conjunções, pronomes, preposições, formam uma classe que não pode ser alargada pelos padrões de formação de palavras e por isso é considerada uma classe fechada (Villalva, 2008).

Muitas línguas têm uma quarta classe aberta, a dos advérbios. Esta classe também pode ser alargada pelas vias regulares. No Inglês, os advérbios podem derivar dos adjectivos ou nomes através da adição de sufixos. No Francês, tal como no Português e na generalidade das línguas Românicas, podem ser construídos a partir de formas adjectivais, por um processo de sufixação.

Os advérbios são especiais, uma vez que eles tendem a não alimentar outros processos de formação de palavras, ao contrário das outras três categorias abertas de palavras, as quais podem também flexionar.

A composição é um processo de formação de palavras que consiste na junção de duas ou mais variáveis lexicais, que podem ser radicais ou palavras. No Português Europeu, há dois tipos de composição: a) a composição morfológica, que relaciona radicais segundo os princípios da formação morfológica de palavras; b) a composição morfo-sintáctica, que é um processo híbrido de formação de palavras, no qual se conjugam propriedades das estruturas sintácticas e morfológicas (Mateus *et al.*, 2003).

A estrutura dos compostos morfológicos é, no entanto, muito diferente da estrutura dos compostos morfo-sintácticos.

A composição morfológica é um processo de junção de dois ou mais radicais, que exige a presença de uma vogal de ligação como delimitador da fronteira entre esses radicais. Os radicais que integram este tipo de compostos podem estabelecer entre si uma relação de modificação ou uma relação de coordenação (Mateus *et al.*, 2003).

Na maioria das línguas, a composição é o processo mais frequente para formar lexemas. Muitas vezes é definida pela propriedade de combinar lexemas e consiste na combinação de duas palavras, em que uma palavra modifica o significado da outra. (Booij, 2007).

A produtividade dos compostos, em muitas línguas, deve-se à transparência e versatilidade semânticas. Quando um novo composto é formado, sabemos automaticamente o significado dos seus constituintes e a única tarefa é encontrar a relação semântica entre eles (Booij, 2007).

No Inglês, os falantes sabem que, quando estão perante uma palavra composta como *bottle factory* (fábrica de garrafas, em Português), é o constituinte que se encontra mais à direita que determina o significado do composto. Efectivamente, o significado de *bottle factory* é um tipo de fábrica (*factory*) e não um tipo de garrafa (*bottle*).

A presença de uma vogal de ligação é um dos aspectos mais característicos da estrutura dos compostos morfológicos. A diferença entre as formas que exibem o mesmo radical à esquerda e à direita da vogal de ligação, mostram que a mesma é um constituinte autónomo e que ocupa uma posição própria na estrutura dos compostos.

Os compostos morfo-sintácticos são unidades lexicais que ocupam posições terminais nas estruturas sintácticas, mas que têm uma estrutura híbrida, exibindo algumas das propriedades das estruturas morfológicas. Para o Português, os compostos morfo-sintácticos podem ser gerados por adjunção, conjunção ou por reanálise de uma expressão sintáctica. Esta classificação afasta-se da tradicional distinção entre justaposição e aglutinação e estabelece uma fronteira entre compostos morfo-sintácticos e expressões sintácticas lexicalizadas (Mateus *et al.*, 2003).

A distinção entre justaposição e aglutinação baseia-se no facto de a estrutura fonológica dos compostos conter ou não conter fronteiras de palavra, sendo que, a fronteira de palavra é o resultado de uma transformação progressiva e gradual. Assim, a distinção entre justaposição e aglutinação não permite identificar duas subclasses de compostos morfo-sintácticos, mas sim dois diferentes estádios de lexicalização destes compostos (Mateus *et al.*, 2003). Assim, os compostos aglutinados são aqueles que sofreram um processo de lexicalização semântica e formal, tratando-se de formas que

perderam a sua estrutura interna. Os compostos classificados como justapostos são aqueles que sofreram apenas um processo de lexicalização semântica.

Os compostos morfo-sintácticos formados por adjunção são estruturas constituídas por dois nomes, com comportamentos semelhantes às estruturas sintácticas de núcleo nominal. O constituinte da direita é um modificador nominal, pelo que a estrutura destes compostos é uma estrutura de adjunção que envolve, exclusivamente, unidades lexicais e gera uma nova unidade lexical (Mateus *et al.*, 2003).

Os compostos morfo-sintácticos que têm uma estrutura de conjunção podem ser nomes ou adjectivos, sendo mais frequentes aqueles que envolvem coordenação de nomes do que coordenação de adjectivos, podendo gerar sequências de mais de dois termos coordenados.

A reanálise é um processo de formação de palavras que consiste na reinterpretação de uma estrutura sintáctica, como uma palavra (Mateus *et al.*, 2003).

Os afixos são constituintes morfológicos que se caracterizam pelo facto de associarem obrigatoriamente a uma base, não podendo ocorrer independentemente. São caracterizados quanto à posição que ocupam relativamente à sua base, designando-se *prefixos* os que ocorrem à sua esquerda e *suffixos* os que ocorrem à sua direita.

Na estrutura das palavras simples do Português, a posição de núcleo é ocupada por um radical morfológicamente inanalísável. A estrutura das palavras complexas formadas por derivação corresponde a uma expansão dessa estrutura. Desta forma, a posição de núcleo é ocupada por radicais complexos, aos quais podem ser adicionados novos afixos (Mateus *et al.*, 2003).

Na morfologia do Português, geralmente o complemento precede o núcleo, envolvendo a derivação quase exclusivamente sufixação.

Uma questão importante neste domínio diz respeito à organização do léxico mental: como é que os falantes reconhecem os morfemas, como é que estes estão organizados no nosso léxico mental, como podem ser identificados e posteriormente, combiná-los com outros morfemas para formar palavras complexas.

O conhecimento lexical de um falante de uma determinada língua, não tem o formato de um dicionário. Existem várias diferenças entre o léxico mental e um dicionário: primeiro, o número de entradas lexicais num bom dicionário é muito maior do que uma lista mental de palavras num indivíduo; segundo, as relações entre palavras existem no léxico mental, mas não no dicionário, isto é, palavras com significados

semelhantes ou formas fonológicas similares aparecem relacionadas no léxico mental (tal como mostrado, por exemplo, pelos *lapsus linguae* observados nas produções de fala), mas não aparecem relacionados no dicionário.

Por outro lado, num dicionário, as relações semânticas entre palavras não estão expressas de forma óbvia. Então, o léxico mental pode ser descrito como uma rede multidimensional de palavras, com todo o tipo de conexões entre as suas palavras, incluindo ligações de formas exibindo semelhanças e relações semânticas, fonológicas e morfológicas (e.g. Booij, 2007).

Existem duas vias para que a informação linguística relativa às palavras complexas possa ser acedida: através da análise ou do armazenamento. Na análise, a palavra complexa é primeiro decomposta nos seus morfemas constituintes, e depois o significado desses constituintes é retido na entrada lexical mental. No armazenamento, o reconhecimento da palavra é realizado através da correspondência directa com a respectiva entrada lexical mental. O primeiro cenário é o único possível para as palavras complexas, com as quais nunca nos deparámos antes, mas que conseguimos interpretar (Booij, 2007).

Os resultados de um recente estudo, incidindo sobre o reconhecimento de compostos, sugerem que o sistema humano lexical não visa a optimização do sistema recorrendo fundamentalmente à computação (ou análise), de modo a armazenar menos, nem o inverso; efectivamente, parece que o cérebro humano não optimiza apenas um mecanismo, mas sim combina duas vias. A mente parece ser caracterizada pela redundância em vez da eficiência, o que pode contribuir para a rapidez com que se desenrola o processo de reconhecimento lexical (Libben & Jarema, 2006).

1.2 Aquisição da Morfologia

De acordo com Clark (2001), as crianças começam a produzir as primeiras palavras entre os 12 e os 20 meses de idade. Elas produzem sistematicamente variações morfológicas de palavras durante o seu primeiro ano de fala. Com o avançar do tempo, as crianças vão aumentando a complexidade dos significados e vão adicionando morfemas gramaticais – prefixos, sufixos, preposições, clíticos, etc. Nos nomes, por exemplo, as crianças começam por adicionar morfemas para marcar distinções como género ou número. No caso dos verbos, adicionam marcadores de aspecto, tempo,

género, número e pessoa. O domínio da linguagem pelas crianças leva vários anos, sendo alguns aspectos de domínio mais tardio. Por exemplo, 1) algumas distinções entre significados podem ser mais complexas do que outras e assim necessitarem de mais tempo de aquisição; 2) alguns paradigmas são menos regulares e também implicam mais tempo; 3) as características individuais de cada língua podem afectar e influenciar o processo de aquisição da morfologia (por exemplo, os sufixos na língua Inglesa são adquiridos mais cedo do que os prefixos) (Spencer and Zwicky, 2001).

Ainda segundo Clark (1995), entre os 2 e os 2,6 anos as crianças normalmente produzem entre 500 e 600 palavras. Nos primeiros meses enquanto falantes, elas falam tipicamente sobre pessoas e animais, brinquedos e utensílios de casa, comida e bebidas e actividades e rotinas do dia-a-dia. Quando vão adicionando palavras ao seu vocabulário, elaboram cada domínio semântico, reorganizando e dividindo cada termo disponível para a produção. Em cada ano que vai passando e com a passagem do tempo, as crianças vão acrescentando novos conceitos ao seu vocabulário e encaixando essas novas palavras em categorias e sub-grupos. As crianças têm um vocabulário mais reduzido que os adultos que as rodeiam, por isso deparam-se com várias lacunas semânticas que têm que ultrapassar (Fletcher & MacWhinney, 1995).

Em relação às categorias gramaticais e de acordo com Boysson-Bardies (1999) (citado por Herschensohn, 2007), as crianças adquirem primeiro os nomes, constituindo esta classe a maior parte do seu vocabulário, na maioria das línguas. Comparativamente, de acordo com um estudo feito pelo mesmo autor, na língua francesa, americana, japonesa e sueca as crianças entre os dez e dezoito meses de idade apresentam percentagens diferentes no uso de nomes, verbos e outras categorias de palavras, mas em todas elas a percentagem de nomes é sempre muito superior.

Clark (1995) refere que, relativamente às classes sintácticas, as crianças começam a fazer a distinção entre as diferentes classes gramaticais quando começam a agrupar as palavras em categorias (acções, animais, ferramentas, ...). As crianças têm acesso a esta informação proveniente dos adultos (que apresentam uma correlação entre a categoria e a classe de palavras), que lhes possibilita um caminho possível para a aquisição da sintaxe. Posteriormente, as crianças são assim capazes de fazer a distinção entre nomes e verbos (Fletcher & MacWhinney, 1995).

No caso do Inglês Americano, a predominância de nomes está directamente relacionada tanto com questões linguísticas como, possivelmente também, aspectos culturais do *input*. No campo linguístico os nomes tendem a ser mais acentuados,

ocupando muitas vezes a posição final na frase, que é a posição mais proeminente. No campo cultural, os cuidadores manuseiam e falam sobre objectos concretos no meio ambiente, o que pode conferir aos nomes que os denotam uma maior saliência (Herschensohn, 2007).

As crianças descobrem nas palavras que extraem a partir do *input* propriedades recorrentes, por exemplo no Inglês, a partir dos dados do *input* elas podem inferir que o plural é feito através da adição de um *-s* a uma forma que ocorre no singular; a partir daí, a criança é capaz de aplicar esta regra a novos casos, flexionando assim novas palavras. Numa primeira fase, a criança faz tipicamente generalizações para todas as palavras quer sejam regulares ou irregulares. Surgem deste modo no léxico infantil formas distintas das do adulto, quando irregulares (Booij, 2007). Estas formas constituem evidência inequívoca de que a criança já possui regras morfológicas, produzindo palavras que não podem estar armazenadas como um todo no seu léxico mental, dado que elas não existem, nessa forma, no *input*.

Relativamente aos estádios da aquisição da morfologia, a informação disponível na literatura é relativamente rara. Booij (2007) refere-se a um artigo clássico de Berko (1958), onde é descrita a aquisição das formas do Pretérito Perfeito dos verbos ingleses (Past Tense), seguindo um processo que recorrentemente se encontra nas línguas: na primeira fase, as crianças aprendem um número de verbos no Pretérito Perfeito de forma automática e não analisada, o que significa que, nesta fase, as crianças produzem verbos regulares e irregulares de forma correcta (e.g. produzem correctamente formas como *asked*, formação regular do Pretérito Perfeito do verbo *to ask* ‘perguntar, pedir’ e *went*, formação irregular deste tempo verbal do verbo *to go* ‘ir’); no segundo estádio, as crianças descobrem a regra para formar este tempo verbal e aplicam-na tanto a verbos regulares como a verbos irregulares, fazendo erros como *goed* em vez de *went*; neste momento dá-se uma sobregeneralização, aplicando-se a regra a todas as formas verbais; no terceiro estádio, a criança adquire a regra mas também as excepções, isto é, consegue produzir correctamente verbos regulares e irregulares. Esta aprendizagem é feita em forma de “curva em U”, pois as crianças apresentam um maior número de palavras na forma correcta na fase Um, depois este número diminui na fase Dois e volta a aumentar na fase Três. Como foi dito, esta curva de aprendizagem é encontrada em várias línguas (Booij,2007).

Uma evidência conclusiva da aprendizagem das regras morfológicas pela criança é a capacidade das crianças de produzirem novas formas de palavras, que não estão

disponíveis no *input*. A partir do momento em que as crianças as conseguem realizar de formas apropriadas, isso significa que adquiriram as regras morfológicas. Do mesmo modo, quando produzem formas irregulares segundo as regras regulares (como no exemplo acima) elas mostram ter adquirido regras morfológicas da língua.

Relativamente às regras morfológicas e de acordo com Ingram (2007), apenas se poderá dizer que os falantes utilizam a decomposição morfológica de palavras a partir do momento em que têm acesso a regras produtivas de composição morfológica, para construir novas formas de palavras. Isto é, apenas a partir do momento em que os indivíduos adquirem as regras para produzirem palavras compostas, podem fazer decomposição morfológica. Note-se, porém, que para inferir regras morfológicas é também preciso que o falante perceba que as palavras são decomponíveis morfológicamente.

Pode ser razoável argumentar que nem todos os falantes perceberão algumas das relações morfológicas (transparentes) entre palavras.

No quadro dos estudos de aquisição das palavras complexas, a composição parece atrair mais a atenção do que a derivação. Uma das razões que pode justificá-lo diz respeito ao papel que esta classe de palavras desempenha em línguas germânicas, como o Inglês, que são habitualmente as línguas mais estudadas.

De acordo com Nicoladis (2006), nas crianças, as novas formas de palavras, bem como os erros na criação dessas mesmas formas, dão-nos algumas pistas relativamente à aquisição de estruturas “subliminares” necessárias para a formação de novas palavras (Libben and Jarema, 2006). Um exemplo dado por aquela autora, é o de uma criança bilingue Francesa-Inglesa que, apontando para uma barra de sabão azul, diz *requin savon* ‘tubarão sabão’ (*shark soap* em Inglês) – a palavra complexa produzida é constituída por duas palavras francesas, mas a ordem das palavras é incorrecta nesta língua

É recorrente na literatura a afirmação de que, em várias línguas, a composição é o processo mais utilizado para formar novas palavras e que as crianças o fazem muito cedo (e.g. Booij, 2007; Nicoladis, 2006). É também frequente a sugestão de que a composição é um processo muito produtivo e frequente (e.g. Nicoladis, 2006), pelo menos nas línguas germânicas como o Inglês ou o Alemão, embora não tenhamos tido acesso a dados quantitativos que o permitam consubstanciar.

De acordo com Libben (2006), a composição pode ser considerada o processo universal fundamental de formação de palavras. Assumindo, que o objectivo da criação

de novas palavras é a comunicação, a composição oferece a mais fácil e efectiva forma de criar e transferir novos significados (Libben and Jarema, 2006).

Para Dressler (2006), os compostos são um dos objectos mais importantes na investigação da morfologia, pois eles estão presentes em todas as línguas do mundo, segundo o que está descrito nas gramáticas do mundo. E existem mesmo línguas, como o Chinês, onde a morfologia quase se esgota na composição (Libben & Jarema, 2006).

Quando uma criança começa a utilizar uma palavra como *candlestick* (castiçal), é pouco provável ela tenha conhecimento que esta palavra pode ser decomposta em dois constituintes *candle* ‘vela’ e *stick* ‘pau’. Quando as crianças crescem, aprendem os princípios das palavras compostas N-N (compostas por dois nomes), percebendo que estes compostos são uma combinação de duas partes: uma parte é considerada a *cabeça* (*stick*) e a outra o *modificador* (*candle*). Este é o momento no desenvolvimento em que a criança pode começar a criar novas palavras compostas (Nicoladis citada em Libben and Jarema, 2006).

Um entendimento generalizado na literatura sobre a aquisição de linguagem é que a compreensão precede a produção. Existem argumentos sugestivos de que os compostos não são excepção à regra geral. Durante a aquisição, existe uma coordenação gradual entre a compreensão e a produção, sendo que esta é precedida pela compreensão.

Vários estudos têm mostrado a precedência da compreensão sobre a produção. Por exemplo, para avaliar a compreensão de palavras compostas, foi realizado um estudo com crianças inglesas com menos de três anos. Pretendia-se que elas diferenciavam dois objectos que estavam interagindo e dois objectos que estavam apenas próximos, com o objectivo de, ao identificarem os objectos em interacção, compreenderem o significado de palavras compostas. Foi também efectuado um estudo com crianças suecas, em que se pretendia que identificassem um objecto correcto descrito por vinte e quatro palavras compostas por dois nomes. As crianças inglesas, obtiveram melhores resultados aos três anos do que aos quatro. Já as crianças suecas, revelaram melhores resultados à medida que a idade ia avançando, desde os 3 aos cinco anos de idade, sendo que o maior avanço foi verificado entre os 2;04 e os 3;00 anos. Estes resultados sugerem que, tanto no inglês como no sueco, as crianças compreendem as regras de subcategorização de compostos e que, apesar de poderem não produzir palavras compostas, compreendem-nas precocemente (Nicoladis citada em Libben and Jarema, 2006).

Clark (1995) refere que crianças com 2 anos ou sensivelmente mais novas, usam e formam novos nomes quando precisam deles. As formações mais precoces numa língua como o Inglês são geralmente compostas, contendo dois nomes. A partir desta idade (2 anos), as crianças começam a usar alguns sufixos derivacionais para construir novos nomes. O primeiro sufixo usado no Inglês parece ser, geralmente, o sufixo *-er* (Fletcher & MacWhinney, 1995).

Como se depreende do que foi dito acima, o facto de uma criança usar uma palavra complexa não implica, necessariamente, que ela a analise internamente. Efectivamente, as crianças podem usar palavras complexas de forma correcta morfológicamente, sem terem um conhecimento morfológico aprofundado, recuperando essas palavras armazenadas no seu léxico mental (Booij, 2007).

Por outro lado, as regras morfológicas devem ser descobertas com base nas palavras que foram formadas através dessas mesmas regras, sendo o primeiro passo na aquisição da morfologia a aquisição individual de palavras complexas.

De acordo com Clark (1995), a formação de palavras na criança implica a análise da palavra em partes (radicais e afixos), “cunhagem” de palavras para objectos, acções e propriedades. Quando as crianças estão a aprender a sua língua, começam por aprender palavras isoladas no fluxo do discurso, adquirindo-as à medida que as ouvem. Posteriormente, elas começam por analisar as partes das palavras, conseguindo identificar radicais, elementos flexionais e outros afixos (Fletcher & MacWhinney, 1995).

De seguida, as crianças terão que identificar potenciais significados. Nesta fase, elas começam a esboçar diferentes tipos de categorias de representações do mundo que as rodeia: categorias de objectos, acções, eventos, relações e propriedades.

Em terceiro lugar, devem traçar possíveis significados para as formas de palavras que identificaram anteriormente.

Como aponta Clark (1995), os primeiros significados das palavras infantis não coincidem, necessariamente, com os do adulto e, não sendo coincidentes, poderão necessitar de ajustamentos posteriores. Esses reajustamentos são, de facto, resultado do uso precoce de padrões de palavras que podem conter extensões, sobreposições parciais ou erros totais dos significados do adulto. Nesta fase - dos significados - as crianças parecem ser guiadas pelos princípios pragmáticos (Fletcher & MacWhinney, 1995).

Para adquirir a morfologia de nomes e verbos, as crianças devem primeiro analisar a estrutura das palavras a que estão expostas e, quando as ouvem, identificar

radicais e afixos, identificar significados consistentes para essas estruturas, e apenas depois começar a utilizar esses mesmos radicais e afixos em novas combinações. Este processo, que consiste em analisar a forma e em atribuir o significado, é um pré-requisito para a aquisição da morfologia (Spencer and Zwicky, 2001).

Ainda segundo Clark (1998), à medida que a criança aprende mais palavras e as memoriza e produz, é também capaz de analisar a sua estrutura morfológica interna. As crianças começam a identificar radicais e afixos dentro das palavras complexas tanto em compostos como em formas derivadas.

Parecem existir princípios básicos na aquisição de processos de formação de palavras: a) transparência – conhecer elementos com apenas uma correspondência entre forma e significado é mais transparente para construir e interpretar novas palavras do que elementos com mais do que uma correspondência; b) simplicidade – formas mais simples são mais facilmente adquiridas do que complexas; c) convencionalidade – nas crianças, a escolha das palavras vai-se adaptando ao uso da gramática convencional, isto é, à medida que crescem, as crianças vão necessitando de palavras que permitam exprimir o que é relevante no seu dia-a-dia; este princípio é motivado pela assumpção de que a criança recorre à formação de novas palavras quando têm uma lacuna lexical; d) produtividade – formas mais frequentes são mais usadas e mais produtivas do que aquelas que são menos utilizadas (Nicoladis citada em Libben and Jarema, 2006).

A “transparência” e previsibilidade de uma palavra composta estão muitas vezes relacionadas com a própria transparência da estrutura interna da sua língua. Por exemplo, em línguas com dois tipos diferentes de compostos, em que um é mais transparente e claro do que outro, aquele que é menos transparente é mais facilmente sujeito a modificações fonológicas e morfológicas. A perda de transparência (formal e interpretativa) pode ser vista num processo em que uma parte de um composto aparece como um afixo.

No entanto Fickkert (2001) observa que as estruturas menos complexas nem sempre são as mais facilmente apreendidas em relação a estruturas mais complexas. A autora argumenta que as estruturas que são mais facilmente apreendidas são aquelas que são mais claras para os falantes, independentemente da sua complexidade gramatical. Por exemplo, Fickkert observa que, na aquisição da estrutura prosódica de palavras e compostos monomorfémicos em nomes e compostos nominais no Holandês, as estruturas mais complexas serão adquiridas mais tarde do que palavras monomorfémicas com o mesmo tamanho, apesar de possuírem o mesmo tamanho, a

simplicidade/complexidade de palavras é um factor decisivo (Weissenborn & Hohle, 2001).

As características de cada língua e a natureza do *input* que as crianças ouvem, interagem com a transparência, simplicidade e produtividade e condicionam os padrões de desenvolvimento de formação de palavras nas crianças, ao longo da sua idade (Fletcher & MacWhinney, 1995). Por exemplo, em línguas em que as palavras compostas sejam altamente frequentes e produtivas, a frequência e a produtividade no *input* das crianças fazem salientar as regras para a formação deste tipo de palavras, reflectindo-se isso no desenvolvimento e produção precoce de compostos (Libben and Jarema, 2006).

Na realidade, existem diferenças na aquisição de padrões de palavras compostas nas várias línguas: enquanto no Inglês a produção deste tipo de palavras aparece aproximadamente aos dois anos de idade, no Francês eles são produzidos pelas crianças mais tarde, por volta dos três anos de idade. As razões para esta diferença podem residir em diferenciais de frequência que as formas compostas apresentam nas duas línguas. Efectivamente, segundo Nicoladis (2006) (citada em Libben and Jarema, 2006), os nomes compostos são frequentes e altamente produtivos no Inglês, sendo também produtivos nas crianças falantes desta língua já aos dois anos. Pelo contrário, na língua francesa, onde os nomes compostos são pouco frequentes, não sendo muito produtivos para as crianças falantes desta língua aos cinco ou seis anos de idade, as crianças raramente escolhem a produção de compostos, na criação de novas estruturas lexicais, usando apenas nomes compostos lexicalizados antes dos três anos; formam apenas novos compostos após os três anos de idade.

Interessantemente, num estudo de caso realizado com uma criança bilingue Inglesa-Francesa, a mesma autora observou a produção de novos compostos nas produções em Francês, antes dos três anos, sugerindo que a exposição a uma língua em que os compostos são frequentes, pode acelerar a aquisição do processo e/ou torná-lo mais produtivo numa outra língua, na qual este tipo de palavras é menos frequente.

Um outro caso interessante no quadro da avaliação do papel da frequência no desenvolvimento morfológico infantil é o do Hebreu. Nesta língua, os compostos têm uma frequência baixa. Porém, crianças que adquirem esta língua produzem este tipo de palavras, espontaneamente, aos 1;06 anos. Contudo, não é nada claro que as crianças compreendam que essas palavras sejam compostas. Isso mesmo é sugerido pelo facto de

estas crianças apenas usarem, espontaneamente, novas palavras por volta dos quatro anos de idade (Nicoladis, 2006 citada em Libben and Jarema, 2006).

Em suma, como recorrentemente apontado na literatura, os padrões de formação de palavras observados nas crianças e nos adultos dependem do que é produtivo em cada língua (e.g. Clark 1995): no Hebreu, os compostos não são produtivos e as crianças raramente produzem novos compostos antes dos seis, sete anos de idade e o mesmo é observado no Francês, onde novamente a composição de modo geral não é uma opção produtiva; estes dados contrastam com os que ocorrem nas línguas germânicas, onde o processo de composição é altamente produtivo, começando as crianças a adquirir esta língua a explorar esta opção mais precocemente (Fletcher & MacWhinney, 1995). Para Clark (1998), citado em Libben and Jarema (2006), o limite para a criatividade linguística das crianças, no que diz respeito aos compostos é a produtividade de formas no input da sua língua. Julgamos que esta observação é válida também para outras áreas da morfologia e da gramática mais em geral.

A questão do impacto da frequência no processo de aquisição e desenvolvimento linguístico é retomada na subsecção seguinte.

1.3 Importância da informação de frequência no input

Plunkett (1997) e Smith (1999) (citados em Herschensohn, 2007), apontam para a importância da frequência e saliência, presentes nas mais diversas teorias de aquisição de linguagem.

De acordo com uma corrente existente durante os anos 20, denominada *Associativismo* (Elman *et al.* 1996; MacWhinney 1999; MacWhinney and Bates 1989), toda a aprendizagem é desencadeada pelo input do meio ambiente em que a criança está inserida, criando associações específicas baseadas na frequência. Os três pontos mais fortes desta abordagem são: um mecanismo unitário de aprendizagem, frequência e saliência no *input* e o papel da interação. Este modelo, que é crucialmente dependente do *input* externo, defende que a mente começa como uma tábua rasa, que é moldada pelas interações estabelecidas. Esta teoria dá uma outra visão sobre a forma como se processa a aquisição de linguagem, não valorizando a predisposição inata da criança para a aquisição da linguagem (Herschensohn, 2007).

Os falantes usam cada palavra em quantidades variáveis nas suas produções. No que respeita às palavras com função gramatical, que pertencem a uma classe fechada, sabe-se que elas apresentam uma elevada frequência de ocorrência. Contudo, estas palavras ocorrem tipicamente em posições não-proeminentes do ponto de vista fonológico, sendo por isso pouco salientes.

Quanto às palavras lexicais, é sabido que palavras que ocorrem mais frequentemente são mais facilmente reconhecidas e retidas do que as de baixa frequência. Uma possível interpretação do efeito da frequência é a existência de uma correlação entre a frequência de exposição de uma palavra e o nível de activação dessa palavra no léxico mental – no processamento, uma palavra com um nível de activação mais elevado será mais rapidamente activada do que uma palavra com um nível de activação baixa (Booij, 2007).

Para além do que já foi dito, existem vários estudos que têm mostrado que a frequência pode desempenhar um papel importante no processo de aquisição e desenvolvimento da linguagem, em vários domínios da gramática, para além do morfológico (Vigário *et al.*, 2006).

A importância da informação de frequência para a compreensão do comportamento linguístico dos falantes, bem como do próprio processo de aquisição do sistema gramatical, tem sido mostrada num leque muitíssimo vasto de trabalhos recentemente, em diferentes campos.

Vigário *et al.* (2006), analisaram a distribuição dos formatos (tamanhos) da palavra na criança (CS) e no adulto, tendo sido estudados dois tipos de fala adulta: a dirigida à criança (CDS) e a entre adultos (ADS). Verificou-se que as produções iniciais das crianças não estão limitadas por restrições de palavra mínima, contrariamente ao que havia sido proposto para outras línguas, produzindo as crianças portuguesas palavras monossilábicas com sílaba aberta com frequência. Quando comparados os dados de emergência e frequência deste formato de palavras na criança, com os dados de frequência no discurso adulto, observou-se que as palavras com este formato são também bastante frequentes na fala adulta, em especial no CDS. Quanto aos restantes formatos, palavras dissilábicas e trissilábicas aparecem também pouco após as primeiras produções com valores de ocorrência altos. Os valores de frequência destes formatos de palavras na fala infantil não muito distante da frequência encontrada no discurso adulto; no que se refere aos valores de frequência das palavras maiores no CS, eles são mais próximos dos encontrados no AS do que do CDS. Efectivamente, o CS tem o dobro das

formas com três ou com mais de três sílabas do que o CDS. Pelo contrário, os padrões de frequência de monossílabos, trissílabos e palavras maiores, assemelham-se àqueles encontrados no AS. A partir destes resultados, Vigário *et al.* (2006) concluem que o AS é um melhor preditor das formas de palavras das crianças do que o CDS e que a frequência dos diversos formatos de palavras na fala da criança está altamente correlacionada com a frequência de formatos de palavras no discurso do adulto, sendo essa relação estatisticamente significativa.

A correspondência quase perfeita entre a frequência de distribuições de formas de palavras prosódicas nas crianças e no adulto (a adquirir o Português), é tanto mais interessante quanto as crianças que adquirem outras línguas, observando-se comportamentos muito diversos. De acordo com Vigário *et al.* (2006), em línguas como o Inglês tem sido observado que desde as primeiras produções as crianças não produzem as chamadas palavras *submínimas*, mantendo-se esse comportamento pelos vários estádios iniciais. O mesmo se verificou nas palavras com mais de duas sílabas, sendo a frequência destes formatos de palavra na língua adulta impossíveis, no caso das palavras mais pequenas do que um pé binário (como as que apresentam uma sílaba terminada em vogal breve), ou muito pouco frequentes, como acontece com as palavras com três ou mais sílabas. O mesmo tipo de correlação entre os dados de frequência e emergência dos vários tipos de palavra na criança e no adulto é encontrada também no Catalão (Prieto 2006), no Francês (Demuth e Johnson 2003), e também no Espanhol, língua que apresenta dados de frequência na fala adulta e dados relativos à emergência e frequência de formatos de palavra aproximáveis aos do Português Europeu (de acordo com os dados apresentados em Lleó, 2004, 2006).

Importa dizer que, para além do interesse que os dados revistos acima têm relativamente ao papel da frequência, eles são também informativos quanto à evolução do tamanho das palavras nas produções iniciais da criança. Esse aspecto é para nós relevante na medida em que a formação de novas palavras, que geralmente resulta da adição de morfemas, poderá estar condicionada pelo aparecimento ou tentativa de produção de palavras maiores, isto é, com um maior número de sílabas. Efectivamente, e como veremos no Capítulo 3, não existem palavras complexas monossilábicas, e são virtualmente inexistentes as palavras complexas dissilábicas.

Também relativamente ao desenvolvimento silábico, variados estudos incidindo sobre várias línguas do mundo têm mostrado que, embora os comportamentos verbais

das crianças sejam condicionados por restrições universais, durante o processo de aquisição, estas podem revelar diferentes percursos e estratégias em função das propriedades do *input*. Por exemplo, Fikkert e Freitas (1998) observam que as produções de crianças a adquirir o Holandês e o Português Europeu (PE) exibem diferenças na cronologia de emergência e domínio de estruturas silábicas com consoantes finais (em posição de Coda silábica). As autoras atribuem a maior rapidez no desenvolvimento de Codas preenchidas em crianças a adquirir o Holandês, quando comparadas com crianças a adquirir o PE, à maior frequência de Codas preenchidas nessa língua na fala adulta (Vigário, Martins & Frota, 2005). É interessante notar que, neste caso, tendo em consideração o grau de complexidade do constituinte no input, a Rima em Holandês é mais complexa do que no PE. Isto significa que, aqui, observa-se uma maior rapidez de aquisição no sistema gramatical mais complexo (Freitas *et al.* 2006). Pode por isso especular-se sobre a importância relativa da informação de frequência e da complexidade do sistema a adquirir.

O efeito acelerador da complexidade do *input* na velocidade de aquisição da estrutura silábica é também observado em crianças bilingues em Espanhol e Alemão, pois verificou-se que estas adquirem mais rapidamente as Codas em Espanhol do que as crianças espanholas monolingues. Estes resultados podem ser explicados pelo facto de as crianças bilingues estarem expostas a uma maior complexidade e frequência de Codas na sua outra língua, o Alemão (veja-se a revisão feita em Freitas *et al.*, 2006).

Num estudo sobre o percurso do desenvolvimento dos tipos silábicos do PE, Freitas *et al.* (2006), os autores avaliam a emergência dos diferentes tipos silábicos na fala da criança, relativamente à sua frequência na fala adulta. Com base na distribuição dos padrões silábicos na fala dos adultos e na fala dirigida às crianças, se apenas a frequência no *input* explicasse a evolução do comportamento linguístico inicial, os autores observam que a predição sobre a ordem de emergência de padrões silábicos do PE é a seguinte CV>V>CVC>CVG/CVN>CCV. Os resultados obtidos por Freitas *et al.* (2006), ao analisarem o discurso de três crianças, confirmam em larga medida esta predição: os formatos silábicos CV e V estão simultaneamente disponíveis desde o início da produção, contrariamente ao que seria esperado de acordo com a escala de complexidade silábica e sobretudo com os dados de frequência exibidos pelo adulto, fazendo destacar-se claramente o tipo CV relativamente a V; por outro lado os padrões CVG e CVN, que se seguem de imediato aos tipos CV e V, precedem (imediatamente) o padrão CVC, inversamente ao que foi predito, tendo em conta a frequência relativa

destes tipos silábicos no adulto. Estudadas as posições que estes tipos silábicos ocupam na fala adulta, Freitas *et al.* (2006) concluem que eles ocorrem em posições proeminentes. Os autores concluem, assim, que o desenvolvimento silábico nas crianças portuguesas não é exclusivamente determinado pela frequência dos vários padrões silábicos no *input* e que factores de natureza gramatical (como a ocorrência em posições proeminentes, designadamente acentuadas ou nas extremidades da palavra), terão que ser considerados na cronologia de eventos associada ao desenvolvimento silábico das crianças.

O papel da frequência está também atestado noutras áreas da aquisição da linguagem. Por exemplo, Kauschke e Klann-Delius (2007) mostram que os dados de desenvolvimento lexical infantil se correlacionam fortemente com a distribuição das palavras no *input*, havendo mesmo alterações no comportamento infantil que acompanham variações na distribuição lexical do *input* (Gülzow & Gagarina, 2007).

Há contudo, referências diversas a casos em que não há correspondência entre os dados de frequência no *input* e o padrão exibido pela criança. Por exemplo, Kupisch (2007) examinou de que forma a frequência de ocorrência de SN composto por nomes, sem determinantes no discurso dirigido às crianças, estava relacionada com o número de omissões de determinantes na linguagem das mesmas. Entre outros resultados, mostra-se que, embora a língua francesa exiba um número mais pequeno de substantivos neste tipo de construção sintáctica do que o Italiano, as crianças francesas omitem tantos determinantes quanto as crianças italianas. Esta autora concluiu que a frequência do *input*, isoladamente, não pode ser o factor que determina a aquisição mais rápida de determinantes. Existem outros factores, nomeadamente as características individuais de cada língua, que juntamente com a frequência influenciam a aquisição dos determinantes. Ainda na mesma linha, Bohnacker (2007) revela que, embora no Sueco seja mais frequente a ocorrência de nomes sem determinante do que no Alemão, as crianças suecas omitem menos determinantes do que as crianças alemãs, comparativamente, nas mesmas idades.

Para além destes, outros casos de ausência de correspondência entre os dados de frequência no adulto e o comportamento linguístico da criança são reportados em estudos compilados em Gülzow e Gagarina (2007).

1.3.1. Factores Externos

Tal como foi descrito anteriormente, e, apesar da frequência ter um papel muito importante na emergência de aquisição feita pelas crianças, não é a única influência na aquisição da linguagem. Um dos possíveis factores que pode influenciar a aquisição de linguagem é o estatuto social, económico e educacional dos seus pais.

Diversos estudos apontam para uma relação clara entre nível socioeconómico (educação e rendimentos dos pais) e o vocabulário das crianças.

Rowe (2008) fez um estudo com crianças com dois anos e seis meses e um ano mais tarde. Observou que o nível que o discurso dirigido à criança aos dois anos e meio prediz o seu vocabulário um ano mais tarde; o discurso que é dirigido à criança em relação ao nível socioeconómico é avaliado pelos rendimentos e educação.

Esta autora conclui que pais com uma educação e rendimentos mais elevados têm crianças com um nível de vocabulário mais diversificado, exibindo um crescimento mais rápido e precoce do vocabulário.

Esta relação estreita entre estes dois conceitos é devida ao discurso que os pais dirigem às suas crianças durante as interações do dia-a-dia. As mães com nível socioeconómico mais elevado, utilizam frases maiores e uma maior diversidade de palavras diferentes com as suas crianças, comparativamente com pais com baixo nível socioeconómico. De acordo com a mesma autora, estima-se que cada criança proveniente de meios socioeconómicos mais altos ouve aproximadamente 11000 frases por dia, comparativamente com crianças que provêm de meios socioeconómicos mais baixos, que ouvem cerca de 700.

Uma possível explicação para estas diferenças poderá ser o diferente conhecimento base sobre o desenvolvimento da criança, o que traduz em diferentes formas de comunicação com a mesma; outra possibilidade é a de que os pais de diferentes grupos têm diferentes estilos e tipos de linguagem; uma terceira possibilidade tem ainda a ver com as capacidades verbais dos pais: a linguagem e o nível de literacia podem influenciar a linguagem e conseqüentemente o discurso destes pais para com os filhos.

Naturalmente, importa notar que, mesmo que factores externos estejam na base das diferenças notadas, é sobre os dados linguísticos a que a criança estará exposta que eles se reflectem e é por esta via que esses factores podem, indirectamente, condicionar

o comportamento linguístico das crianças nas diversas fases do percurso de aquisição e desenvolvimento da linguagem.

II. METODOLOGIA

Neste capítulo, fazemos uma descrição do tipo de estudo desenvolvido, dos objectivos e hipóteses colocadas e dos *corpora* utilizados neste trabalho. É também realizada uma descrição dos critérios utilizados para a análise das bases de dados envolvidas.

2.1 Tipo de Estudo

Esta investigação é um estudo de caso, do tipo descritivo. Este tipo de estudos desenvolve uma investigação qualitativa, assumindo o papel de uma investigação particular.

O presente trabalho tem como principal objectivo analisar a emergência de palavras complexas no decurso do processo de aquisição e desenvolvimento linguístico da criança. Incidiu nas produções de uma criança entre os 1;01 e 3;01 anos de idade, constituídas numa base de dados pré-existente, de uso público.

Para além dos dados da criança, foi também estudado o discurso do adulto, com o objectivo de avaliar o impacto de um conjunto de aspectos caracterizadores da fala adulta no Português Europeu no comportamento linguístico inicial da criança.

2.2 Objectivos

São várias as questões que norteiam o presente trabalho. Desconhecendo-se no essencial como se processa a aquisição e desenvolvimento da morfologia derivacional no Português Europeu, procuraremos, antes de mais, responder às seguintes questões:

- Em que fase do desenvolvimento lexical a criança começa a produzir palavras complexas;
- Qual a ordem relativa de emergência de palavras derivadas e compostas;
- Nas palavras derivadas, quais os afixos que emergem em primeiro lugar;

- Quais as categorias de palavras que a criança utiliza para formar novas palavras complexas;
- Como se correlacionam a frequência de palavras no adulto e a frequência e emergência do mesmo tipo de palavras na criança (e.g. frequência/emergência de palavras simples/complexas; diferentes tipos de palavras complexas; afixos utilizados);
- Qual o papel do tamanho das palavras na emergência de palavras complexas (estará a produção de palavras complexas dependente do tamanho de palavra? E nesse caso, estará dependente do tamanho do que a criança *quer* produzir, o *alvo*, ou o tamanho do que efectivamente *consegue* produzir?).

2.3 Os dados da Criança (Base de Dados LumaLiDa)

Os dados da criança analisados nesta dissertação são os que integram a base de dados *LumaLiDa*, versão 1 (Frota, Vigário & Jordão (2008) *LumaLiDaOn*. Lisbon: Laboratório de Fonética da FLUL (<http://www.fl.ul.pt/laboratoriofonetica/lumalidaon.htm>). Esta base de dados é construída com os primeiros dados do Diário Linguístico de uma criança a adquirir o Português Europeu. O Diário é composto por transcrições fonéticas, quase diárias e *online*, do discurso infantil da criança feitas pelos dois familiares mais próximos, especialistas em fonética e fonologia, devidamente treinados. A versão 1 da base de dados tem 5191 entradas, correspondentes a enunciados (de uma ou mais palavras) produzidos entre os 01;01.26 e os 03;03.07 de idade¹ contendo um total de 17.229 palavras. O banco de dados encontra-se registado em *Access*.

Uma vez que a base de dados, *LumaLiDaOn*, está organizada por idade (anos e meses da criança) e meses do ano civil, para facilitar a visualização e análise de palavras, elaborou-se uma tabela de conversão entre o mês de produção e a idade da criança (e.g. todas as palavras produzidas pela criança em Novembro de 2004 correspondem à idade 1;00; todas as palavras produzidas em Dezembro de 2004 correspondem, à idade 1;01, ... todas as palavras produzidas em Fevereiro de 2007, correspondem à idade 3;03). Seguidamente apresentamos a conversão de idades em meses de produções:

1;00 – Novembro de 2004

1;01 – Dezembro de 2004
1; 02 – Janeiro de 2005
1;03 – Fevereiro de 2005
1;04 – Março de 2005
1;05 – Abril de 2005
1;06 – Maio de 2005
1;07 – Junho de 2005
1;08 – Julho de 2005
1;09 – Agosto de 2005
1;10 – Setembro de 2005
1;11 – Outubro de 2005
2;00 – Novembro de 2005
2;01 – Dezembro de 2005
2; 02 – Janeiro de 2006
2;03 – Fevereiro de 2006
2;04 – Março de 2006
2;05 – Abril de 2006
2;06 – Maio de 2006
2;07 – Junho de 2006
2;08 – Julho de 2006
2;09 – Agosto de 2006
2;10 – Setembro de 2006
2;11 – Outubro de 2006
3;00 – Novembro de 2006
3;01 – Dezembro de 2006
3;02 – Janeiro de 2007
3;03 – Fevereiro de 2007

Neste intervalo de tempo, a criança produziu, como referido acima, um total de 17.229 palavras (*tokens*), sendo o número de palavras diferentes produzidas 2079 (*types*). O banco de dados **LumaLiDaOn** (no formato *Excel*), é composto por várias colunas: (A) transcrição fonética do enunciado da criança; (B) transcrição ortográfica adaptada da transcrição anterior; (C) transcrição fonética do alvo; (D) transcrição ortográfica do alvo; (E) significado do enunciado; (F) contexto da produção; (G) data da

produção da criança; (H) a idade da criança. As colunas de (I) a (R) estão vazias e destinam-se a incluírem nova informação, de diferentes aspectos da gramática, e que são estudadas por diferentes pesquisadores (por enquanto, somente a coluna R inclui algumas informações, tais como o tipo de frase). Na coluna (S) são incluídos eventuais comentários.

A seguinte figura mostra a imagem de um extracto da base de dados tal como surge no monitor do computador.

	A	B	C	D	E	F	G	H
1	Utterance	Orthography	Target Utterance	Target Orthography	Meaning	Context	Date	Age
2	t"at6	tata	p"atu	pato	pato		19-Dez-04	0;01.26
3	d"a	dá	d"a	dá	dá		19-Dez-04	0;01.26
4	p"ap6	papa	p"atu	pato	pato		19-Dez-04	0;01.26
5	m6m"6	mamá	m"ar6	Mara	Mara		18-Jan-05	0;02.25
6	p"ap6	papa	"Dpl6	hopla	hopla (desenhos animados)		18-Jan-05	0;02.25
7	m6m6m6	mamama	m6m"6"	mamá	mamá		18-Jan-05	0;02.25
8	m6m6m6	mamama	mam"i	mami	mami		18-Jan-05	0;02.25
9	p"ap6	papa	p"ap6	papa	papa		18-Jan-05	0;02.25
10	m6m"6	mamá	m6m"6"	mamá	mamá		18-Jan-05	0;02.25
11	d"a	dá	d"a	dá	toma		18-Jan-05	0;02.25
12	dad"a	dadá	d"a	dá	toma		18-Jan-05	0;02.25
13	t"at6	tata	p"atu	pato	pato		18-Jan-05	0;02.25
14	t"(w)	tio	g"atu	gato	gato		18-Jan-05	0;02.25
15	"i	i	i:	hii	relinchar do cavalo		18-Jan-05	0;02.25
16	"a	a	"ag_w6	água	água		18-Jan-05	0;02.25
17	dad"a	dadá	d"a	dá	dá		18-Jan-05	0;02.25

Figura II.1. Extracto da Base de Dados LumaLiDaOn.

A partir deste ficheiro, criámos um novo documento (denominado *Léxico Mensal Target LumaLiDa*), também em formato Excel, dividido em várias páginas (“folhas”) para as diferentes idades da criança. A base de dados da criança (neste ficheiro) contém, assim, vinte e cinco folhas, correspondentes às diferentes idades (1;01; 1;02.....3;02; 3;03). Em cada uma destas folhas encontram-se duas colunas: na coluna (A) estão listadas todas as palavras diferentes (*types*) produzidas pela criança no período a que a folha se reporta; na coluna (B) aparece a frequência de cada uma dessas palavras, isto é o número de vezes que a palavra ocorreu (*tokens*). Os dados constantes nesta lista (listagem dos *types* e contagem de *tokens*) foram obtidos através da ferramenta electrónica *FreP* (Martins,Vigário e Frota 2004-2010), fornecido pelo Laboratório de Fonética no âmbito das linhas de acção em investigação (Fonética Forense, Recursos Linguísticos e Fonologia Teórica e Laboratorial).

Porque a ferramenta foi também utilizada por nós para determinar a frequência dos diferentes formatos (i.e. tamanhos) de palavra na criança e no adulto, descrevemo-la

sumariamente de seguida (veja-se para uma apresentação mais detalhada Vigário *et al.* 2005, 2006). O FreP é uma ferramenta electrónica que permite extrair automaticamente, a partir de textos escritos, informação sobre a frequência de unidades e padrões fonológicos desde o nível da palavra prosódica até ao traço. Entre outras valências, permite determinar (i) o número total de palavras prosódicas, clíticos, sílabas, segmentos, e também de palavras ortográficas e caracteres ortográficos; (ii) o número de palavras prosódicas com os diferentes tamanhos (tendo em conta o número de sílabas e tendo em conta o número de segmentos), o número de enclíticos e proclíticos fonológicos, o número dos diferentes tipos silábicos (CV, V, CVC...), e isso em função da posição na palavra e da presença/ausência de acento; (iii) o número de ocorrência dos diferentes segmentos ([p], [t]...), também tendo em conta as diferentes posições e a frequência dos diversos padrões acentuais em geral e em função do tamanho da palavra. Para além disso, as versões mais recentes da ferramenta listam todas as palavras diferentes que ocorrem no *corpus* analisado e o respectivo número de ocorrências. Finalmente, a ferramenta permite separar os dados relativos às contagens sobre palavras prosódicas e sobre clíticos, gerando um ficheiro de texto apenas com as palavras prosódicas do *corpus* analisado e outro só com os clíticos fonológicos.

Mais concretamente, os dados *type* correspondem ao número de palavras diferentes que a criança tentou produzir (*palavras-alvo*) pela criança, independentemente da sua frequência. Nos dados *token*, é considerada a frequência de cada *palavra-alvo* diferente produzida pela criança. Isto significa que na primeira coluna das folhas deste documento são listadas as diferentes palavras que a criança tenta produzir (e.g. *mãe, pai, sapato...*, mesmo que o diga de formas algo distintas das dos adultos, como *pato* para *sapato*), na segunda coluna é indicado quantas vezes a criança disse cada uma destas *palavras-alvo*.

A partir deste ficheiro original foi adicionado para cada idade um conjunto de informações resultante da análise de cada palavra. Foram adicionadas três colunas: na coluna (C) estão os dados relativos à estrutura interna da palavra; na (D), a classe gramatical a que pertencem; e na (E) os morfemas concretos que integra, no caso de ser uma palavras derivada. Importa dizer que, para o preenchimento da coluna (D) se teve em conta a base de dados *LumaLiDaOn*, a partir da qual se pode saber em que contexto foram produzidas as palavras e assim ter quase sempre pistas para determinar a categoria da palavra produzida, quando uma forma pode pertencer a mais de uma categoria.

A seguinte figura mostra a imagem de um extracto da base de dados *Léxico_Mensal_Target_LumaLiDa*, criada por nós, tal como surge no monitor do computador.

	A	B	C	D	E
1	Palavra (type)	Frequência	Processo Morfológico	Categoria Morfossintática	Morfemas incluídos
2	não	1	0	Adv	0
3	um	1	0	Outr	0
4	milho	1	0	N	0
5	bebé	1	5	N	0
6	Tito	1	0	N	0
7					
8	Número total de palavras únicas	5			
9					
10	Total frequência	5			

Figura II.2. Extracto da Base de Dados Léxico_Mensal_Target_LumaLiDa.

Como já indicado acima, para a listagem lexical foram consideradas as palavras *alvo*, independentemente dos formatos variáveis que as palavras podem assumir nas produções da criança nas diferentes idades. Os dados que estão na base desta contabilização são os que aparecem no ficheiro *LumaLiDaOn*, na coluna referente à transcrição ortográfica do alvo das produções da criança. Contudo, o *alvo* apenas foi tido em conta para efeitos da contabilidade das palavras efectivamente produzidas. Porque a coluna relevante também inclui palavras gramaticais que a criança nas fases iniciais tipicamente não produz, essa coluna contém palavras que não foram efectivamente produzidas. Por esta razão, procedemos a uma verificação sistemática das palavras constantes neste campo da base de dados, considerando as palavras produzidas na forma correcta ou de um modo suficientemente aproximado para se identificar a *palavra-alvo* e excluindo as que a criança não chegou a tentar produzir. Assim, por exemplo, se bem que os contextos de ocorrência de artigo *a* numa dada idade possam ter sido dez, se a criança apenas produziu essa forma, efectivamente, duas vezes, então o valor de frequência do artigo contabilizado foi 2.

As classes de palavras foram verificadas de acordo com o contexto em que cada palavra foi utilizada, de acordo com a classificação convencional, tendo como instrumentos auxiliares fundamentais Cunha e Cintra (1994), Villalva (2001), Mateus *et al.* (2004), o *Dicionário da Academia das Ciências de Lisboa* e ainda a base de dados *MorDebe*, disponibilizada no Portal da Língua Portuguesa.ⁱⁱ

Sequências envolvendo enclíticos foram separadas e as suas partes tratadas independentemente.

Com o intuito de dispormos de informação relativa à evolução dos formatos (tamanhos) de palavra, foram constituídos ficheiros de texto contendo as produções infantis por cada idade, sobre os quais a ferramenta FreP correu. Obtivemos deste modo o número de palavras dos diferentes tamanhos nas produções da criança em cada mês. Os dados da criança considerados para esta análise foram não apenas (i) as *palavras-alvo* mas também (ii) as palavras efectivamente realizadas pela criança. A versão do FreP usada para esta contagem é uma das mais recentes (versão 2046 de Setembro de 2009).

Os resultados obtidos através das produções realizadas pela criança fazem parte de um ficheiro comum, denominado *Resultados* e que contém todos os dados obtidos quer da criança quer do adulto. Este ficheiro em formato *Excel*, encontra-se dividido em folhas.

Os resultados da criança correspondem a três folhas: “Criança *type*”, “Criança *token*” que contém todos os resultados das produções da criança e “*Sílaba Criança*” que corresponde à informação do número de sílabas para cada palavra.

2.4 Os dados do Adulto – O Corpus do Português Falado

Com o objectivo de obtermos informação sobre as produções adultas comparável à analisada na criança, foi utilizada uma base de dados do adulto que contém uma lista de palavras que ocorrem no Corpus do Português Falado. Documentos Autênticos, editado em CR-ROM pelo Centro de Linguística da Universidade de Lisboa e Instituto Camões, CD 1, Portugal Gravações da década de 90. Estes dados são parte integrante da base de dados *FrePOP*, em construção no Laboratório de Fonética da FLUL (Frota, Vigarrio, Martins & Cruz, em curso).

Esta base de dados continha duas colunas: a primeira (A), referente a todas as diferentes palavras únicas (*types*) presentes neste *corpus*, num total de 3204 palavras, e a segunda (B), com a frequência de ocorrência dessas palavras (frequência de *tokens*), com um total de 20075 ocorrências.

A partir desta base de dados, foram adicionadas quatro novas colunas: na coluna (C) foi identificado o tipo de palavras quanto a processos morfológicos exibidos; na coluna (D) identificou-se a classe gramatical da palavra; a coluna (E) foi preenchida com os morfemas derivacionais concretos, no caso de a palavra ser derivada; e a coluna (F) inclui observações adicionais relevantes.

As classes de palavras foram todas verificadas com o auxílio dos instrumentos descritos na secção anterior (secção 2.3).

De seguida mostramos a imagem de um extracto da base de dados *LEXICOD1Potugalanos90Adulto*, criada por nós, tal como surge no monitor do computador.

	A	B	C	D	E	F
1	Palavra (type)	Frequência (token)	Processo Morf	Classe Gramatis	incluídos	Observações
2	que	1025	0 Adv,Outr		0	
3	e	742	0 N,Outr		0	
4	não	730	0 Adv,N		0	
5	a	689	0 N,Outr		0	Obs: 2 -a
6	é	685	0 V		0	
7	de	574	0 Outr		0	Obs: 2 -de
8	o	519	0 N,Outr		0	Obs: 3 -o

Figura II.3. Extracto da Base de Dados LEXICOD1Portugalanos90Adulto.

Tal como aconteceu no ficheiro correspondente com os dados da criança, e uma vez que no ficheiro original, verbos e enclíticos apareciam numa única entrada, estas sequências foram manualmente separadas, sendo as suas partes tratadas independentemente.

Quanto aos processos morfológicos, foram estabelecidos critérios de análise das bases de dados, descritos na secção seguinte.

Relativamente às classes gramaticais, foram colocadas na base de dados todas as categorias possíveis de cada palavra, uma vez que, neste caso, não tivemos acesso ao contexto em que cada uma foi utilizada e tal trabalho, dada a quantidade de dados, não seria, nesta fase, exequível. Por esta razão, a informação, referente às classes gramaticais de palavras, não foi utilizada na nossa análise. Contudo, pareceu-nos importante incluir essa informação complementar, pois poderá vir a revelar-se importante para o desenvolvimento de estudos futuros, dado que a base de dados ficará disponível para outros trabalhos.

Para a análise das classes gramaticais, utilizámos antes dados a que tivemos acesso através da Investigadora Amália Mendes, e que se encontram disponíveis em anexo (Anexo 2). Os dados podem ser consultados no ficheiro de *Resultados*. O *corpus* a partir do qual foi extraída a informação foi o *corpus* do Português incluído no C-ORAL-ROM (*Integrated Reference Corpora for Spoken Romance Languages*), um *corpus* multilingue de língua falada espontânea, para quatro línguas românicas (espanhol, francês, italiano e português). O *corpus* português foi desenvolvido pelo grupo RePort do Centro de Linguística da Universidade de Lisboa, em parceria com três instituições europeias (projecto financiado pela Comissão Europeia, IST – 2000 – 26228). O *corpus* foi editado em CD-ROM, a acompanhar um livro: CRESTI, Emanuela and Massimo Monegnia (eds.) (2005) C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages. Amsterdam/Philadelphia: John Benjamins Publishing Company, Studies in Corpus Linguistics nº 15 (com DVD).

Posteriormente a esta análise, foi também criado um ficheiro com todas as palavras simples e outro com todas as palavras complexas do discurso do adulto. A partir destes dois ficheiros foram determinados os valores de frequência dos diversos formatos de palavras em cada um dos ficheiros, assim se obtendo os valores relativos à frequência de palavras simples e de palavras complexas em função do seu tamanho (em número de sílabas). Para determinar o tamanho das palavras (prosódicas) em causa, usou-se a ferramenta FreP (versão 2046 de Setembro de 2009), já apresentada na secção precedente. Para este efeito, e dado poder entender-se que não constituem classes morfológicas claras quanto ao estatuto de palavras simples ou complexas, optou-se por excluir das contagens as palavras reduplicadas e as formas verbais no infinitivo, participípio passado e gerúndio.

Todos os resultados foram compilados no ficheiro Excel *Resultados* já referido na secção anterior. Como aí referido, este ficheiro está dividido em folhas, que incluem os diversos tipos de dados da criança e do adulto.

No que respeita aos resultados do adulto, foram criadas quatro folhas dentro do ficheiro Excel *Resultados*: 1^a) *Adulto type*, corresponde à análise feita dos processos morfológicos e afixos e são considerados tanto o número de palavras diferentes como a sua frequência; 2^a) *Adulto token* - corresponde aos resultados das categorias de palavras utilizadas pelo adulto; 3^a) *Afixos Adulto* – que disponibiliza todos os afixos concretos utilizados pelo adulto e a sua frequência; 4^a) *Sílabas Adulto* - o tamanho das palavras simples e complexas (em número de sílabas).

2.5 Critérios para a classificação das palavras

O trabalho classificatório que permitiu o preenchimento das duas bases de dados descritas anteriormente implicou o estabelecimento de critérios, que descrevemos sumariamente nesta subsecção.

Antes de mais, importa realçar que o objectivo desta classificação é a sua funcionalidade para efeitos de comparação entre os dados da criança e do adulto.

Lembramos, para além disso que, no que respeita à classificação morfológica dos dados da criança, foram consideradas as formas tendo em conta o que a criança visava produzir, isto é, o *alvo* (a forma no *input*), e não aquilo que efectivamente produziu. Em todos os casos em que as produções não são exactamente iguais aos alvos, contudo, existem semelhanças fonológicas que permitem identificá-los com as formas do adulto.

Para a contabilização dos tamanhos de palavra, contudo, foram consideradas não apenas os formatos das *palavras-alvo*, mas também os formatos das palavras efectivamente produzidas

Para a classificação morfológica, foram estabelecidos os seguintes critérios, na generalidade dos casos em conformidade com o que pode ser encontrado na literatura no domínio da morfologia do Português:

- 1) Relativamente à classificação quanto à estrutura morfológica interna (primeira coluna de cada uma das bases de dados (*LumaLiDaOn* e *LEXICODIPotugalanos90Adulto*), foram definidas as seguintes classes: (i) ausência de afixo, radical ou radical com marca de classe ou vogal temática (e.g. *que, não, casa, homem*) – código 0; (ii) simples com sufixo flexional, considerando o plural dos adjetivos e nomes e os morfemas de tempo-modo-aspecto e pessoa-número dos verbos (e.g. *peixes, podem, querem, havia*) – código 1; (iii) formas no Infinitivo, Particípio Passado e Gerúndio (e.g. *estudar, caída, trazer, chegado*) – código 11; (iv) formas com sufixação derivacional (e.g. *bastante, exactamente, tratamento*) – código 21; (v) formas com prefixação (e.g. *bicicleta, desculpa, triângulo*) – código 22; (vi) formas com sufixação derivacional e prefixação (mas não simultânea) (e.g. *desligado, engraçado*) – código 23; (vii) formas com sufixação derivacional e prefixação simultânea

(circunfixação) (e.g. *acalmo, rebola, empoleirada*) – código 24; (viii) derivação regressiva (e.g. *balanço*) – código 25; 3) forma com dois radicais ou palavras (e.g. *televisão, coca-cola, fotografia*) – código 3; (ix) formas com sufixos (z-) avaliativos (e.g. *chuchinha, bolinho, ursinha, cremezinho*) – código 4; (x) formas com reduplicação – consideraram-se aqui todas as instâncias de palavras constituídas pela repetição de sílabas, independentemente de existirem lexicalizadas no léxico adulto (ex: *chichi, cocó, chuchu, bebê, vovô, vovó, bibi*) – código 5; não se aplica se as formas repetidas corresponderem a palavras existentes (ex: *pisca-pisca, chupa-chupa, dói-dói*), pois nesses casos considerou-se que as formas pertencem à classe 3. Este tipo de palavras foi considerado como uma classe autónoma, apesar de a reduplicação não se tratar de um processo morfológico produtivo no Português Europeu. Contudo, pela sua especificidade relativamente ao restante vocabulário da língua, pela sua frequência na fala familiar e no discurso dirigido à criança, e pela sua frequência nas primeiras produções da criança (e.g. Correia 2009), julgámos útil considerar este tipo de palavras numa classe independente.

- 2) Quanto à classificação das palavras relativamente às suas categorias morfossintáticas (segunda coluna de cada uma das bases de dados), foram consideradas as seguintes classes: Nomes (código N) (e.g. *mamã, carro*); Adjectivos (código Adj) (e.g. *linda, verde*); Advérbios (código Adv) (e.g. *ali, ainda*); Verbos (código V) (*está, pára*); Palavras que são resultado da lexicalização de duas (ou mais) palavras, originando uma amálgama e em que uma das partes não pertence a nenhuma das classes anteriores (sendo aqui classificadas como do tipo Outr) (*dele, daquele*) (código Amlg-Outr); Palavras que são resultado da lexicalização de duas (ou mais) palavras, originando uma amálgama e em que uma das partes é uma palavra do tipo Adv (código Amlg-Adv) (*daqui, dali*); palavras de classe fechada com função gramatical e sem conteúdo lexical, incluindo as interjeições e algumas palavras tradicionalmente classificadas como advérbios, incluindo todas as expressões “QU”, independente da sua função adverbial, interrogativa, relativa ou outra, bem como preposições, conjunções, artigos, determinantes, numerais, etc. (código Outr). Essencialmente, as categorias com amálgamas e Outr incluem palavras que não são susceptíveis de entrar em processos de formação de palavras, pelo que se considerou supérfluo para os objectivos do presente estudo estabelecer

subclasses no interior delas. Isso permitiu minimizar os problemas que, inevitavelmente, se colocam às tarefas de classificação.

- 3) Para efeitos da contabilização dos processos originadores de palavras complexas, considerou-se que o tipo de processo morfológico exibido por uma palavra se correlaciona com sua composição interna morfológica. Concretamente, uma palavra constituída por um radical e um prefixo conta como uma instância de prefixação; uma palavra constituída por um radical, ao qual foram adicionados simultaneamente um prefixo e um sufixo, conta como uma instância de circunfixação; uma palavra constituída por um radical e um sufixo, conta como uma instância de derivação por sufixação; uma palavra constituída por um prefixo e um sufixo não adicionados simultaneamente, conta como uma instância de prefixação e uma instância de derivação por sufixação; uma palavra constituída por dois prefixos, conta como duas instâncias de prefixação; uma palavra com dois sufixos conta como duas instâncias de derivação por sufixação; tudo isto, independentemente de as palavras poderem não ser (inteiramente) morfológicamente transparentes.
- 4) A coluna (E) corresponde à listagem dos afixos concretos presentes nas palavras derivadas;
- 5) No caso da base de dados do adulto, em que não tivemos acesso ao contexto sintáctico das palavras, existiram problemas de classificação relativamente às classes gramaticais, envolvendo palavras homógrafas e homónimas. Uma vez que uma opção teria que ser tomada, optou-se sempre pela forma intuitivamente considerada mais provável no discurso do adulto (e.g. no caso de *colher*, que pode ter como categoria *nome* ou *verbo*, optámos pela categoria nominal). As formas com o sufixo *-do* (e.g. *casado*, *passado*) foram, deste ponto de vista, especialmente difíceis de classificar, uma vez que, frequentemente, cada forma pode corresponder à categoria adjectival, nominal ou participial. De modo a ultrapassar o problema, decidimos, mais uma vez, optar por uma classificação baseada no que intuitivamente é a forma mais frequentemente usada. Embora este seja um procedimento com algum grau de subjectividade, existem estudos que mostram que os sujeitos possuem e usam conhecimento sobre a frequência de ocorrência dos objectos da língua (e.g. Schmitt and Dunham, 1999; McCrostie, 2007).

- 6) Como é sabido, nem sempre é claro quando uma palavra complexa foi reanalisada como simples. A nossa opção foi considerar que uma palavra é complexa se é decomponível em partes, de forma a que, pelo menos um dos elementos tenha um formato e um significado aproximável a um morfema existente (e.g. *galheteiro*, *televisões*); isto, mesmo que outra parte não se aproxime de um morfema existente. Sempre que aconteceu (apenas) uma das partes não poder ser interpretada como um morfema existente, ou quando o significado já dificilmente se recupera a partir do significado dos morfemas exibidos pela palavra, a forma foi classificada como podendo ter perdido transparência (o código adoptado é PT, colocado na coluna das observações, a coluna F). Esta situação pode ocorrer, por exemplo, com palavras que foram formadas noutras fases da língua mas que podem ser tratadas pelos falantes também como complexas, dada a sua forma e significado. Por exemplo, na palavra *professor*, considerámos que a terminação é analisável como um sufixo (*or*), a partir da relação que se estabelece em termos formais e semânticos, com palavras como *canalizador*, *aviador*, *agricultor*, etc.; o mesmo acontecendo com a palavra *comunicativo*, formada já no Latim (cf. Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa, 2001): a palavra exhibe morfemas identificáveis no estado actual da língua, a partir da comparação com formas como *indicativo*, *facultativo*, *putativo*, etc.
- 7) No caso de palavras nas quais existiram dúvidas sobre a existência de sufixos derivacionais (verdadeiros diminutivos ou aumentativos) ou avaliativos, optou-se sistematicamente por esta última classificação.
- 8) Nos casos em que existem palavras que, em função do contexto, pertencem a diferentes categorias, a palavra aparece na coluna das palavras tantas vezes quantas as categorias diferentes.
- 9) Os sufixos derivacionais foram considerados sem a vogal de classe e o sufixo flexional, contrariamente à forma como são referidos normalmente nas gramáticas da língua, com o objectivo de facilitar as contagens do que pertence à mesma classe.
- 10) Como já foi dito atrás, nos dados da criança, não foram contabilizadas omissões e/ou substituições, tendo sido apenas considerado o que a criança produziu de facto.

- 11) Não considerámos as terminações *o/a* como uma categoria flexional mas sim como marcadores de classe. Esta distinção é apresentada e justificada por Villalva (1994, 2000) mas também por Huber (1933, 1986), que consideram que desaparecida a flexão casual, a flexão dos adjectivos e dos nomes se reduz à realização do plural.
- 12) Na base de dados do adulto, a flexão apenas foi contabilizada nas palavras simples; isto é, o que se pretendeu marcar foi, dentro das palavras simples, qual a proporção de palavras flexionadas e de palavras não-flexionadas.
- 13) Os compostos estrangeiros foram ignorados, no caso dos dados do adulto.
- 14) De acordo com Villalva (em Mateus et al. 2003) “bi e tri” são considerados prefixos de quantificação, pelo que as palavras “bicicleta ou triciclo” foram classificadas como derivadas por prefixação.

ⁱ (Na apresentação das idades os primeiros dígitos correspondem aos anos, os segundos aos meses e os terceiros aos dias de idade da criança.

ⁱⁱ A informação lexical que o Portal integra provém da MorDebe, uma base de dados morfológicos do português desenvolvida nos últimos quatro anos no ILTEC. Contém actualmente mais de 135000 palavras e integra um conjunto de informações relativas à morfologia das mesmas e à sua ortografia.

III. ANÁLISE DE RESULTADOS

Este capítulo tem como objectivo a apresentação e análise dos resultados. Encontra-se dividido em duas secções maiores: I) análise de resultados na criança e II) análise de resultados no adulto.

Cada uma destas secções inclui as mesmas subsecções. Iniciamos esta análise de resultados com a subsecção 3.1., sobre *Palavras simples versus palavras complexas*, que tem como objectivo identificar quais os tipos de palavras, simples ou complexas, que a criança e o adulto utilizam no seu discurso, qual a sua frequência e, no caso da criança, qual o seu padrão de emergência.

Na subsecção 3.2 mostramos e analisamos os resultados relativos aos *tipos de palavras em função da sua estrutura interna*, com o intuito de conhecer qual a constituição morfológica de palavras utilizadas quer pela criança, quer pelo adulto.

Seguidamente, na subsecção 3.3, *classes de palavras*, apresentamos os resultados relativos à frequência de ocorrência das diferentes classes de palavras utilizadas pela criança e pelo adulto com o objectivo, tendo em conta, no caso da criança, a evolução do comportamento da criança no tempo a este respeito.

Na subsecção 3.4, referente ao *tamanho de palavra simples e complexas*, analisamos o tamanho das palavras produzidas pelo adulto e pela criança e das *palavras-alvo* das produções da criança, em ambos os casos separando os resultados relativos às palavras simples dos referentes às palavras complexas.

Por fim, na secção 3.5 são analisados os *afixos mais utilizados* pela criança e pelo adulto.

III.I. Análise de Resultados na Criança

Antes de mais importa lembrar aqui que todos os dados da criança analisados correspondem às *palavras-alvo*, isto é, do que a criança pretende dizer, e não às suas efectivas produções, tal como já foi dito na secção 2.5 do capítulo *Metodologia*. Isto com a excepção da análise do tamanho da palavra em função de ser ou não uma forma complexa, caso em que se analisaram não apenas as *palavras-alvo*, mas também as palavras tal como efectivamente produzidas.

Os valores referem-se a contagens sobre o total de ocorrências das palavras (que designamos como *tokens*) e ao total de ocorrências de palavras únicas (que designamos *types*) – ver secção 2.3 no capítulo *Metodologia*.

I) 3.1 Palavras simples *versus* palavras complexas

Nesta subsecção procedemos à observação dos dados relativos à ocorrência de palavras simples e complexas nas produções da criança.

Para além das categorias *palavras simples* e *palavras complexas* foi ainda considerada a categoria *reduplicação*, que inclui palavras constituídas por duas sílabas idênticas. Tal como referido na secção 2.5 do capítulo *Metodologia*, não sendo um tipo de palavras resultante de um processo morfológico regular no Português (fala adulta), julgamos contudo, ser útil para os presentes propósitos diferenciar estes casos dos restantes incluídos na classe das palavras simples.

Os primeiros resultados que reportamos referem-se ao número de palavras simples e palavras complexas produzidas por esta criança, entre os 1;01 e os 3;03 anos.

Por uma questão de facilidade na visualização dos dados, os resultados são apresentados em três gráficos, em função da idade: no primeiro gráfico são apresentados os resultados relativos ao primeiro ano de idade, no segundo os relativos ao segundo ano, no terceiro os resultados relativos aos primeiros três meses do terceiro ano de vida da criança.

Os resultados aqui apresentados correspondem aos valores computados sobre *types*, isto é, sobre as palavras únicas produzidas pela criança e aos valores computados sobre *token*, isto é o total das palavras, incluído as várias ocorrências de cada item lexical.

O gráfico seguinte revela os tipos de palavras (simples, complexas e com reduplicação) utilizados pela criança, em valores absolutos, durante o primeiro ano de idade.

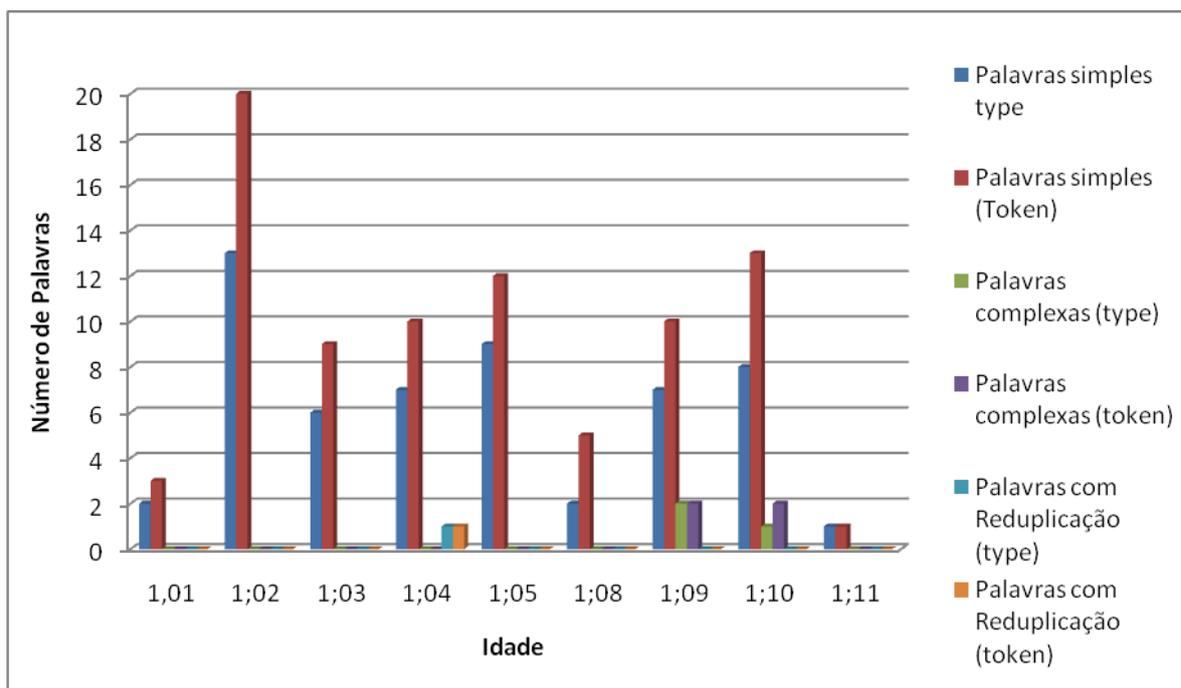


Figura III.1. Palavras simples, complexas e com reduplicação no 1º ano de idade – valores absolutos (type e token).

A seguinte tabela apresenta os mesmos dados do gráfico anterior (III.1), mas em valores percentuais.

Idade	Palavras simples (type)	Palavras simples (token)	Palavras complexas (type)	Palavras complexas (token)	Palavras com reduplicação (type)	Palavras com reduplicação (token)
1,01	100%	100%	0%	0%	0%	0%
1;02	100%	100%	0%	0%	0%	0%
1;03	100%	100%	0%	0%	0%	0%
1;04	87,50%	90,91%	0%	0%	12,50%	9,09%
1;05	100%	100%	0%	0%	0%	0%
1;08	100%	100%	0%	0%	0%	0%
1;09	77,78%	83,33%	22,22%	16,67%	0%	0%
1;10	88,89%	86,67%	11,11%	13,33%	0%	0%
1;11	100%	100%	0%	0%	0%	0%

Tabela III.1. Palavras simples, complexas e com reduplicação no 1º ano de idade – valores percentuais (type e token).

As palavras complexas são inexistentes nos primeiros meses das produções da criança e quando surgem as primeiras ocorrências (1;09 e 1;10) são em número diminuto. Em todo o período correspondente ao primeiro ano de idade, as palavras simples dominam o discurso da criança.

Podemos ainda observar que a criança utiliza a primeira palavra com reduplicação de sílabas aos 1;04 de idade, mas de modo episódico (apenas uma palavra, *bóing-bóing*), correspondendo a 12,50% do total de palavras naquele mês.

Observa-se ainda que produz as primeiras palavras complexas ao 1;09 (duas palavras (22,22% (type)) *obrigada* e *bivó* e ao 1;10A produz mais uma palavra complexa *bivó* (11,11% (type)).

Analisemos a distribuição de palavras simples, complexas e com reduplicação produzidas durante o segundo ano de idade. O gráfico (III.2) refere-se a valores absolutos e o tabela seguinte (III.2) refere-se aos valores percentuais.

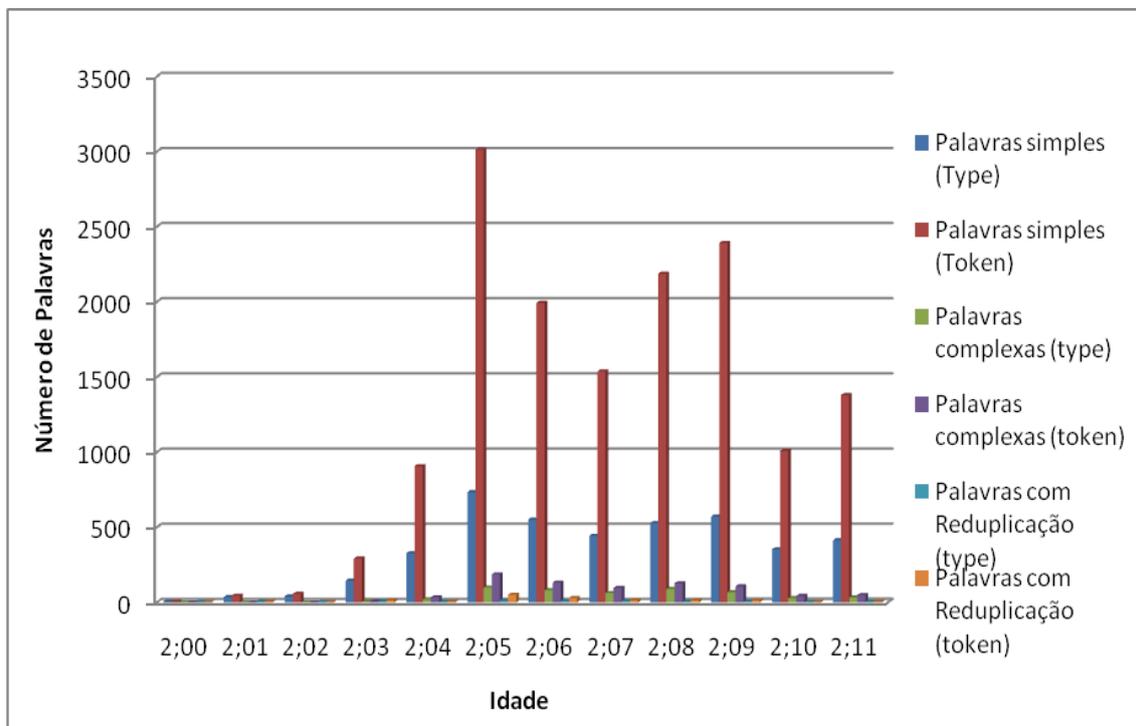


Figura III.2. Tipos de palavras no 2º ano de idade – valores absolutos (type/token).

Idade	Palavras simples (type)	Palavras simples (token)	Palavras complexas (type)	Palavras complexas (token)	Palavras com Reduplicação (type)	Palavras com Reduplicação (token)
2;00	80%	80%	0%	0%	20%	20%
2;01	86,05%	88,46%	0%	0%	13,95%	11,54%
2;02	95,35%	96,61%	0%	0%	4,65%	3,39%
2;03	91,72%	93,29%	3,18%	2,24%	5,10%	4,47%
2;04	93,45%	95,78%	5,70%	3,70%	0,85%	0,53%
2;05	86,86%	92,68%	11,72%	5,75%	1,42%	1,57%
2;06	85,98%	92,49%	12,62%	6,17%	1,40%	1,35%
2;07	86,05%	93,21%	12,21%	5,88%	1,74%	0,91%
2;08	84,75%	93,91%	14,45%	5,45%	0,80%	0,64%
2;09	88,68%	95,22%	10,54%	4,26%	0,78%	0,52%
2;10	91,93%	95,64%	7,55%	4,17%	0,52%	0,19%
2;11	91,80%	96,17%	7,32%	3,41%	0,89%	0,42%

Tabela III.2. Tipos de palavras no 2º ano de idade – valores percentuais (type/token).

Durante o segundo ano de idade da criança, observa-se que o número de palavras complexas mantém-se proporcionalmente muito abaixo do número de palavras simples.

Neste intervalo de tempo a criança utiliza palavras complexas, como *desculpa*, *fotografia*, *obrigada*, *desligar* (2;03); *chupa-chupa*, *bicho-da-conta*, *comichão*, *ouvido* (2;04); *rebolar*, *coca-cola*, *calções*, *ginástica* (2;08); *bombeira-doutora*, *simpática*, *triciclo*, *cuidado* (2;11).

Ao longo deste segundo ano de idade, o número de palavras complexas vai aumentando de acordo com o crescimento da criança. Pelo contrário, o número de palavras com reduplicação diminui ao longo do crescimento da criança e com o aumento do léxico, atingindo a frequência máxima aos 2;00. Nessa altura, a ocorrência de reduplicações atinge os 20% (token), mas esses valores estão associados apenas à produção de uma palavra única, *bebé*. Esta alta frequência está relacionada com a quantidade reduzida do léxico.

Tanto no caso das palavras simples como no caso das palavras complexas, embora em menor quantidade neste último tipo de palavras, existe uma grande diferença nas frequências de *types* e *tokens*, sugerindo que as palavras únicas são reutilizadas com alguma frequência, como, de resto, na fala adulta.

Seguidamente apresentamos os gráficos correspondentes aos resultados no terceiro ano de idade da criança. Mais uma vez, o primeiro gráfico (III.3) corresponde aos valores absolutos e a segunda tabela (III.3) aos valores percentuais.

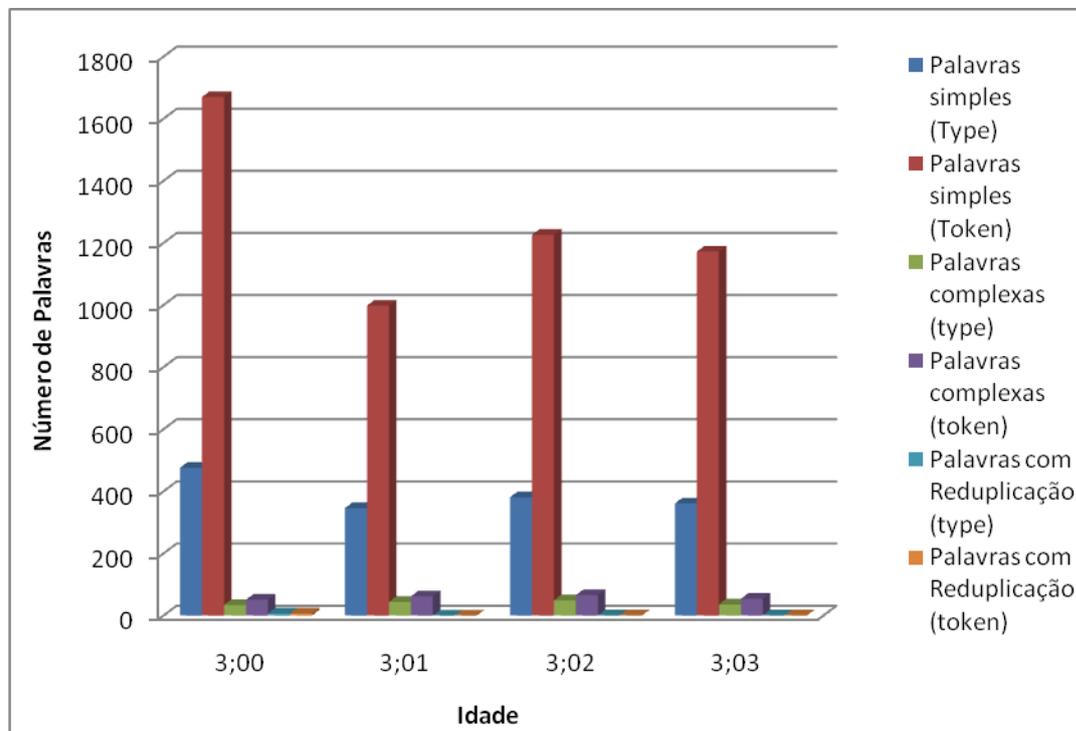


Figura III.3. Tipos de Palavras no 3º ano de idade – valores absolutos (type/token).

Idade	Palavras simples (type)	Palavras simples (token)	Palavras complexas (type)	Palavras complexas (token)	Palavras com Reduplicação (type)	Palavras com Reduplicação (token)
3;00	92,07%	96,64%	6,58%	2,95%	1,35%	0,41%
3;01	88,72%	94,15%	11,28%	5,85%	0%	0%
3;02	88,40%	94,82%	11,37%	5,10%	0,23%	0,08%
3;03	90,70%	95,52%	9,05%	4,40%	0,25%	0,08%

Tabela III.3. Tipos de Palavras no 3º ano de idade – valores percentuais (type/token).

Neste terceiro ano de idade da criança, em que as suas produções estão limitadas a apenas quatro meses, tal como anteriormente, a ocorrência de palavras complexas é muito diminuta. O mesmo é verificado para as palavras com reduplicação de sílabas.

Quanto às palavras complexas, nestas idades, apesar de se encontrarem em número inferior ao das palavras simples), elas apresentam valores absolutos superiores

aos apresentados em idades mais precoces (41, 44, 50 e 37 palavras complexas *type*, para as idades 3:00; 3:01; 3:02 e 3:03, respectivamente).

Em termos percentuais, tal como no período anteriormente considerado, as palavras complexas correspondem a uma média de 10% (valor arredondado de 9,57%) do léxico mensal total da criança (*type*), sendo esse valor bastante inferior quando considerados os *tokens* (em média 4,5%).

Os valores computados, tanto para *type* como para *token* aos três anos de idade, são muito semelhantes aos encontrados aos dois anos para o mesmo tipo de dados. De acordo com *type*, as palavras complexas atingem uma média de 9.48% (aos dois anos) e 9.57% (aos três anos); quanto aos dados *token*, uma média 4.5% aos dois e três anos.

Da observação dos dados salienta-se a evolução da dimensão do léxico da criança no tempo. Como é de esperar, há um aumento bastante significativo do número de palavras e da sua frequência, em termos de *token e type*, à medida que a criança cresce. Enquanto aos 2;00 a criança produz 4 palavras simples e 1 complexa (*type/token*), passados três meses este valor sobe para as 144 palavras simples e 5 complexas (*type*) e 292 palavras simples e 7 complexas (*token*). Aos 2;05 produz 734 palavras simples e 111 palavras complexas (*type*), e 3014 palavras simples e 238 complexas (*token*), atingindo nesta idade o maior número de palavras simples e complexas (*token*) no período observado. A partir desta idade, o número de palavras simples e complexas diminui subtilmente, sendo que aos 2;09 produz 572 palavras simples e 68 palavras complexas (*type*), e 2392 palavras simples e 107 complexas (*token*). Na última idade observada (3;03), a criança produziu 361 palavras simples e 36 palavras complexas (*type*), e 1172 palavras simples e 54 complexas (*token*),

I) 3.2 Tipos de palavras em função da sua estrutura

Analisamos as palavras simples e complexas realizadas pela criança, de acordo com a sua estrutura interna, na qual podem surgir morfemas de diferentes tipos, utilizados na formação de novas palavras.

Esta secção analisa os resultados do uso de palavras pela criança que evidenciam diferentes processos morfológicos. Tal como referido na secção 2.5 do capítulo *Metodologia*, as palavras foram codificadas quanto aos processos de formação da

seguinte forma – 0: palavras formadas por radical ou radical com marca de classe ou vogal temática; 1: palavras simples com sufixo flexional; 11: formas no Infinitivo, Particípio Passado e Gerúndio; 21: formas com sufixação derivacional; 22: formas com prefixação; 23: formas com sufixação derivacional e prefixação (mas não simultânea); 24: formas com sufixação derivacional e prefixação simultânea (circunfixação); 25: derivação regressiva; 3: forma com dois radicais/palavras; 4: formas com sufixos avaliativos; 5: formas com reduplicação.

Os quadros que se seguem correspondem à distribuição de palavras únicas (*types*) em função da sua estrutura interna, quer em valores absolutos como em valores percentuais.

Idade	0	1	11	2	21	22	23	24	25	3	4	5
1;01	1	1										
1;02	12	1										
1;03	6											
1;04	6	1										1
1;05	8	1										
1;08	2											
1;09	5	1	1	2	1	1						
1;10	8			1		1						
1;11	1											
2;00	4											1
2;01	36	1										6
2;02	34	7										2
2;03	113	30	1	3	1	2				2		8
2;04	231	92	5	5	5					6	9	3
2;05	458	252	24	23	19	4				11	60	12
2;06	322	181	49	29	26	3			1	8	42	9
2;07	279	123	42	29	23	6				6	27	9
2;08	297	171	60	52	46	2	2	2		7	30	5
2;09	329	187	56	35	30	4		1		7	26	5
2;10	203	121	29	16	10	4		2		2	11	2
2;11	243	137	34	15	12	3				4	14	4
3;00	235	195	46	12	11	1				7	15	7
3;01	201	120	25	17	14	3				5	20	
3;02	192	161	28	21	15	6				7	21	1
3;03	185	142	34	14	10	3		1		7	15	1

Tabela III.4. Tipo de palavras em função da estrutura interna (valores absolutos) - Criança (*type*).

Idade	0	1	11	2	21	22	23	24	25	3	4	5
1,01	50%	50%										
1,02	92,31%	7,69%										
1,03	100%											
1,04	75%	12,50%										12,50%
1,05	88,89%	11,11%										
1,08	100%											
1,09	55,56%	11,11%	11,11%	22,22%	11,11%	11,11%						
1,10	88,89%					11,11%						
1,11	100%											
2,00	80%											20%
2,01	83,72%	2,33%										13,95%
2,02	79,07%	16,28%										4,65%
2,03	71,97%	19,11%	0,64%	1,91%	0,64%	1,27%				1,27%		5,10%
2,04	65,81%	26,21%	1,42%	1,42%	1,42%					1,71%	2,56%	0,85%
2,05	54,20%	29,82%	2,84%	2,72%	2,25%	0,47%				1,30%	7,10%	1,42%
2,06	50,16%	28,19%	7,63%	4,68%	4,05%	0,47%			0,16%	1,25%	6,54%	1,40%
2,07	54,07%	23,84%	8,14%	5,62%	4,46%	1,16%				1,16%	5,23%	1,74%
2,08	47,67%	27,45%	9,63%	8,34%	7,38%	0,32%	0,32%	0,32%		1,12%	4,82%	0,80%
2,09	51,01%	28,99%	8,68%	5,43%	4,65%	0,62%		0,16%		1,09%	4,03%	0,78%
2,10	52,86%	31,51%	7,55%	4,16%	2,60%	1,04%		0,52%		0,52%	2,86%	0,52%
2,11	53,88%	30,38%	7,54%	3,33%	2,66%	0,67%				0,89%	3,10%	0,89%
3,00	45,45%	37,72%	8,90%	2,32%	2,13%	0,19%				1,35%	2,90%	1,35%
3,01	51,54%	30,77%	6,41%	4,36%	3,59%	0,77%				1,28%	5,13%	
3,02	44,55%	37,35%	6,50%	4,87%	3,48%	1,39%				1,62%	4,87%	0,23%
3,03	46,48%	35,68%	8,54%	3,51%	2,51%	0,75%		0,25%		1,76%	3,77%	0,25%

Tabela III.5. Tipo de palavras em função da estrutura interna (valores percentuais) - Criança (*type*).

O discurso da criança, nas idades mais precoces é marcado por palavras simples, sem afixo ou com sufixo flexional.

Podemos verificar que aos 1;04 anos esta criança usa a primeira palavra com reduplicação *bóing-bóing*, correspondendo a 12,50% (*type*) e 9,09% (*token*) do seu léxico mensal e que ainda não usa palavras complexas. A partir dos dois anos de idade, a criança utiliza sempre este tipo de palavras (e.g. *piu-piu* aos 2;02; *pimpim* aos 2;06; *bibi* aos 2;09; *chuchu* aos 3;03), com exceção da idade 3;01 onde não produz nenhuma destas palavras. A utilização deste tipo de palavras vai diminuindo à medida que a criança cresce, tomando os valores mais baixo quando atinge o terceiro ano de idade.

As formas no Infinitivo, Particípio Passado e Gerúndio surgem aos 1;09, *marchar*. Contudo, a produção sistemática deste tipo de palavras inicia-se apenas aos 2;03 (*arranjar* aos 2;03; *caindo* aos 2;07; *dormir* aos 2;11; *comprar* aos 3;02). A evolução do uso deste tipo de palavras ao longo da idade da criança é positiva, pois a criança produz mais palavras deste tipo à medida que cresce.

Aos 1;09 anos produz as primeiras palavras complexas e utiliza simultaneamente derivação por sufixação *obrigada* e por prefixação *bivó*.

A partir dos 2;03 anos começa a produzir palavras complexas através de composição, de uma forma sistemática, apresentando palavras com esta estrutura em todas as idades seguintes (e.g. *fotografia* aos 2;03; *saco-de-dormir* aos 2;05; *coca-cola* aos 2;07; *jardim zoológico* aos 2;11; *televisão* aos 3;03).

Da mesma forma, as palavras derivadas são produzidas pela criança em todas as idades a partir dos 2;03 anos (e.g. *desculpa* aos 2;03; *portão* aos 2;06; *ocupado* aos 2;10; *ventania* aos 3;01).

A partir da idade de dois anos e quatro meses a criança usa as primeiras palavras com sufixação (z-)avaliativa, e tal como se tem verificado anteriormente para os restantes tipos de palavras, utiliza-as em todas as idades subsequentes (e.g. *fimbinho* aos 2;04; *comidinha* aos 2;07; *passarinho* aos 3;00; *pequeninos* aos 3;03).

Relativamente às palavras derivadas, existe uma maior frequência das palavras sufixadas em relação às prefixadas, mas, como foi dito, estes dois tipos de palavras surgem na mesma idade (1;09) e são produzidos, em ambos os casos, de forma constante pela criança a partir dos 2;03. As ocorrências de formas com sufixação derivacional e prefixação, mas não simultânea (23) (e.g. *desligado* aos 2;08) e as formas com sufixação derivacional e prefixação simultânea, circunfixação (24), (e.g. *esfomeado* aos 2;09) são raras e tardias.

Os dois gráficos seguintes apresentam os resultados das palavras cuja estrutura interna evidencia mais do que um processo morfológico, quer em valores absolutos, quer em valores percentuais (*type*).

Idade	5+4	3+11	22+11	3+4
2;05	5			
2;06	1			
2;07		1		
2;08			1	
3;01	1			1

Tabela III.6. Ocorrência de palavras cuja estrutura interna revela mais de um processo morfológico (valores absolutos) - Criança (*type*).

Idade	5+4	3+11	22+11	3+4
2;05	0,59%			
2;06	0,16%			
2;07		0,19%		
2;08			0,16%	
3;01	0,26%			0,26%

Tabela III.7. Ocorrência de palavras cuja estrutura interna revela mais de um processo morfológico (valores percentuais) - Criança (*type*).

A partir da idade de dois anos e cinco meses a criança usa as primeiras palavras com reduplicação e afixação (sufixação (z-)avaliativa) *bebezinhas* (2;05); *bebezinha* (2;06) e *piu-piuzinho* (3;01).

Aos 2;07 utiliza uma palavra composta flexionada *telefonar* e aos 2;08 uma palavra derivada flexionada *desmontada*. Aos 3;01 usa uma palavra composta e com sufixação (z-)avaliativa *pontapezinho*.

De um modo geral, e como visto na secção precedente, a criança produz predominantemente palavras simples. Entre as palavras complexas que produz, as palavras derivadas emergem mais cedo.

A distribuição dos diferentes tipos de palavras complexas é apresentada no quadro seguinte, em valores absolutos e valores percentuais - dados *type*.

Idade	Derivação	Composição	Sufixação (z-) Avaliativa
1;01			
1;02			
1;03			
1;04			
1;05			
1;08			
1;09	2		
1;10	1		
1;11			
2;00			
2;01			
2;02			
2;03	3	2	
2;04	5	6	9
2;05	23	11	60
2;06	30	8	42
2;07	29	6	27
2;08	52	7	30
2;09	35	7	26
2;10	16	2	11
2;11	15	4	14
3;00	12	7	15
3;01	17	5	20
3;02	21	7	21
3;03	14	7	15

Tabela III.8. Distribuição das palavras resultantes de processos morfológicos (derivação, composição e sufixação (z-)avaliativa) (valores absolutos) - Criança (type).

Idade	Derivação	Composição	Sufixação (z-) Avaliativa
1;01			
1;02			
1;03			
1;04			
1;05			
1;08			
1;09	22,22%		
1;10	11,11%		
1;11			
2;00			
2;01			
2;02			
2;03	1,91%	1,27%	
2;04	1,42%	1,71%	2,56%
2;05	2,72%	1,30%	7,10%
2;06	4,68%	1,25%	6,54%
2;07	5,62%	1,16%	5,23%
2;08	8,34%	1,12%	4,82%
2;09	5,43%	1,09%	4,03%
2;10	4,16%	0,52%	2,86%
2;11	3,33%	0,89%	3,10%
3;00	2,32%	1,35%	2,90%
3;01	4,36%	1,28%	5,13%
3;02	4,87%	1,62%	4,87%
3;03	3,51%	1,76%	3,77%

Tabela III.9. Distribuição das palavras resultantes de processos morfológicos (derivação, composição e sufixação (z-)avaliativa) (valores percentuais) - Criança (type).

Dentro das palavras complexas, as formas mais utilizadas pela criança vão variando, de acordo com a idade, entre as que exibem morfologia derivacional e sufixação (z-)avaliativa, que apresentam quase sempre valores aproximados: aos 2;03 e entre os 2;07 e os 2;11 anos predominam as palavras derivadas; dos 2;04 aos 2;06 e dos 3;00 aos 3;03 anos a criança usa mais palavras com sufixação (z-)avaliativa. Regista-se, mesmo assim, a ocorrência mais precoce de palavras derivadas aos (1;09 e 1;10) e um maior destaque no número de palavras com sufixação (z-)avaliativa quando comparada com a encontrada no período dos 2;04 aos 2;06.

As palavras compostas são em geral claramente menos frequentes do que os dois restantes tipos de palavras. Contudo, elas ocorrem na mesma altura em que

aparecem as palavras que exibem os outros dois processos de formação de palavras de modo constante, ocorrendo em todos os meses a partir desta altura.

Nos últimos quatro meses observados, as palavras complexas atingem uma média de 9,44% (*type*), sendo as palavras derivadas 3,77%, as palavras compostas 1,50% e as palavras com sufixação (z-) avaliativa 4,17% do total de palavras produzidas.

Os quadros que se seguem correspondem às ocorrências totais de palavras (*token*) em função da sua estrutura interna, quer em valores absolutos como em valores percentuais.

Idade	0	1	11	2	21	22	23	24	25	3	4	5
1;01	2	1										
1;02	16	4										
1;03	9											
1;04	7	3										1
1;05	10	2										
1;08	5											
1;09	7	2	1	2	1	1						
1;10	13			2		2						
1;11	1											
2;00	4											1
2;01	45	1										6
2;02	48	9										2
2;03	245	46	1	4	1	3				3		14
2;04	663	239	5	8	8					14	13	5
2;05	2031	929	54	41	34	7				28	112	51
2;06	1309	580	105	44	37	6			1	17	71	29
2;07	1051	391	96	50	44	6				13	33	15
2;08	1512	546	130	78	69	4	2	3		7	41	15
2;09	1636	622	134	53	47	5		1		10	44	13
2;10	697	265	47	28	16	10		2		2	14	2
2;11	906	412	62	20	14	6				6	23	6
3;00	1047	525	98	20	17	3				9	22	7
3;01	657	292	49	25	17	8				8	27	
3;02	761	408	57	32	19	13				7	27	1
3;03	717	381	74	21	12	8		1		12	21	1

Tabela III.10. Tipo de palavras em função da estrutura interna (valores absolutos) - Criança (*token*).

Idade	0	1	11	2	21	22	23	24	25	3	4	5
1;01	66,67%	33,33%										
1;02	80,00%	20,00%										
1;03	100%											
1;04	63,64%	27,27%										9,09%
1;05	83,33%	16,67%										
1;08	100%											
1;09	58,33%	16,67%	8,33%	16,66%	8,33%	8,33%						
1;10	86,67%			13,33%		13,33%						
1;11	100%											
2;00	80%											20%
2;01	86,54%	1,92%										11,54%
2;02	81,36%	15,25%										3,39%
2;03	78,27%	14,70%	0,32%	1,28%	0,32%	0,96%				0,96%		4,47%
2;04	70,01%	25,24%	0,53%	0,84%	0,84%					1,48%	1,37%	0,53%
2;05	62,45%	28,57%	1,66%	1,27%	1,05%	0,22%				0,86%	3,44%	1,57%
2;06	60,71%	26,90%	4,87%	2,05%	1,72%	0,28%			0,05%	0,79%	3,29%	1,35%
2;07	63,70%	23,70%	5,82%	3,03%	2,67%	0,36%				0,79%	2,00%	0,91%
2;08	64,89%	23,43%	5,58%	3,35%	2,96%	0,17%	0,09%	0,13%		0,30%	1,76%	0,64%
2;09	65,13%	24,76%	5,33%	2,11%	1,87%	0,20%		0,04%		0,40%	1,75%	0,52%
2;10	66,07%	25,12%	4,45%	2,66%	1,52%	0,95%		0,19%		0,19%	1,33%	0,19%
2;11	63,14%	28,71%	4,32%	1,40%	0,98%	0,42%				0,42%	1,60%	0,42%
3;00	60,59%	30,38%	5,67%	1,15%	0,98%	0,17%				0,52%	1,27%	0,41%
3;01	61,98%	27,55%	4,62%	2,35%	1,60%	0,75%				0,75%	2,55%	
3;02	58,86%	31,55%	4,41%	2,48%	1,47%	1,01%				0,54%	2,09%	0,08%
3;03	58,44%	31,05%	6,03%	1,71%	0,98%	0,65%		0,08%		0,98%	1,71%	0,08%

Tabela III.11. Tipo de palavras em função da estrutura interna (valores percentuais) - Criança (token).

Tendo em consideração os valores de frequência computados sobre *token*, os resultados relativos ao tipo de palavra em função da estrutura interna são similares aos apresentados tendo em conta a contagem sobre *types*, sendo que as palavras simples, sem afixos ou com sufixo flexional, prevalecem no discurso da criança. As palavras complexas são sempre em número diminuto.

Os dados vistos mostram, assim, que o padrão de evolução dos diferentes tipos de palavras ao longo da idade é semelhante nos dados contados sobre *types* e sobre *tokens*. O mesmo não se pode dizer, contudo, dos valores de frequência dos diversos tipos de palavras nestes dois tipos de contagem.

Os dados percentuais desta categoria de palavras em *type* são sempre mais elevados do que em *token*, em todas os tipos de palavras: formas com Particípio Passado e Gerúndio; palavras derivadas e compostas; palavras com sufixação (z-) avaliativa, e formas reduplicadas apresentam valores de *tokens* proporcionalmente inferiores aos de *types*.

Nas tabelas seguintes apresentamos os valores de ocorrência das palavras obtidas pela aplicação de mais de um processo morfológico, quer em valores absolutos, quer em valores percentuais (*token*).

Idade	5+4	3+11	22+11	3+4
2;05	6			
2;06	1			
2;07		1		
2;08			1	
3;01	1			1

Tabela III.12. Ocorrência de palavras cuja estrutura interna revela mais de um processo morfológicos (valores absolutos) - Criança (*token*).

Idade	5+4	3+11	22+11	3+4
2;05	0,18%			
2;06	0,05%			
2;07		0,06%		
2;08			0,04%	
3;01	0,09%			0,09%

Tabela III.13. Ocorrência de palavras cuja estrutura interna revela mais de um processo morfológicos (valores percentuais) - Criança (*token*).

Os resultados aqui mostrados são idênticos aos relativos à contagem sobre *types*. A partir da idade de dois anos e cinco meses a criança usa as primeiras palavras com reduplicação e afixação (sufixação (z-)avaliativa), aos 2;07 utiliza uma palavra composta flexionada e aos 2;08 uma palavra derivada flexionada. Aos 3;01 usa uma palavra composta e com sufixação (z-)avaliativa.

De seguida, são apresentadas as distribuições dos diferentes tipos de palavras complexas, em valores absolutos e valores percentuais - dados *token*.

Idade	Derivação	Composição	Sufixação (z-) Avaliativa
1;01			
1;02			
1;03			
1;04			
1;05			
1;08			
1;09	2		
1;10	2		
1;11			
2;00			
2;01			
2;02			
2;03	4	3	
2;04	8	14	13
2;05	41	28	112
2;06	44	17	71
2;07	50	13	33
2;08	78	7	41
2;09	53	10	44
2;10	28	2	14
2;11	20	6	23
3;00	20	9	22
3;01	25	8	27
3;02	32	7	27
3;03	21	12	21

Tabela III.14. Distribuição das palavras resultantes de processos morfológicos de formação de palavras (derivação, composição e sufixação (z-)avaliativa) (valores absolutos) - Criança (*token*).

Idade	Derivação	Composição	Sufixação (z-) Avaliativa
1;01			
1;02			
1;03			
1;04			
1;05			
1;08			
1;09	16,66%		
1;10	13,33%		
1;11			
2;00			
2;01			
2;02			
2;03	1,28%	0,96%	
2;04	0,84%	1,48%	1,37%
2;05	1,27%	0,86%	3,44%
2;06	2,05%	0,79%	3,29%
2;07	3,03%	0,79%	2,00%
2;08	3,35%	0,30%	1,76%
2;09	2,11%	0,40%	1,75%
2;10	2,66%	0,19%	1,33%
2;11	1,40%	0,42%	1,60%
3;00	1,15%	0,52%	1,27%
3;01	2,35%	0,75%	2,55%
3;02	2,48%	0,54%	2,09%
3;03	1,71%	0,98%	1,71%

Figura III.15. Distribuição das palavras resultantes de processos morfológicos de formação de palavras (derivação, composição e sufixação (z-)avaliativa) (valores percentuais) - Criança (token).

Relativamente às palavras complexas, os resultados do processo mais recorrente nas palavras utilizadas pela criança variam de acordo com a idade tal como foi observado para os valores *type*, entre as palavras derivadas e as palavras com sufixação (z-)avaliativa, que apresentam valores muito idênticos.

Aos 2;03 e entre os 2;07 e os 2;10 anos a criança usa mais palavras derivadas; dos 2;04 aos 2;06 e dos 2;11 aos 3;01 predominam as palavras com sufixação (z-)avaliativa; aos 3;02 voltam a predominar as palavras derivadas e aos 3;03 estes dois tipos de palavras apresentam valores iguais.

Diferenciando os dados *type* e *token*, observamos que os valores das palavras derivadas por prefixação não revelam uma diferença tão significativa quanto as restantes palavras nestes dois tipos de computação dos valores de frequência.

Tal como foi observado com base nos *types*, as palavras complexas são claramente menos frequentes, encontrando-se em número diminuto em relação às palavras simples.

Seguidamente é apresentado um gráfico que compara os dados *type* e *token*, relativamente à distribuição de palavras complexas, em valores absolutos e percentuais.

	Derivação		Composição		Sufixação (z-) avaliativa	
	<i>type</i>	<i>token</i>	<i>type</i>	<i>token</i>	<i>type</i>	<i>token</i>
1,01						
1;02						
1;03						
1;04						
1;05						
1;08						
1;09	2	2				
1;10	1	2				
1;11						
2;00						
2;01						
2;02						
2;03	3	4	2	3		
2;04	5	8	6	14	9	13
2;05	23	41	11	28	60	112
2;06	30	44	8	17	42	71
2;07	29	50	6	13	27	33
2;08	52	78	7	7	30	41
2;09	35	53	7	10	26	44
2;10	16	28	2	2	11	14
2;11	15	20	4	6	14	23
3;00	12	20	7	9	15	22
3;01	17	25	5	8	20	27
3;02	21	32	7	7	21	27
3;03	14	21	7	12	15	21

Tabela III.16. Distribuição das palavras resultantes de processos morfológicos de formação de palavras (derivação, composição e sufixação (z-)avaliativa) (valores absolutos) - Criança (*type/token*).

	Derivação		Composição		Sufixação (z-)avaliativa	
	<i>type</i>	<i>token</i>	<i>type</i>	<i>token</i>	<i>type</i>	<i>token</i>
1;01						
1;02						
1;03						
1;04						
1;05						
1;08						
1;09	22,22%	16,66%				
1;10	11,11%	13,33%				
1;11						
2;00						
2;01						
2;02						
2;03	1,91%	1,28%	1,27%	0,96%		
2;04	1,42%	0,84%	1,71%	1,48%	2,56%	1,37%
2;05	2,72%	1,27%	1,30%	0,86%	7,10%	3,44%
2;06	4,68%	2,05%	1,25%	0,79%	6,54%	3,29%
2;07	5,62%	3,03%	1,16%	0,79%	5,23%	2,00%
2;08	8,34%	3,35%	1,12%	0,30%	4,82%	1,76%
2;09	5,43%	2,11%	1,09%	0,40%	4,03%	1,75%
2;10	4,16%	2,66%	0,52%	0,19%	2,86%	1,33%
2;11	3,33%	1,40%	0,89%	0,42%	3,10%	1,60%
3;00	2,32%	1,15%	1,35%	0,52%	2,90%	1,27%
3;01	4,36%	2,35%	1,28%	0,75%	5,13%	2,55%
3;02	4,87%	2,48%	1,62%	0,54%	4,87%	2,09%
3;03	3,51%	1,71%	1,76%	0,98%	3,77%	1,71%

Tabela III.17. Distribuição das palavras resultantes de processos morfológicos de formação de palavras (derivação, composição e sufixação (z-)avaliativa) (valores percentuais) - Criança (*type/token*).

A evolução destes três tipos de palavras que compõem as palavras complexas, são muito similares nos dados *type* e *token*. O mesmo não se passa em relação à frequência. Estes três tipos de palavras, palavras derivadas, compostas e com sufixação (z-)avaliativa são menos frequentes nos *tokens* do que nos *types*.

Nos últimos quatro meses observados, as palavras complexas (palavras derivadas, compostas e com sufixação (z-) avaliativa) atingem uma média de 9,4% (*type*) e de 4.5% (*token*) do discurso total da criança.

De seguida apresentamos um conjunto de palavras cujas produções efectivas da criança mostram diferenças relativamente às palavras alvo que ultrapassam os meros aspectos fonológicos na produção de palavras complexas (a realização efectiva da criança, apresentada na coluna da direita, é transcrita ortograficamente).

Coluna1	Coluna2	Coluna3
Idade	Alvo	Realização
2;04	chuchinha	chuchazinha
2;05	sozinha	sozazinha
	bolinha	bolazinha
	baixinho	baixoginho
	sozinho	sozuzinho
	malinha	malazinha
	jeitinho	jeitozinho
	cuequinha	quecazinha
	pomadinha	pomadazinha
	gatinho	gatozinho
	pouquinha	poucazinha
	quentinha	quentizinha
	vaquinha	vacazinha
	fuminhos	fumozinho
	fofinho	fofozinho
	chuchinha	chuchazinha
	pintinhas	pintazinha
	molinho	monzinho
	tontinha	tontazinha
	ervinhas	vevazinha
	sininho	sinozinho
	limpinha	mimpazinha
	sequinho	secozinho
	pianinho	pinhãozinho
	bolsinho	boçozinho
	bolinhas	bolazinha
	pombinho	bombozinho
	porquinhos	poucozinho
	pontinha	pontazinha
	sentadinho	sentazinha
2;06	bonequinha	bonecazinha

Tabela III.18. Diferenças entre palavras *alvo* e realização - Criança (*type/token*).

Na generalidade dos casos verifica-se, para além de eventuais outras particularidades, a utilização do sufixo z-avaliativo em formas que no adulto são tipicamente formadas com sufixos avaliativos. Tal acontece a partir dos 2;04 de idade. Estes dados irão ser importantes para a discussão da inferência de regras relativas à formação de palavras complexas.

I) 3.3 Classes de Palavras

Nesta subsecção é investigada a evolução da ocorrência das diversas classes de palavras nas produções da criança. Tal como dito no início deste capítulo, a importância desta secção prende-se com o facto de apenas um subconjunto de todas as classes de palavra admitir processos de formação de palavras. Por esta razão, é importante determinar qual a relação entre a emergência e evolução de palavras complexas na fala da criança e a emergência e desenvolvimento das classes de palavras a que as palavras complexas pertencem.

Tal como referido na secção 2.5 do capítulo *Metodologia*, foram consideradas as seguintes classes gramaticais de palavras: N *Nomes*; Adj *Adjectivos*; Adv *Advérbios*; V *Verbos*; Amlg-Outr (palavras que são resultado da lexicalização de duas (ou mais) palavras, originando uma amálgama, e em que uma das partes não pertence a nenhuma das classes anteriores (sendo aqui classificadas como do tipo *Outr*); Amlg-Adv (palavras que são resultado da lexicalização de duas (ou mais) palavras, originando uma amálgama, e em que uma das partes é uma palavra do tipo *Adv*); e foram colocadas na classe *Outr* todas as palavras de classe fechada, com função gramatical e sem conteúdo lexical, incluindo as interjeições e algumas palavras tradicionalmente classificadas como advérbios, incluindo todas as expressões “QU”, independente da sua função adverbial, interrogativa, relativa ou outra, bem como preposições, conjunções, artigos, determinantes, numerais, etc.

Observemos de seguida os quadros com os dados correspondentes às classes de palavras utilizadas pela criança, em valores absolutos e percentuais, computados sobre *types*.

Idade	N	V	Outr	Adv	Adj	Amlg-Adv	Amlg-Outr
1;01	1	1					
1;02	11	1	1				
1;03	6						
1;04	7	1					
1;05	8	1					
1;08	2						
1;09	4	2	1	2			
1;10	9						
1;11	1						
2;00	3		1	1			
2;01	35	1	4		3		
2;02	23	6	6	2	6		
2;03	86	27	16	9	18		1
2;04	182	81	43	16	23	2	4
2;05	444	224	72	25	68	2	9
2;06	293	187	63	24	62	1	12
2;07	218	150	68	26	40	3	11
2;08	253	206	68	23	58	2	15
2;09	277	210	82	25	37	2	12
2;10	142	135	59	18	25		5
2;11	178	146	67	24	25	1	10
3;00	210	193	68	22	18		6
3;01	141	115	63	26	37		8
3;02	152	139	68	25	35	1	11
3;03	150	132	56	22	21	2	15

Tabela III.19. Classes de Palavras (valores absolutos) – Criança (type).

Idade	N	V	Outr	Adv	Adj	Amlg- Adv	Amlg- Outr
1;01	50,00%	50,00%					
1;02	84,62%	7,69%	7,69%				
1;03	100%						
1;04	87,50%	12,50%					
1;05	88,89%	11,11%					
1;08	100%						
1;09	44,44%	22,22%	11,11%	22,22%			
1;10	100%						
1;11	100%						
2;00	60%		20%	20%			
2;01	81,40%	2,33%	9,30%		6,98%		
2;02	53,49%	13,95%	13,95%	4,65%	13,95%		
2;03	54,78%	17,20%	10,19%	5,73%	11,46%		0,64%
2;04	51,85%	23,08%	12,25%	4,56%	6,55%	0,57%	1,14%
2;05	52,54%	26,51%	8,52%	2,96%	8,05%	0,24%	1,07%
2;06	45,64%	29,13%	9,81%	3,74%	9,66%	0,16%	1,87%
2;07	42,25%	29,07%	13,18%	5,04%	7,75%	0,58%	2,13%
2;08	40,61%	33,07%	10,91%	3,69%	9,31%	0,32%	2,41%
2;09	42,95%	32,56%	12,71%	3,88%	5,74%	0,31%	1,86%
2;10	36,98%	35,16%	15,36%	4,69%	6,51%		1,30%
2;11	39,47%	32,37%	14,86%	5,32%	5,54%	0,22%	2,22%
3;00	40,62%	37,33%	13,15%	4,26%	3,48%		1,16%
3;01	36,15%	29,49%	16,15%	6,67%	9,49%		2,05%
3;02	35,27%	32,25%	15,78%	5,80%	8,12%	0,23%	2,55%
3;03	37,69%	33,17%	14,07%	5,78%	5,28%	0,50%	3,77%

Tabela III.20. Classes de Palavras (valores percentuais) – Criança (type).

Verificamos que a categoria de palavras mais utilizada pela criança, independentemente da estrutura interna das palavras, é a categoria *Nomes*. Seguidamente, a segunda categoria mais utilizada é a dos *Verbos*.

Até aos dois anos de idade, a criança utiliza maioritariamente *Nomes* e *Verbos*, sendo que, precocemente, introduz palavras de novas categorias *Outr* e *Advérbios* no seu léxico (e.g. *hii* aos 1;02; *já, lá* aos 1;09 anos).

A classe dos *Adjectivos* é aquela que surge mais tardiamente, aos 2;01 anos (e.g. *escuro, verde*).

Quanto à classe dos Advérbios, é utilizada pela primeira vez aos 2;00, e de forma constante a partir dos 2;02, onde observamos que se trata da classe (com exceção das amálgamas) menos frequente na maioria das idades (exceção feita para os 3;00 e 3;03), revelando valores sempre inferiores a 7%.

A partir desta idade (2;01) a criança utiliza todas as classes gramaticais no seu discurso, aumentando a frequência destas ao longo do tempo e conseqüentemente diminuindo a frequência de *Nomes*.

As palavras constituídas por amálgamas surgem por volta dos dois anos e meio e são muito pouco frequentes na fala da criança (e.g. *dali, daqui* (2;04); *depressa, nela* (2;05); *neste, noutra* (3;03)).

Apresentamos de seguida, três gráficos que sintetizam os valores percentuais referentes às classes gramaticais de palavras, em cada ano de produções da criança (dados *type*).

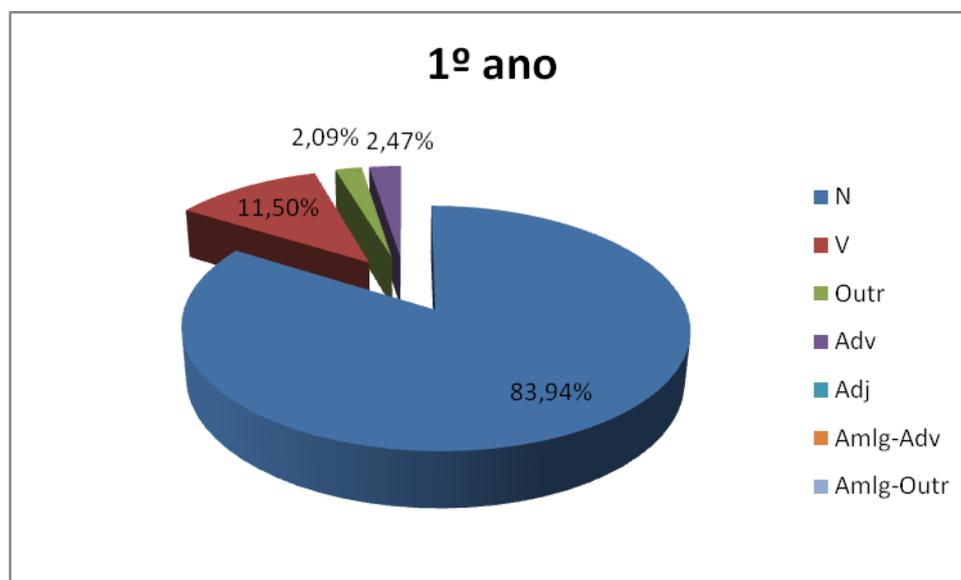


Figura III.4. Classes de Palavras (síntese por ano) – 1º ano (valores percentuais) – Criança (type).

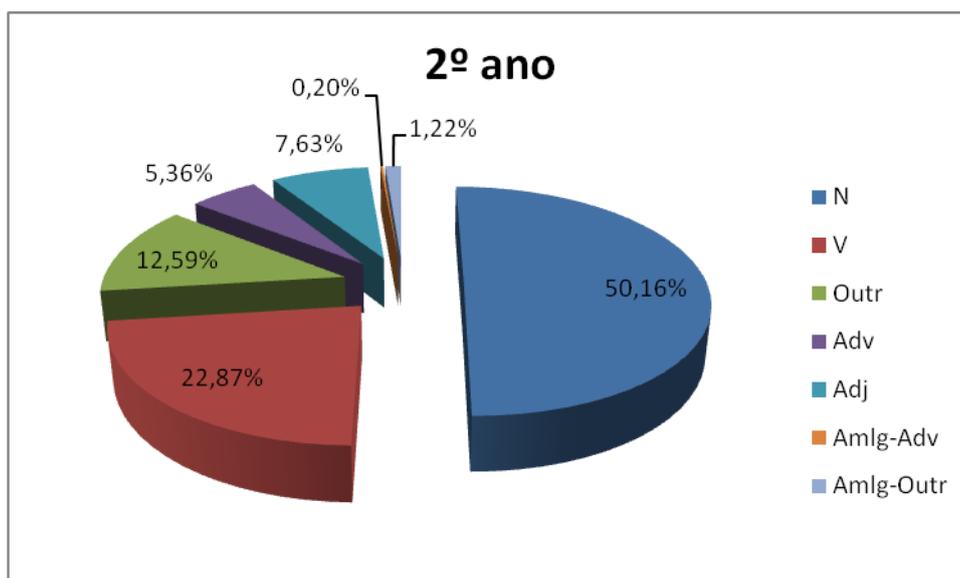


Figura III.5. Classes de Palavras (síntese por ano) – 2º ano (valores percentuais) – Criança (type).

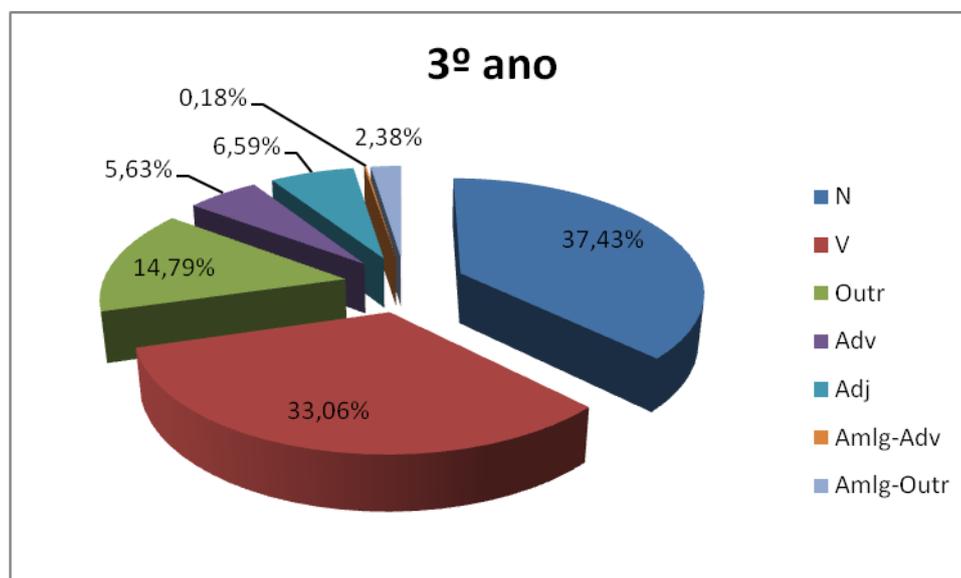


Figura III.6. Classes de Palavras (síntese por ano) – 3º ano (valores percentuais) – Criança (type).

Verificamos que, no primeiro ano existem apenas quatro classes gramaticais de palavras nas produções da criança (*Nomes, Verbos, Outros, Advérbios*) e que existem uma clara predominância de *Nomes*.

No segundo ano, todas as classes gramaticais estão presentes e observamos que os *Nomes* correspondem a 50% do discurso da criança, sendo que os restantes 50% são divididos por todas as classes, embora os *Verbos* correspondam a 33% do discurso da criança.

No terceiro ano, existe uma divisão de classes gramaticais usadas pela criança, com um predomínio significativa de *Nomes* e *Verbos*.

De seguida, disponibilizamos os quadros correspondentes às classes de palavras utilizadas pela criança, em valores absolutos e percentuais, agora computados sobre *token*.

Idade	N	V	Outr	Adv	Adj	Amlg- Adv	Amlg- Outr
1:01	2	1					
1:02	15	4	1				
1:03	9						
1:04	8	3					
1:05	10	2					
1:08	5						
1:09	5	3	1	3			
1:10	15						
1:11	1						
2:00	3		1	1			
2:01	41	1	6		4		
2:02	33	8	9	2	7		
2:03	161	44	46	29	29		4
2:04	458	215	113	103	46	2	10
2:05	1477	894	378	286	175	4	38
2:06	905	624	342	147	110	1	27
2:07	566	468	348	161	56	7	44
2:08	622	634	683	211	106	5	69
2:09	608	663	918	203	69	2	49
2:10	258	284	363	92	36		22
2:11	318	416	473	169	31	1	27
3:00	367	520	577	202	33		29
3:01	244	296	376	76	49		19
3:02	278	374	470	106	48	2	15
3:03	289	360	405	110	23	5	35

Tabela III.21. Classes de Palavras (valores absolutos) – Criança (token).

Idade	N	V	Outr	Adv	Adj	Amlg-Adv	Amlg-Outr
1;01	66,67%	33,33%					
1;02	75,00%	20,00%	5,00%				
1;03	100%						
1;04	72,73%	27,27%					
1;05	83,33%	16,67%					
1;08	100%						
1;09	41,67%	25,00%	8,33%	25,00%			
1;10	100%						
1;11	100%						
2;00	60%		20%	20%			
2;01	78,85%	1,92%	11,54%		7,69%		
2;02	55,93%	13,56%	15,25%	3,39%	11,86%		
2;03	51,44%	14,06%	15,97%	9,27%	9,27%		
2;04	48,36%	22,70%	12,99%	10,88%	4,86%	0,21%	
2;05	45,42%	27,49%	12,61%	8,79%	5,38%	0,12%	0,18%
2;06	41,98%	28,94%	17,12%	6,82%	5,10%	0,05%	
2;07	34,30%	28,36%	23,88%	9,82%	3,39%	0,24%	
2;08	26,70%	27,21%	32,32%	9,23%	4,55%		
2;09	24,20%	26,39%	38,50%	8,16%	2,75%		
2;10	24,45%	26,92%	36,49%	8,72%	3,41%		
2;11	22,16%	28,99%	34,84%	11,85%	2,16%		
3;00	21,24%	30,09%	35,07%	11,69%	1,91%		
3;01	23,02%	27,92%	37,26%	7,17%	4,62%		
3;02	21,50%	28,92%	37,20%	8,35%	3,71%		0,31%
3;03	23,55%	29,34%	33,58%	9,29%	1,87%	0,08%	2,28%

Tabela III.22. Classes de Palavras (valores percentuais) – Criança (token).

Os resultados relativos aos *token* revelam que, até aos 2;07 anos, a categoria mais utilizada pela criança é a categoria dos *Nomes*, seguida da classes verbal, que apresenta também valores desde o início das produções infantis. Importa dizer, de todo o modo, que no primeiro ano existem meses em que apenas são produzidas palavras da classe nominal. Apenas aos 2;02, numa altura em que o léxico sofre grande expansão, as palavras da classe verbal surgem consistentemente e com alguma frequência. A terceira categoria mais frequente, mas que até aos 2 anos aparece esporadicamente, é aquela designámos de *Outr*, tal como verificámos anteriormente para os dados *type*. A classe adverbial aparece consistentemente apenas a partir dos 2;02.

A partir dos 2;07 e até aos 3;03 anos, os resultados são bastante diferentes dos descritos para os *types* – as palavras mais frequentes pertencem agora à categoria *Outr*, onde estão incluídas as palavras de classe fechada. Importa clarificar, de todo o modo, que isto não significa, naturalmente, que o discurso da criança é dominado por palavras desta categoria genérica, uma vez que, se somadas as restantes classes, a percentagem de palavras pertencentes a classes abertas é de 84% e 65% (média para dados *type*) e fechadas é de 14% e 35% (média para dados *token*) respectivamente, para este intervalo de idade.

Dentro das classes de palavras abertas, a criança utiliza neste período mais *Verbos* e apenas em terceiro lugar usa palavras da classe *Nome*.

Outra diferença observada relativamente a *types* e *tokens*, relaciona-se com a classe menos frequente. Enquanto que nos *types*, a classe menos frequente é a dos *Adverbios*, nos *tokens*, a classe menos frequente é a dos *Adjectivos* (com excepção das amálgamas).

Apresentamos de seguida, três gráficos que sintetizam os valores percentuais referentes às classes gramaticais de palavras, em cada ano de produções da criança, dados *token*.

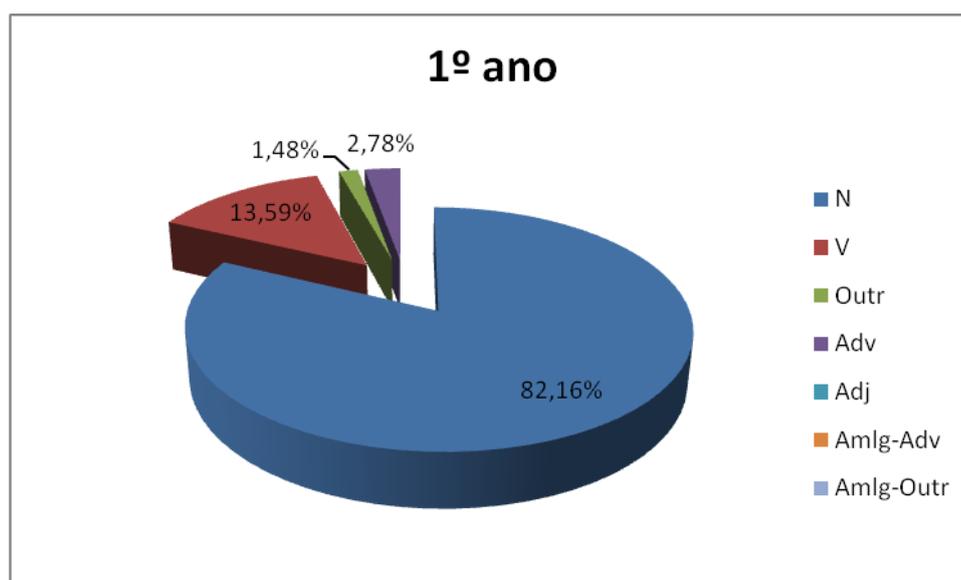


Figura III.7. Classes de Palavras (síntese por ano) – 1º ano (valores percentuais) – Criança (token).

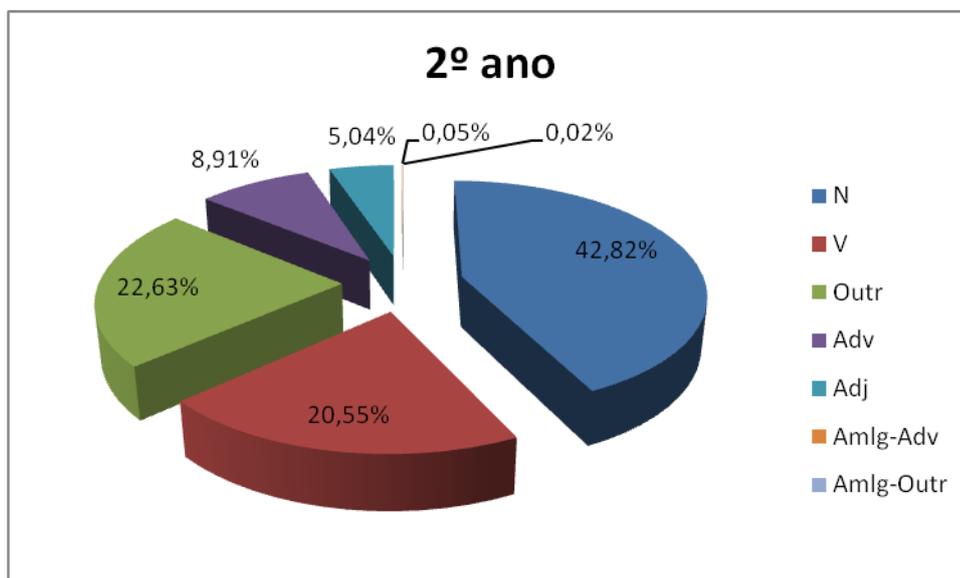


Figura III.8. Classes de Palavras (síntese por ano) – 2º ano (valores percentuais) – Criança (token).

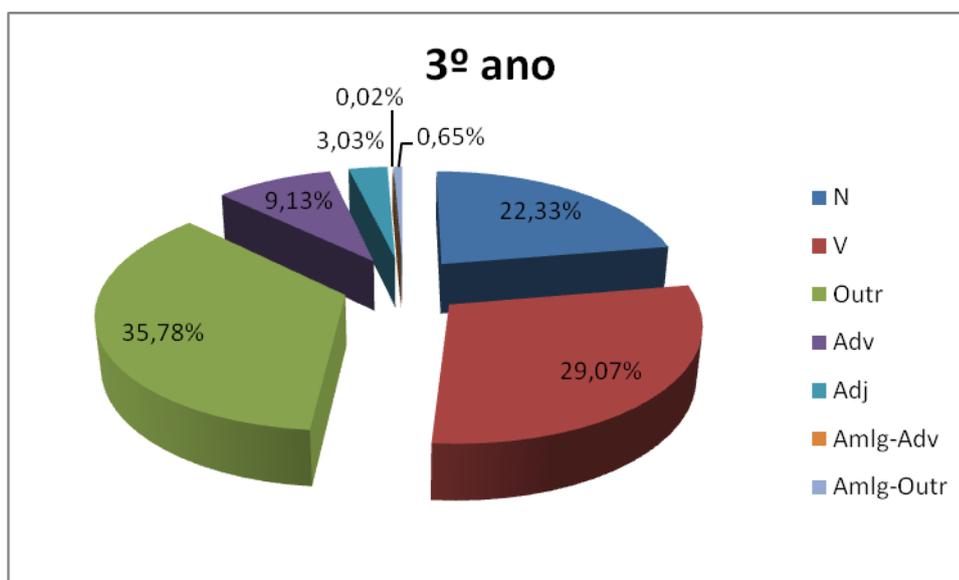


Figura III.9. Classes de Palavras (síntese por ano) – 3º ano (valores percentuais) – Criança (token).

Verificamos, mais uma vez, que no primeiro ano, os valores dos dados *type* e *token* são muito similares, mas no 2º e 3º ano existem diferenças significativas entre os dados *type* e *token*, mais visíveis nestes gráficos síntese.

I) 3.4 O tamanho das palavras simples e complexas

Analisamos de seguida o tamanho de palavras que constituem o discurso da criança, relativamente às formas efectivamente produzidas pela criança e às formas *alvo*, isto é, as formas que as palavras produzidas têm na fala adulta.

A importância do estudo do tamanho de palavra prende-se com o facto de as palavras complexas terem necessariamente *um certo* tamanho (superior a duas sílabas, como veremos na subsecção II.3.4, desta mesma secção, relativa ao adulto), pelo que é concebível que a criança produza formas complexas apenas quando produzir palavras maiores, isto é, palavras com maior número de sílabas.

Seguidamente, apresentamos os resultados provenientes da análise do número de sílabas das palavras produzidas e das palavras *alvo* da criança, divididas em dois pontos.

Nas duas subsecções seguintes, os resultados não estão divididos em dados *type* e *token*, como tem vindo a acontecer nas subsecções anteriores. Isto porque, não nos interessa fazer uma distinção entre estes dois tipos dados, mas sim comparar os dois tipos de palavras: *alvo* e realização. São sempre apresentados os dados que dizem respeito à ocorrência de todas as palavras, incluído repetições da mesma palavra (*token*), pois a mesma palavra realizada pode corresponder a *alvos* diferentes e o mesmo *alvo* pode ser realizado, de diferentes maneiras, ao longo das idades e numa mesma idade.

3.4.1 O tamanho das palavras na criança - produções *alvo*

Nesta subsecção analisamos o tamanho das palavras *alvo* do discurso da criança, isto é, as palavras que a criança pretende produzir.

Para cada idade da criança, verificámos quantas palavras com diferentes números de sílabas constituíam as palavras *alvo* da criança.

Os resultados são apresentados em valores percentuais e valores absolutos.

De seguida é apresentada uma tabela respeitante à distribuição dos diferentes formatos de palavra (*alvo*) da criança, em valores absolutos. Posteriormente são apresentados três gráficos que correspondem aos valores percentuais dos mesmos formatos de palavras, divididos em anos da criança, por uma questão de facilidade de visualização de dados.

Nr. Sílabas	1	2	3	4	5	6	Total
1;01	1	2					3
1;02	6	13	1				20
1;03	1	4	4				9
1;04	6	5	1				12
1;05		6	4	2			12
1;08		7	2				9
1;09	7	9	1	1			18
1;10	2	13					15
1;11		1					1
2;00	2	3					5
2;01	26	23	6	1			56
2;02	33	33	8	1			75
2;03	161	188	29	6	1		385
2;04	418	620	91	15	1		1145
2;05	1471	1996	427	64	2	1	3961
2;06	1045	1150	301	89	7	1	2593
2;07	701	869	199	51	6		1826
2;08	953	1055	242	86	11		2347
2;09	1079	1030	339	84	5		2537
2;10	450	437	129	38	4	2	1060
2;11	585	632	196	30	9		1452
3;00	790	692	204	58	9		1753
3;01	466	419	144	46	1	1	1077
3;02	567	520	159	50	9		1305
3;03	509	536	146	35	11		1237

Tabela III.23. Número de sílabas - alvo (valores absolutos) - Criança.

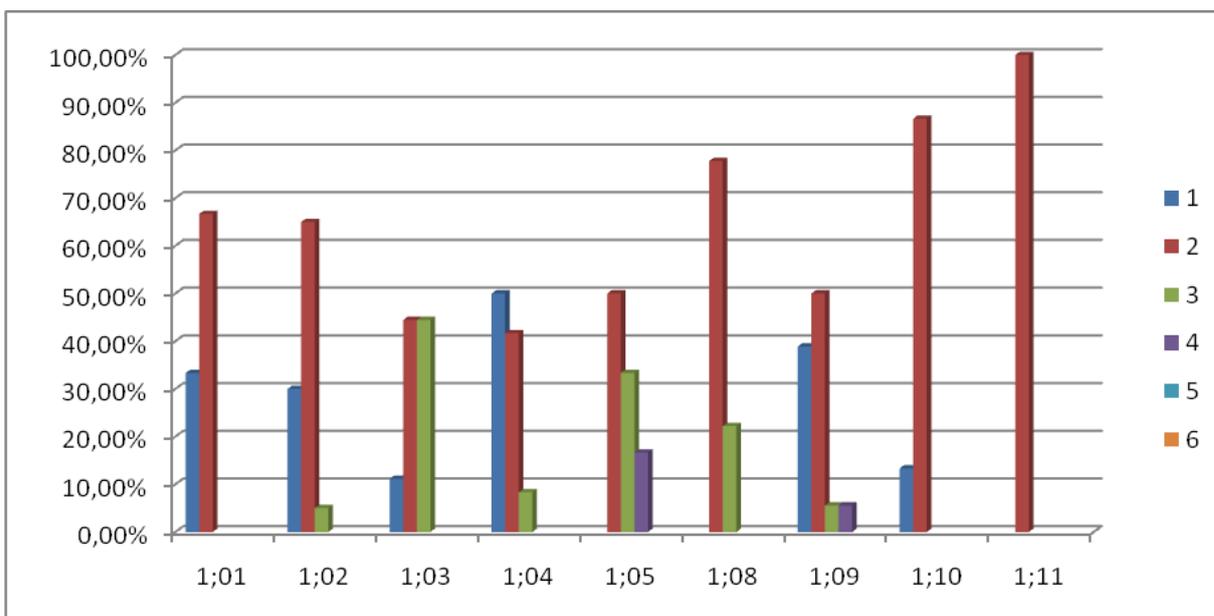


Figura III.10. Número de sílabas - alvo (valores percentuais) – 1º ano - Criança.

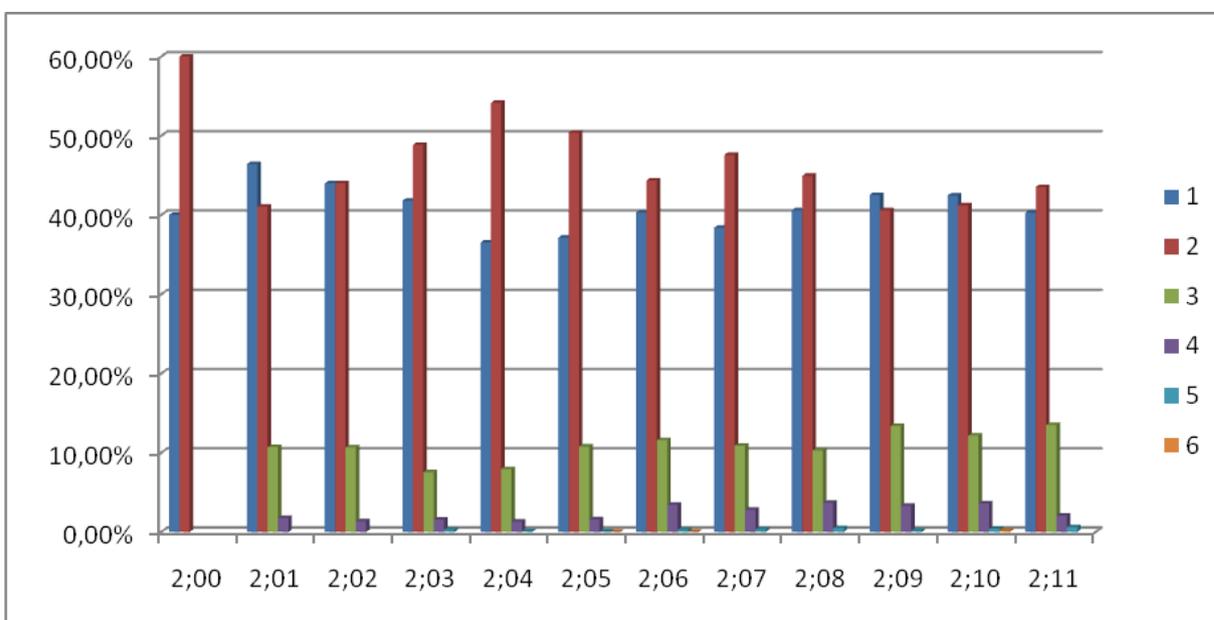


Figura III.11. Número de sílabas - alvo (valores percentuais) – 2º ano - Criança.

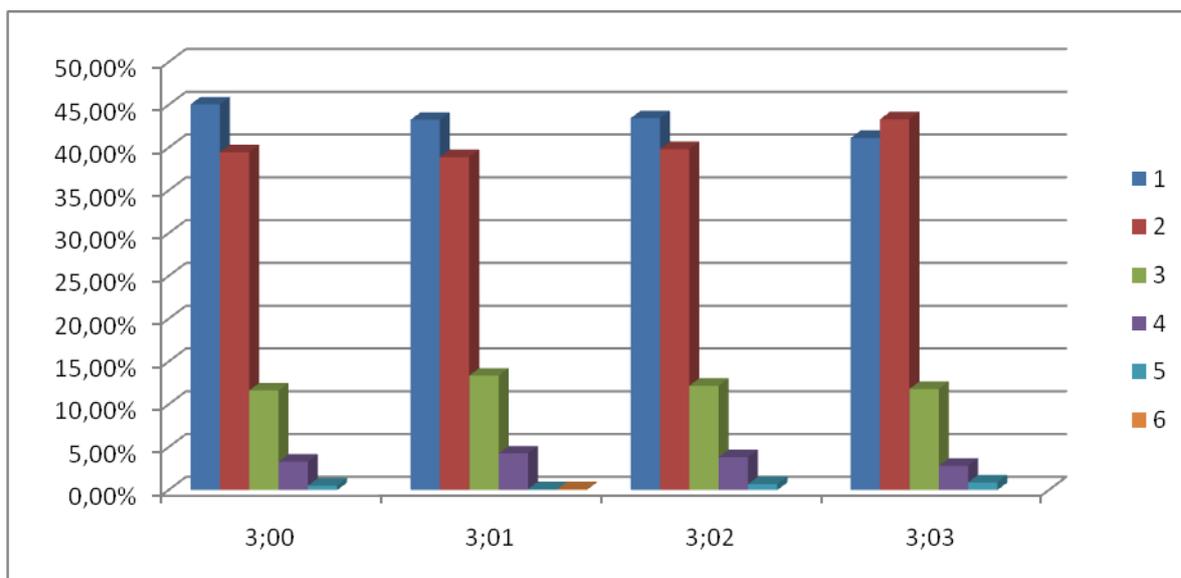


Figura III.12. Número de sílabas - alvo (valores percentuais) – 3º ano - Criança.

De um modo geral, as palavras com duas sílabas são as mais utilizadas pela criança, excepto em idades pontuais: 1;04, 2;01, 2;09, 2;10 e 3;00-3;02. Este formato de palavras apresenta valores elevados de frequência no discurso da criança, para todas as idades.

As palavras monossilábicas estão sempre presentes no discurso da criança, com excepção de três idades (1;05, 1;08 e 1;11). Estas palavras apresentam valores elevados de frequência entre os 30% e os 50%, ressalvando as idades 1;03 (11,11%) e 1;10 (13,33%), em que apresentam valores mais baixos.

Quanto às palavras trissilábicas, estas surgem precocemente na criança, aos 1;02 e encontra-se de forma constante no discurso da criança, excepto nas idades compreendidas entre os 1;10 e os 2;00, em que não existem palavras *alvo* com três sílabas, no discurso da criança. Esta é a terceira forma de palavras mais utilizada pela criança.

As palavras com quatro sílabas estão presentes pela primeira vez aos 1;05 e de seguida aos 1;09, contudo a presença deste tipo de palavras de forma constante no discurso da criança, surge aos 2;01. A partir desta idade, esta forma de palavras apresenta valores reduzidos, entre os 1,31% e o valor máximo de 4,27% aos 3;01.

Relativamente às palavras com cinco sílabas surgem aos 2;03 no discurso da criança, mantém-se de forma constante em todas as idades mas nunca atingem 1% do número total de palavras.

De um modo geral, as palavras *alvo* mais frequentes são as palavras dissilábicas, seguidas das palavras monossilábicas e das palavras trissilábicas.

A partir da idade 2;03 e até aos 3;03, as produções da criança incluem sempre palavras *alvo* até cinco sílabas.

3.4.2 O tamanho das palavras na criança – realização

Esta subsecção apresenta os resultados relativos ao número de sílabas das palavras efectivamente realizadas pela criança em função do tamanho. Como é sabido, as primeiras produções das crianças não apresentam, tipicamente, a forma adulta, podendo ser realizadas, nomeadamente, com truncamento de sílabas. Isto tem, naturalmente, consequências nos formatos das primeiras palavras. Tal como na subsecção anterior, importa-nos verificar se existe uma correlação entre a evolução dos diversos formatos de palavra das formas efectivamente produzidas pela criança e a emergência e desenvolvimento de palavras complexas.

Apresentamos, de seguida, uma tabela correspondente ao número de sílabas, referentes à realização efectiva de cada palavra, em valores absolutos. Posteriormente são apresentados três gráficos que correspondem aos valores percentuais dos mesmos formatos de palavras, divididos em anos da criança.

Nr. Sílabas	1	2	3	4	5	6	7	Total
1;01	1	2						3
1;02	7	11	2					20
1;03	5	3	1					9
1;04	7	5	1					13
1;05	5	8						13
1;08	1	8						9
1;09	6	6						12
1;10	6	10	1					17
1;11		1						1
2;00	2	3						5
2;01	30	26						56
2;02	22	37						59
2;03	126	182	6					314
2;04	219	650	38	5				912
2;05	758	2330	106	14				3208
2;06	686	1143	277	41				2147
2;07	594	793	206	29	1			1623
2;08	932	958	233	69	1			2193
2;09	1077	925	310	74	2			2388
2;10	443	406	120	38	3			1010
2;11	588	579	182	36	9			1394
3;00	702	648	200	69	15		5	1639
3;01	440	386	150	53	5			1034
3;02	518	490	170	57	10			1245
3;03	483	482	147	52	9			1173

Tabela III.24. Número de sílabas - *realização* (valores absolutos) - Criança.

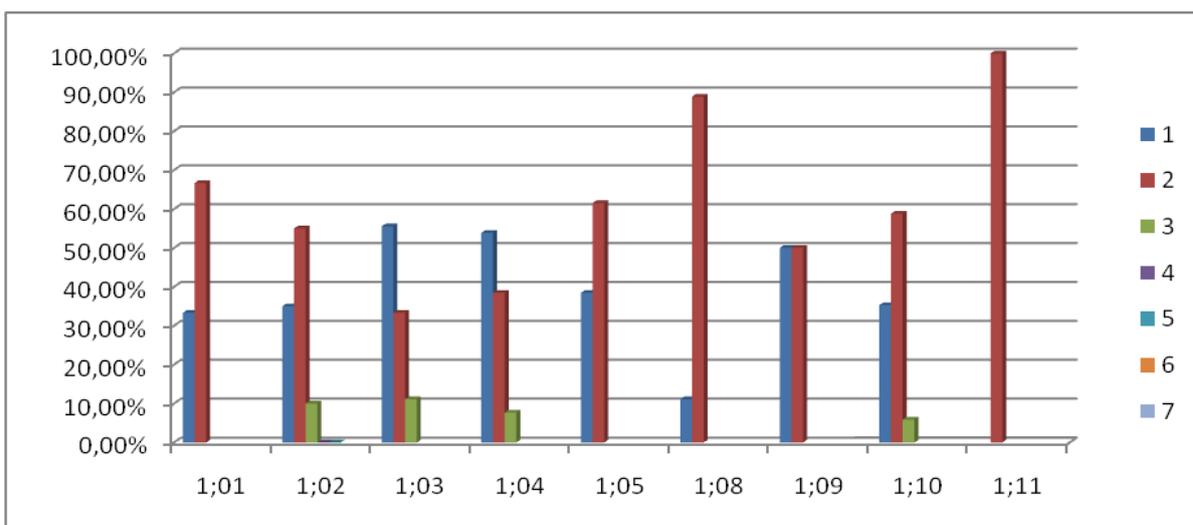


Figura III.13. Número de sílabas - realização (valores percentuais) – 1º ano - Criança.

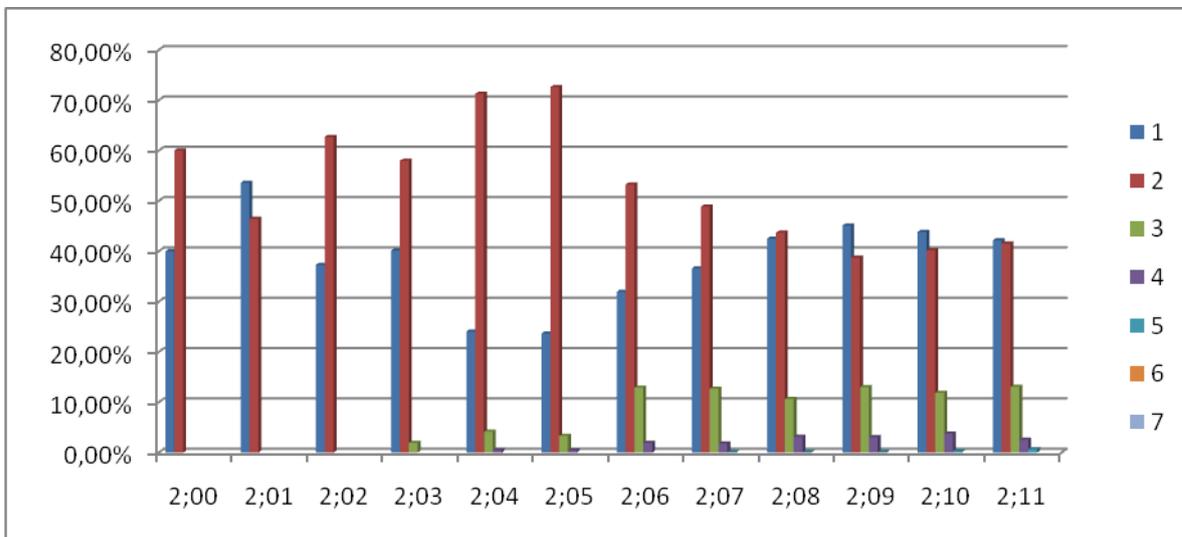


Figura III.14. Número de sílabas - realização (valores percentuais) – 2º ano - Criança.

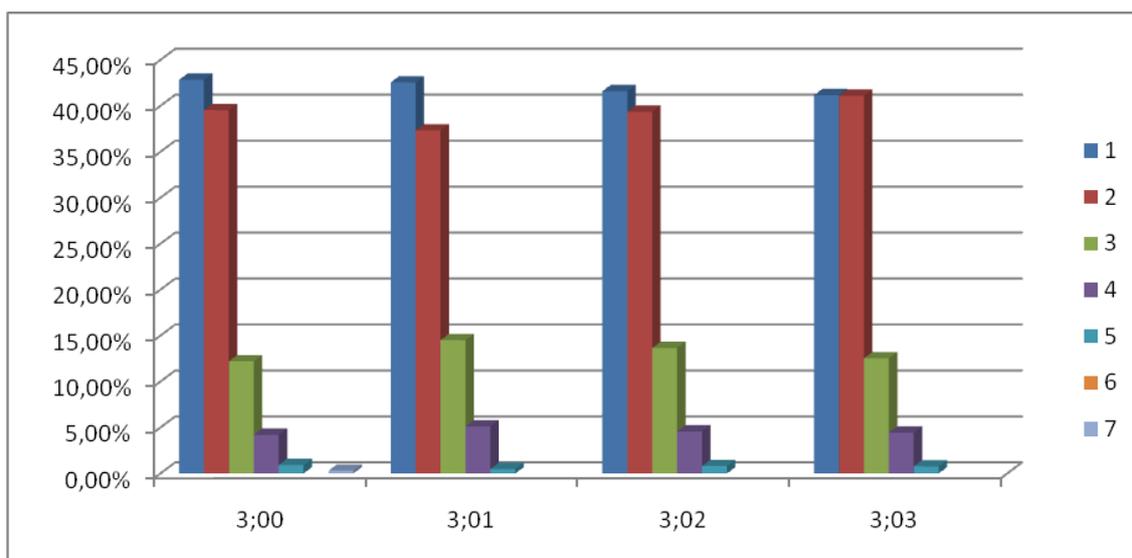


Figura III.15. Número de sílabas - realização (valores percentuais) – 3º ano - Criança.

O formato de palavras realizadas pela criança, quanto ao número de sílabas, encontra-se fundamentalmente dividido entre as palavras monossilábicas e dissilábicas.

Se bem que nas idades mais precoces as palavras dissilábicas são mais frequentes, nas idades mais tardias as palavras monossilábicas apresentam valores bastantes próximos das dissilábicas, as mais frequentes. Até à idade 2;08, as palavras dissilábicas sobrepõem-se às palavras monossilábicas com exceção das idades 1;03, 1;04 e 2;01. A partir dos 2;09 e até aos 3;03 as palavras com uma sílaba são mais frequentes, mas apresentam valores próximos das dissilábicas. Estas duas formas de palavras são altamente frequentes, correspondendo a cerca de 80 a 90% das palavras realizadas pelas crianças.

As palavras trissilábicas são realizadas pelas crianças em idades bastante precoces, 1;02, 1;03, 1;04 e 1;10. Porém, apenas são realizadas de modo constante em todas as idades a partir dos 2;03 anos. Estas palavras correspondem à terceira forma mais utilizada pela criança, apresentando o valor máximo de 14,51% aos 3;01 anos.

Quanto às palavras com quatro sílabas, verificamos que emergem em idades mais tardias comparativamente com as restantes formas de palavras, aos 2;04 anos. Apresentam valores compreendidos entre os 0,44% e o valor máximo de 5,13%. Este tipo de palavras está sempre presente no discurso da criança a partir desta idade.

As palavras com cinco sílabas são realizadas a partir dos 2;07 anos no discurso da criança, embora em número muito reduzido, nunca atingindo um ponto percentual.

Relativamente a palavras com mais do que cinco sílabas, é apenas observada uma palavras, repetida cinco vezes, com sete sílabas realizada aos 3;00 anos (e.g. *abetesimimagi (realização); apetece mais (alvo)*).

De um modo geral, as palavras realizadas pela criança são marcadas por palavras mono e dissilábicas que caracterizam o discurso, em terceiro lugar realizam palavras trissilábicas. As palavras com quatro e cinco sílabas ocupam um espaço reduzido no discurso da criança mas mantêm-se presentes de forma constante. É apenas observada uma palavra com mais do que cinco sílabas que corresponde a uma única palavra de sete sílabas.

3.5 Afixos utilizados

Nesta subsecção verificamos quais os cinco afixos não flexionais mais utilizados pela criança em cada mês.

Conhecer quais os afixos flexionais concretos mais produtivos no léxico infantil tem, quanto a nós e dado o foco desta tese, um interesse intrínseco. Para além disso, permite-nos, pela comparação com os dados do adulto, avaliar a este respeito o papel da informação de frequência no *input* nas produções infantis.

Por uma questão de simplificação na apresentação dos resultados e à semelhança do que já aconteceu na apresentação de resultados anteriores, dividimos os resultados em três gráficos, por ano da criança. Em cada gráfico, são apresentados os valores mensais correspondentes ao respectivo ano.

O número de afixos, corresponde aos valores absolutos, número de afixos utilizados pela criança.

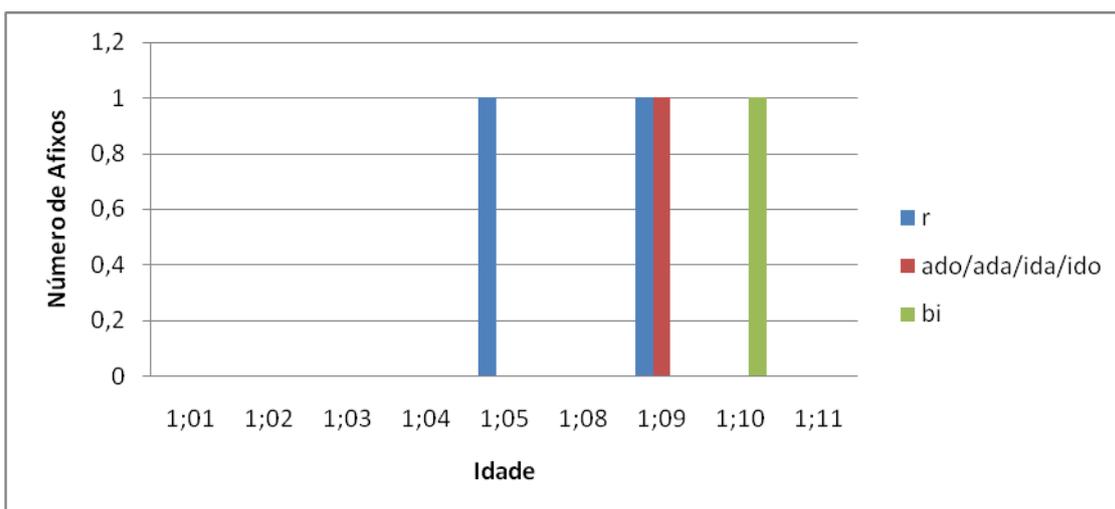


Figura III.16. Afixos utilizados durante o 1º ano (valores absolutos) - Criança (type).

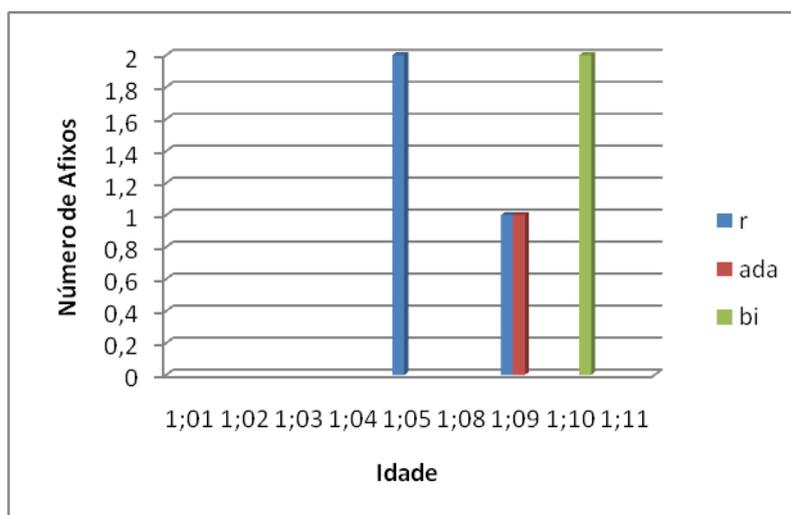


Figura III.17. Afixos utilizados durante o 1º ano - Criança (token).

Durante o primeiro ano, a criança apenas utiliza palavras com afixos em três idades (1;05, 1;09 e 1;10) e de forma limitada, isto é, a criança apenas utiliza um afixo aos 1;05A /r/; aos 1;09A utiliza novamente o afixo /r/ e também um afixo /ada/ e aos 1;10 anos utiliza um afixo /bi/.

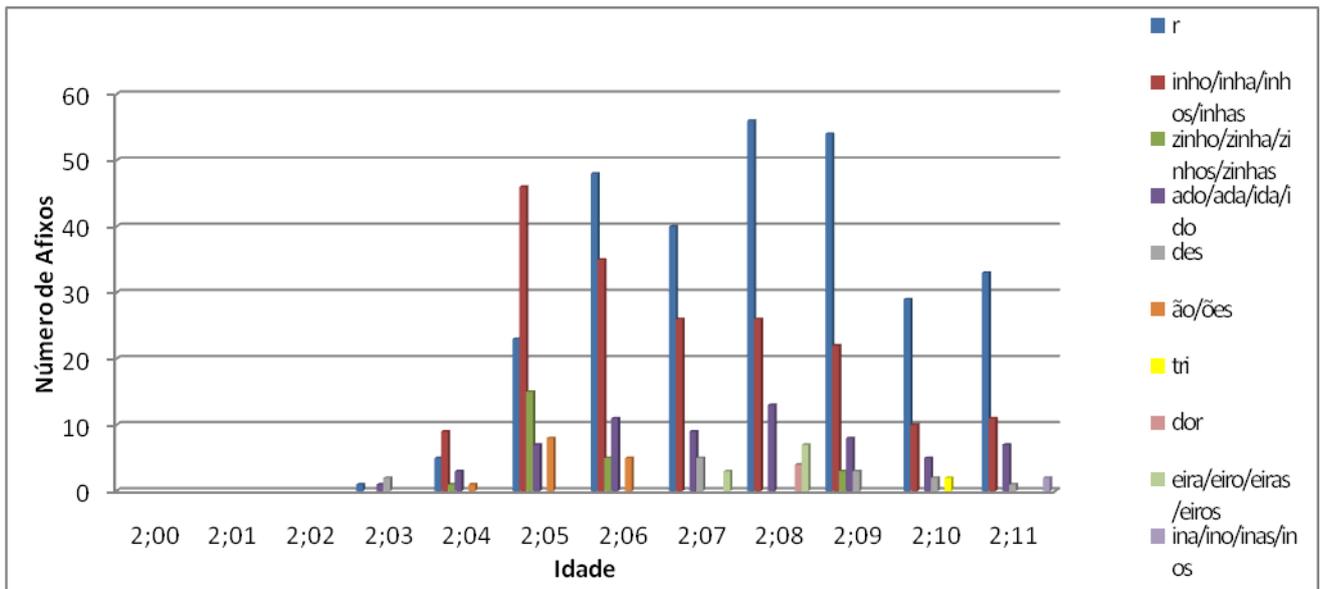


Figura III.18. Cinco afixos mais utilizados em cada mês, durante o 2º ano (valores absolutos) - Criança (type).

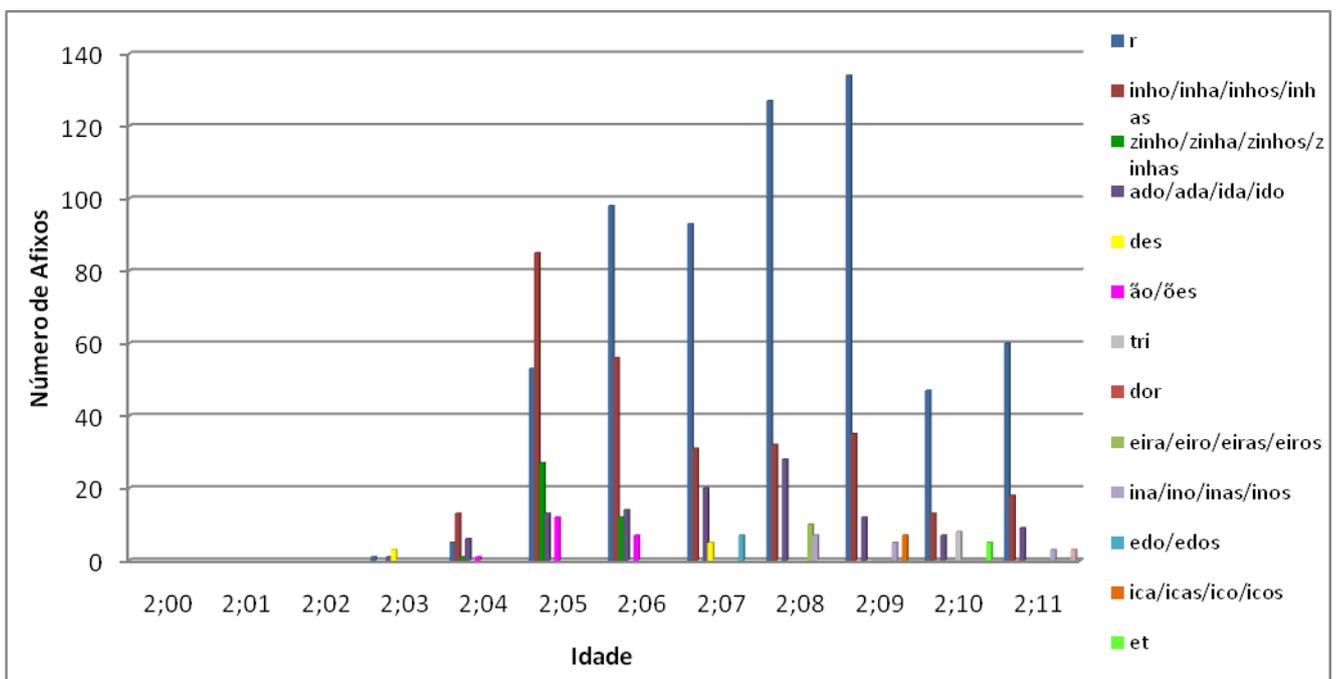


Figura III.19. Cinco afixos mais utilizados em cada mês, durante o 2º ano (valores absolutos) - Criança (token).

Durante o segundo ano de idade da criança, os afixos presentes nas palavras produzidas aumentam de forma considerável – tanto em número como em diversidade – a partir dos 2;03 anos. Aos 2;04 e 2;05 anos os afixos mais utilizados são os avaliativos (-inho, -inha, -inhos, -inhas). A partir dos 2;06 anos o afixo mais utilizado é -r. Estes valores são comuns a *types* e a *tokens*.

Relativamente à variedade de afixos utilizados, aos 2;06 anos a criança utiliza palavras com 19 afixos diferentes, aos 2;08 utiliza palavras com 28 afixos diferentes e aos 2;09 utiliza palavras com 25 afixos distintos. São estas as idades com palavras com maior variedade de afixos nos *types*.

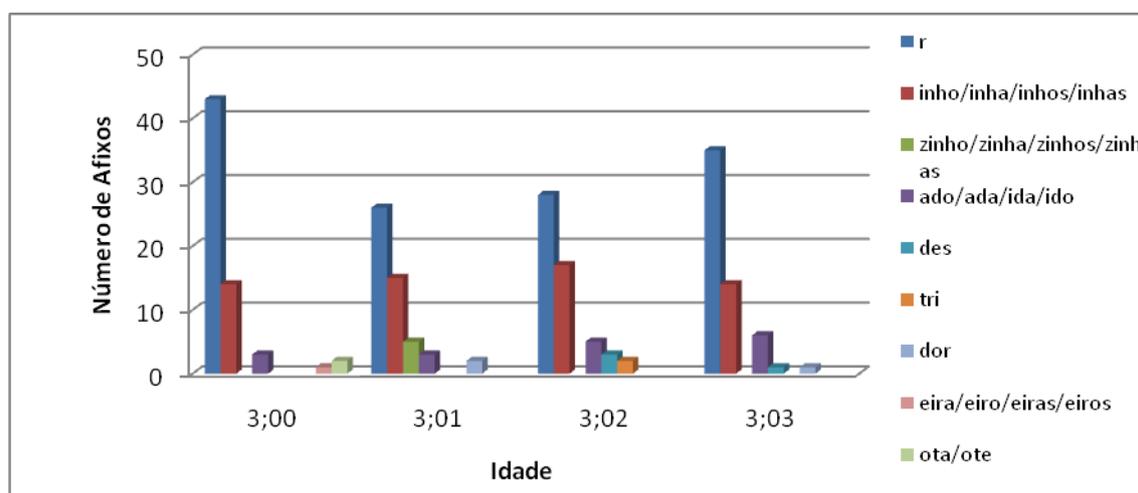


Figura III.20. Cinco afixos mais utilizados em cada mês, durante o 3º ano (valores absolutos) - Criança (type).

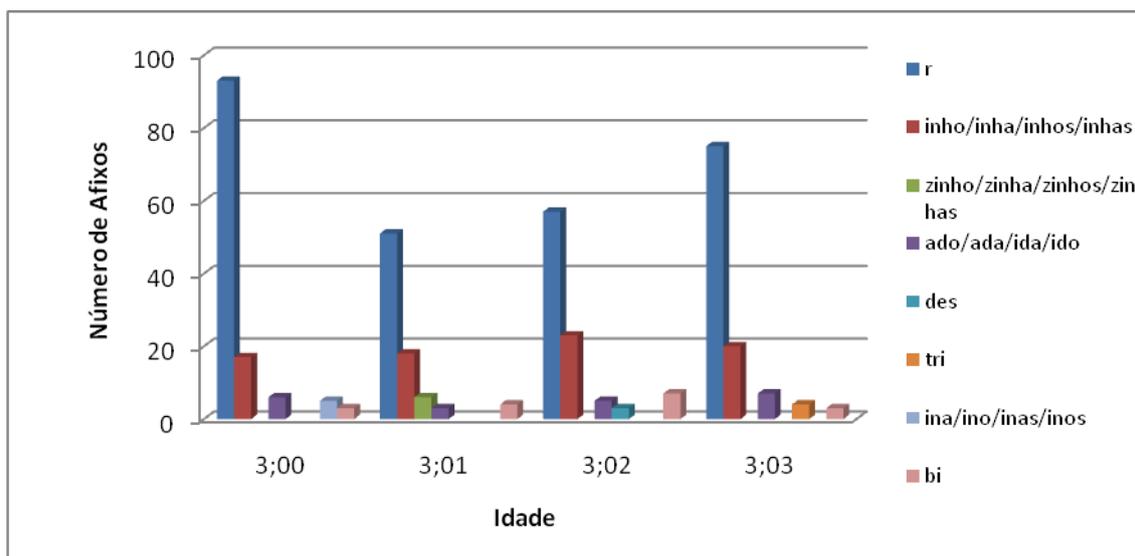


Figura III.21. Cinco afixos mais utilizados em cada mês, durante o 3º ano (valores absolutos) - Criança (token).

Os afixos utilizados no terceiro ano de vida da criança, seguem o mesmo padrão que os relativos ao período anterior. O afixo /r/ é o mais utilizado em todas as idades.

Apesar de aqui só estarem representados os cinco afixos mais frequentes, a criança utiliza uma grande variedade de afixos. A lista completa de afixos utilizados pela criança encontra-se em anexo (Anexo 3 e 4).

III.II. Análise de Resultados no Adulto

Dada a importância dos dados do *input* para a emergência e desenvolvimento da linguagem da criança, analisamos nesta secção os dados relativos ao adulto, tendo por base na generalidade dos casos a Base de Dados *Corpus* do Português Falado, tal como foi referido na secção 2.4 do capítulo *Metodologia*. Como também aí foi dito, apenas no caso da determinação da distribuição das diferentes classes de palavras não foi esta a fonte dos dados usada (subsecção II) 3.3 Classes de Palavras.

II) 3.1 Palavras simples *versus* palavras complexas

Nesta subsecção são analisadas as palavras simples e complexas utilizadas pelo adulto. O nosso objectivo é determinar a relação existente entre os valores encontrados na fala adulta e os verificados na fala da criança.

Apresentamos baixo os valores relativos à frequência de palavras em função da sua complexidade morfológica, em valores absolutos e valores percentuais, na fala adulta.

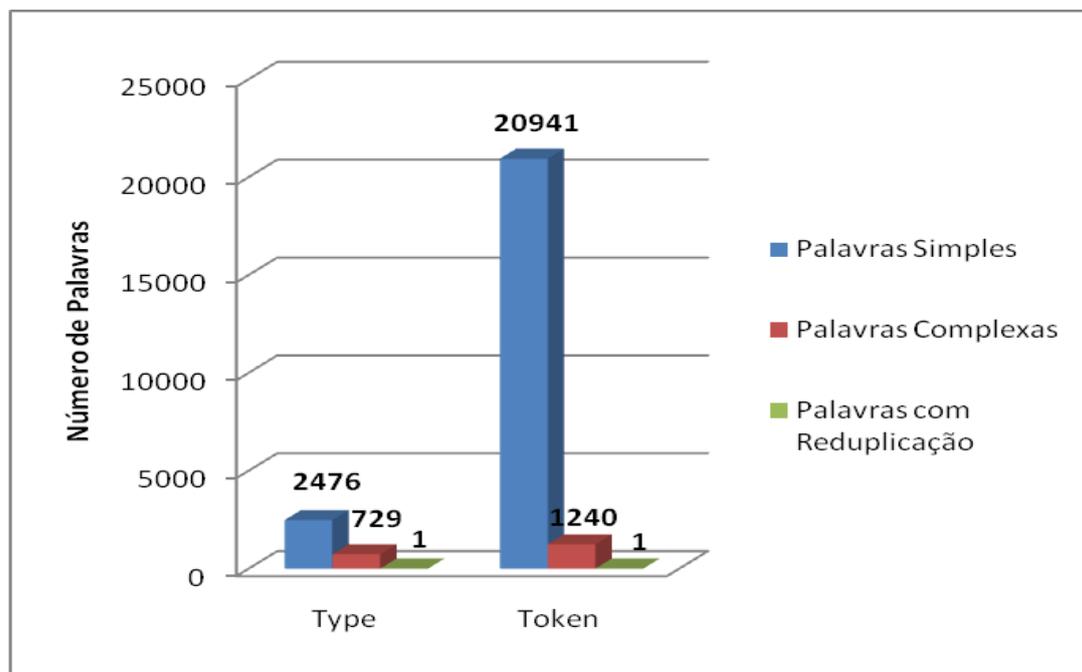


Figura III.22. Palavras simples, complexas e com reduplicação no adulto – valores absolutos (type e token).

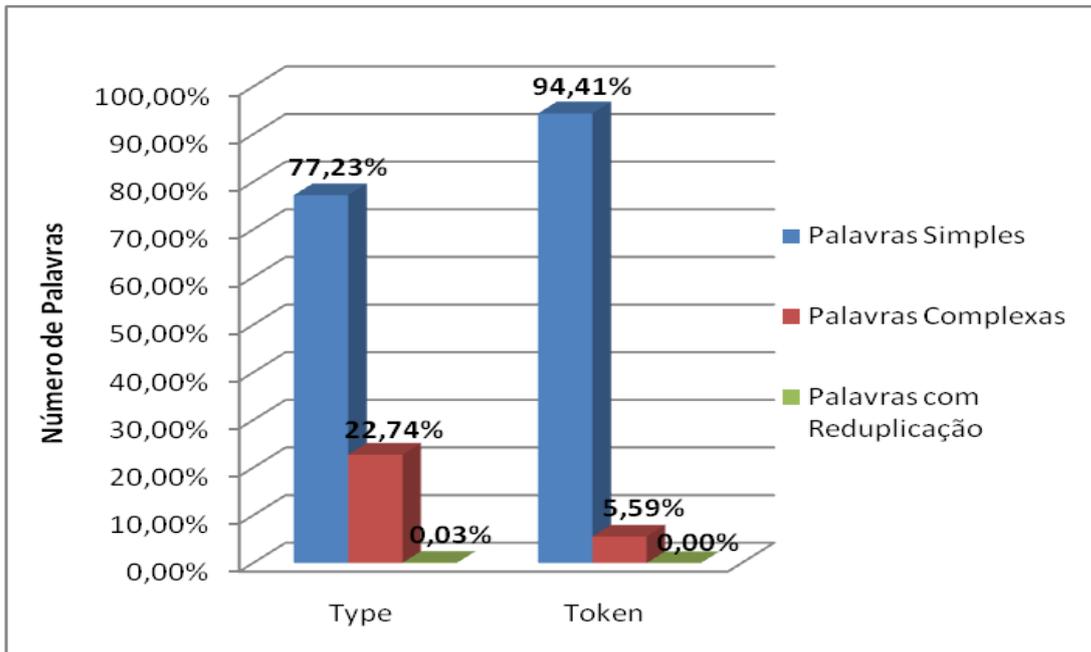


Figura III.23. Palavras simples, complexas e com reduplicação no adulto – valores percentuais (type e token).

Os valores apresentados para o adulto, vão no mesmo sentido que os resultados obtidos na criança: um número diminuto de palavras complexas relativamente às palavras simples.

Analisando o discurso do adulto, verificámos que entre as palavras complexas mais frequentes para o adulto encontramos (*altura, normalmente, qualquer, televisão*), e entre as palavras simples (*que, e, não, a*).

O discurso do adulto contém apenas uma palavra com reduplicação (*bebê*), sendo por isso claro que as palavras reduplicadas não caracterizam o discurso do adulto.

Observam-se outras diferenças relativamente à criança na proporção entre palavras simples e complexas, em especial quando considerados os *types*. Nestes dados as palavras complexas atingem os 22,74%.

Os valores de frequência nos *tokens* aproximam-se mais dos resultados encontrados para a criança. As palavras complexas correspondem a cerca de 6% do total

do discurso do adulto. As palavras simples dominam o discurso do adulto em cerca de 94%.

II) 3.2 Tipos de palavras em função da sua estrutura

Nesta subsecção são apresentados os resultados respeitantes à constituição das palavras na fala adulta. Estes resultados irão ser relacionados com os da criança, de modo a avaliarmos a sua relação com os dados infantis.

São de seguida observadas as distribuições percentuais do tipo de palavras utilizadas em função da sua constituição morfológica, quer nas contagens sobre *types* quer nas sobre *tokens*, no adulto.

Apresentamos de seguida os valores correspondentes aos diferentes tipos de palavras consoante a sua estrutura nos *types* do adulto, quer em valores percentuais como absolutos.

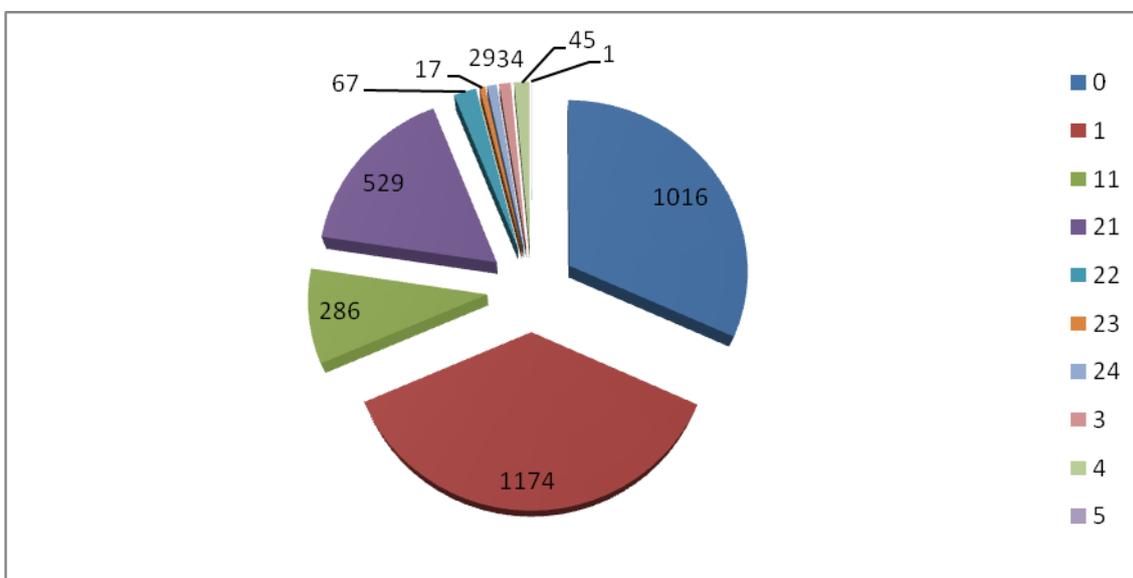


Figura III.24. Tipo de palavras em função da estrutura interna (valores absolutos) - Adulto (type).

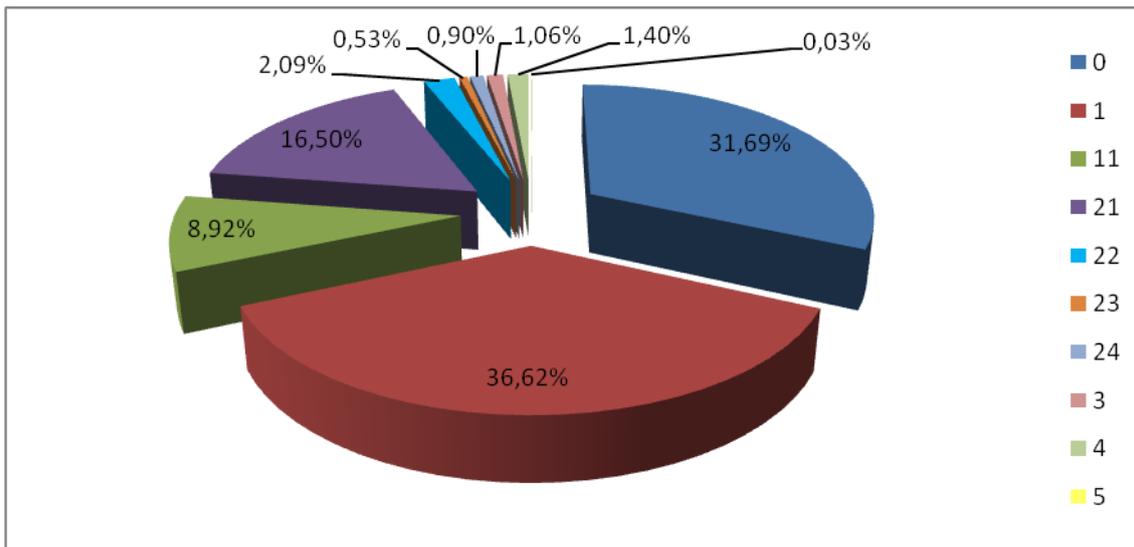


Figura III.25. Tipo de palavras em função da estrutura interna (valores percentuais) - Adulto (type).

De acordo com os resultados representados nestes dois gráficos, observamos que existe uma predominância de formas simples, e, dentro destas, são mais frequentes as palavras com sufixo flexional (36,62%) do que as não flexionadas (31,69%).

As formas complexas em conjunto com as formas nominais dos verbos atingem cerca de um terço do total de palavras (31,69%).

Observamos ainda que, excluindo as palavras simples, as palavras com derivação sufixal (16,50%) e as formas verbais infinitivas e participiais (8,92%) são as que mais se destacam no discurso do adulto.

Dentro das palavras complexas, verificamos que existe uma predominância das palavras derivadas relativamente às palavras compostas e às com sufixação (z-) avaliativa. Os restantes tipos de palavras, com a exceção das palavras derivadas por prefixação, que ainda constituem 2% do total de palavras, são muito pouco frequentes.

As palavras com reduplicação, tal como foi referido anteriormente, apresentam uma ocorrência única no *corpus* do adulto em observação.

Seguidamente, apresentamos os resultados dos tipos de palavras em função da estrutura interna, mas relativamente aos *tokens*, valores percentuais e absolutos.

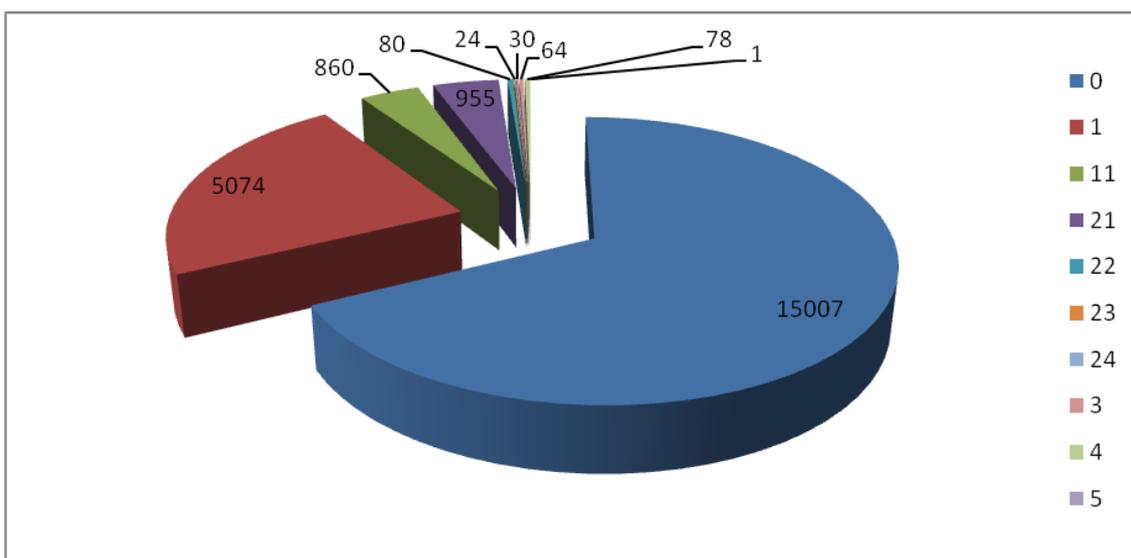


Figura III.26. Tipo de palavras em função da estrutura interna (valores absolutos) - Adulto (token).

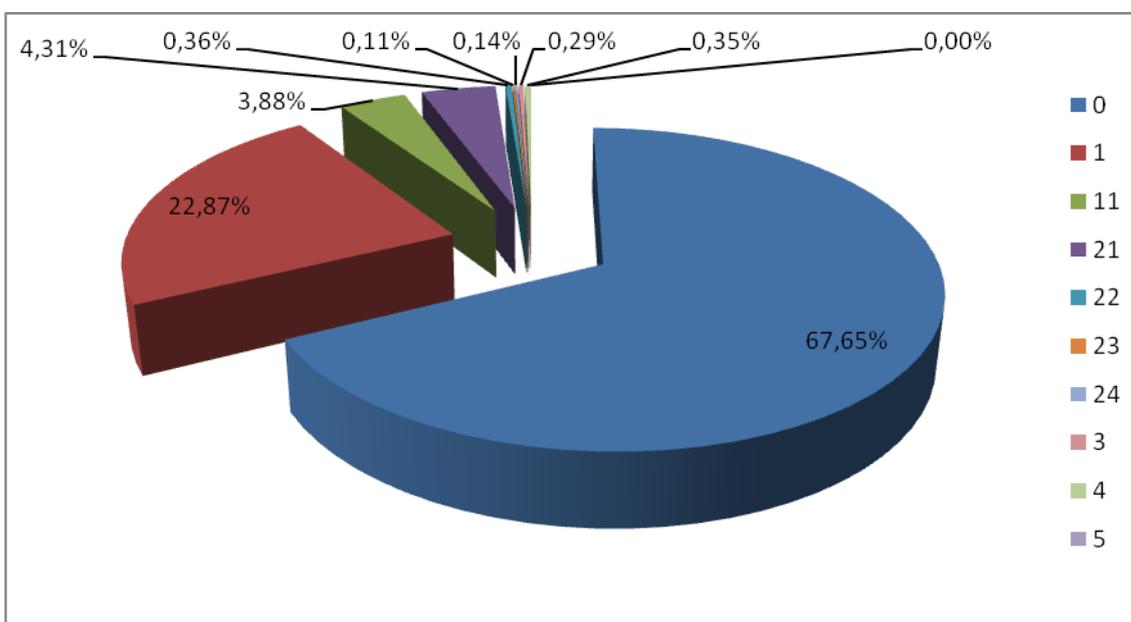


Figura III.27. Tipo de palavras em função da estrutura interna (valores percentuais) - Adulto (token).

Analisando os gráficos acima, observamos que nos dados *token* a preponderância das formas simples é muito significativa, tal como a das formas simples não flexionadas (67,65%) em relação às formas flexionadas (22,87%).

As palavras complexas (derivadas, compostas e com sufixação (-z)avaliativa) atingem, nestes dados *token*, um valor diminuto, 5,56%, destacando-se a predominância da derivação relativamente às restantes formas.

As formas de palavras com reduplicação mantêm-se virtualmente inexistentes.

De todos os tipos de palavras analisadas e, retirando as duas formas mais utilizadas pelo adulto, formas simples flexionadas e não flexionadas, apenas as palavras com sufixação derivacional e as formas verbais infinitivas e participiais apresentam valores de ocorrência acima de 1%. Todas as outras formas apresentam valores inferiores.

Os gráficos seguintes permitem comparar os resultados computados sobre *types* e sobre *tokens*, em valores percentuais e absolutos, para os tipos de palavras em função da sua estrutura morfológica.

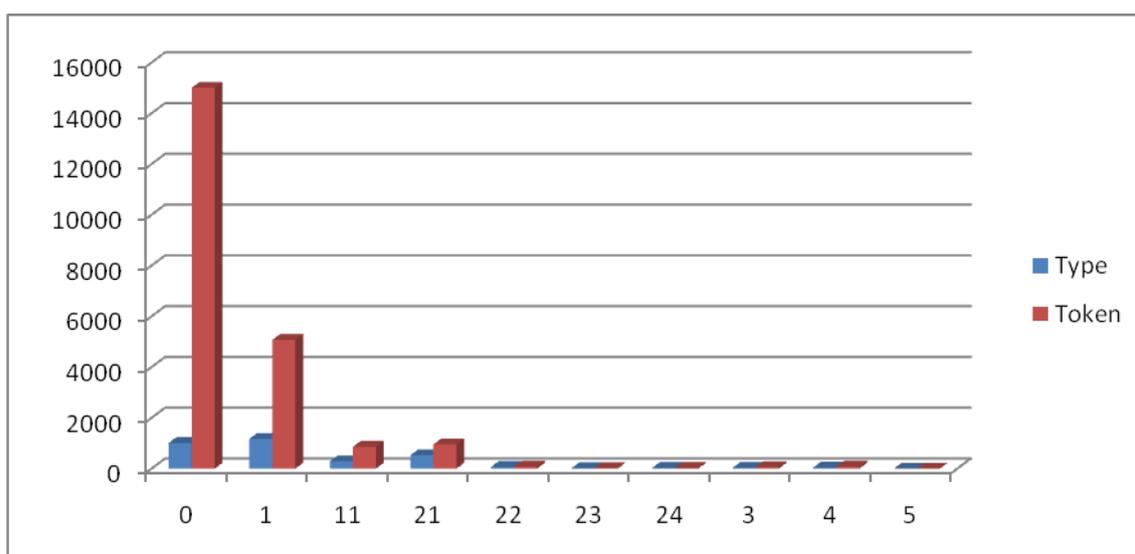


Figura III.28. Tipo de palavras em função da estrutura interna (valores absolutos) - Adulto (type/token).

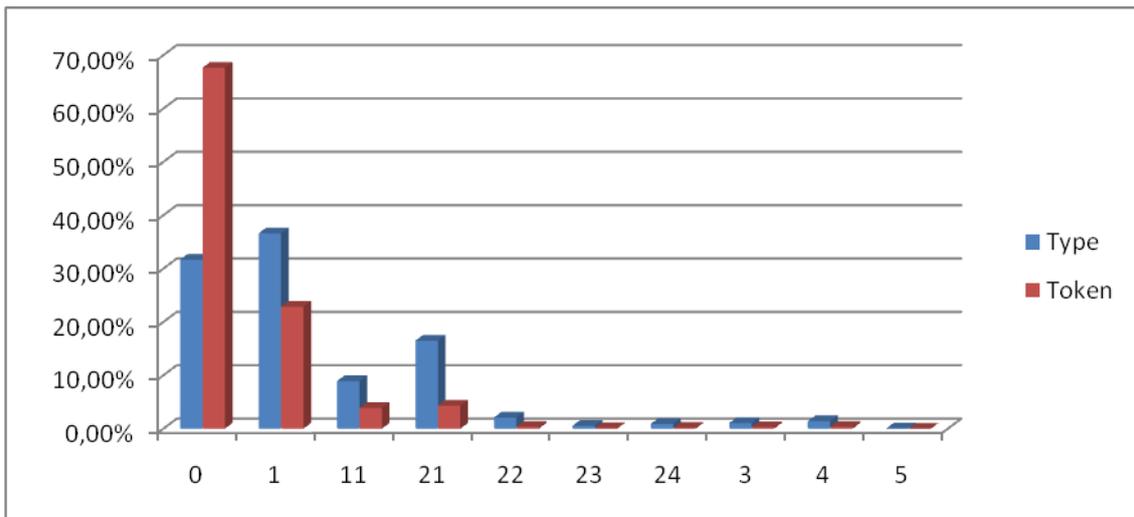


Figura III.29. Tipo de palavras em função da estrutura interna (valores percentuais) - Adulto (type/token).

Verifica-se que existem diferenças significativas nas distribuições das frequências de ocorrência dos tipos de palavras em função da sua estrutura, comparando *types* e *tokens*. Antes de mais, dentro das palavras simples, as palavras mais frequentes nos *types* são as flexionadas, enquanto são as não flexionadas nas contagens obtidas sobre *tokens*.

Mas da comparação dos resultados em função da contagem em termos de *token* e *type*, destaca-se sobretudo o aumento da predominância de palavras simples nos *tokens*, quando comparada com os *types* e a consequente diminuição na frequência relativa de palavras complexas naquele tipo de contagem. Efectivamente, mesmo se incluirmos na categoria de palavras complexas as formas nominais dos verbos, estas palavras não ultrapassam os 10% do total de *tokens*, contra os 31,43% se a mesma contagem for feita considerando os *types*.

Observando especificamente, as palavras derivadas, compostas e com sufixação (z-)avaliativa, verificamos que as primeiras são mais frequentes do que as restantes nos dados *type* e *token*. Contudo, a diferença entre estes três tipos de palavras é bastante mais significativa nas contagens em termos de *types* relativamente aos *tokens*: 20,02% de palavras derivadas contra 1,06% de palavras compostas e 1,40% palavras com sufixação (z-) avaliativa (*type*) versus 4,92% de palavras derivadas, 0,29% de palavras compostas e 0,35% de palavras com sufixação (z-) avaliativa (*token*).

Esta diferença está também patente na percentagem de ocorrência do tipo de palavras complexas mais frequente: considerando apenas as palavras com derivação sufixal, os valores nos *types* são muitíssimo superiores aos encontrados nos *token* (16,50% contra 4,31%, respectivamente).

Note-se ainda que as palavras prefixadas, compostas e com sufixação (z-) avaliativa ocorrem em percentagens inferiores a 0,4% na contagem em termos de *tokens*, enquanto nas contagens em termos de *types* a ocorrência destas palavras apresenta valores acima de 1%.

II) 3.3 Classes de Palavras

Os dados das classes de palavras não foram obtidos pela nossa análise, foram facultados pela Investigadora Amália Mendes e correspondem ao *corpus* do Português incluído no C-ORAL-ROM, como descrito na secção 2.5 do capítulo *Metodologia*. Tal como aí referido, foi possível deste modo obter dados correspondentes à frequência em termos de *token*; efectivamente, tal não seria possível fazer utilizando a base de dados do adulto usada nas restantes subsecções, por esta incluir apenas as listagens de palavras e respectivas frequências, sem informação sobre o contexto sintáctico em que ocorrem, inviabilizando a sua categorização.

A informação sobre as classes gramaticais no adulto permite-nos determinar a que classes pertencem as palavras complexas no adulto e sua frequência relativa. Tal como nas subsecções anteriores a partir destes resultados podemos avaliar a relação entre os dados do adulto e os descritos para a criança.

De seguida apresentamos as classes de palavras mais utilizadas pelo adulto, em valores percentuais e absolutos.

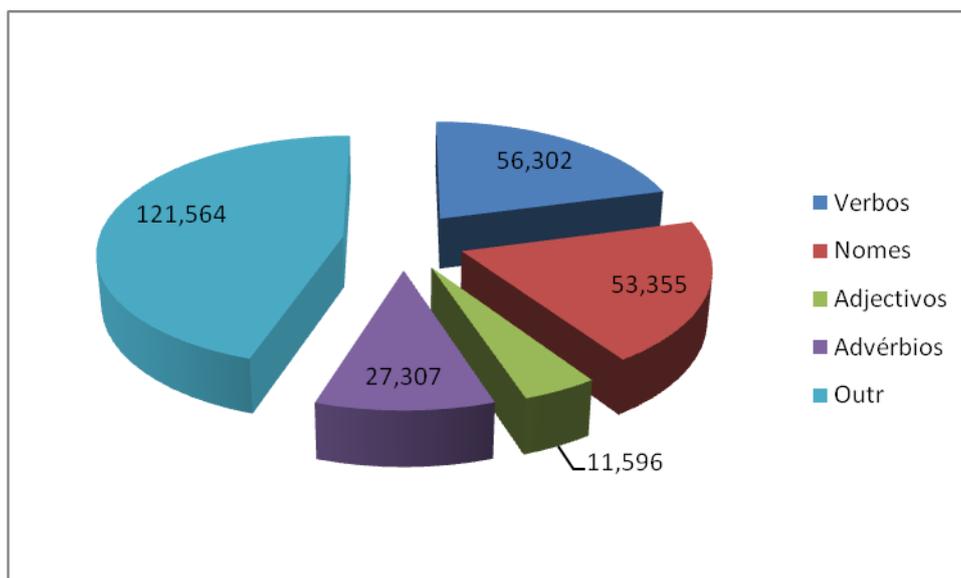


Figura III.30. Classes de Palavras (valores absolutos) - Adulto (token)

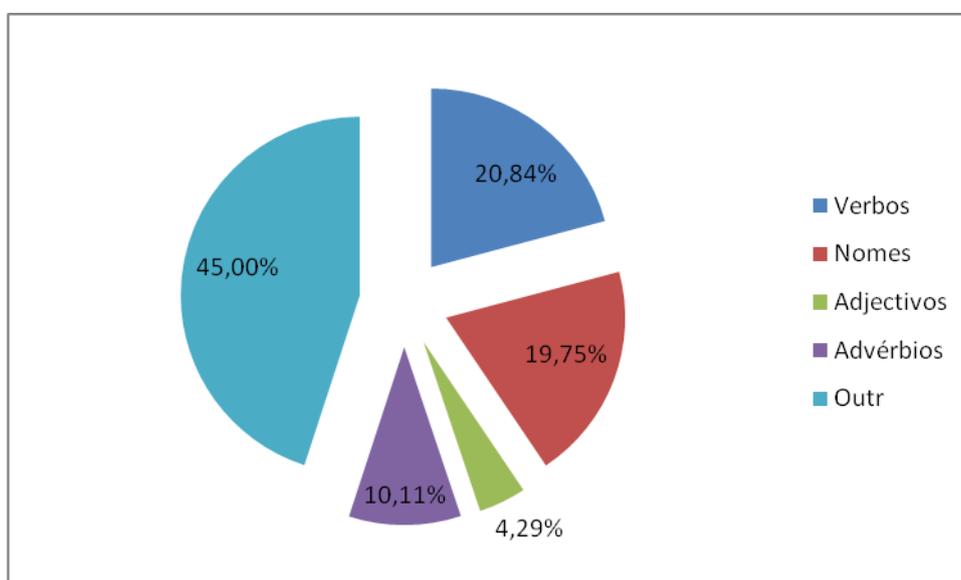


Figura III.31. Classes de Palavras (valores percentuais) - Adulto (token).

Os resultados de frequência das categorias de palavras utilizadas pelo adulto revelam que a categoria mais utilizada é a categoria de palavras *Outr*, correspondendo a

45%, quase metade das palavras produzidas pelo adulto. Esta categoria contém todas as palavras de classe fechada, tais como preposições, artigos e conjunções.

Dentro das classes abertas, o adulto utiliza mais frequentemente *Verbos* e *Nomes*, em proporções muito semelhantes (21% e 20% do total de palavras, respectivamente).

Os *Advérbios* representam 10% do total de palavras, sendo a classe dos *Adjectivos* a menos frequente na fala do adulto, com uma percentagem de 4%.

II) 3.4 O tamanho de palavras simples e complexas no adulto

Nesta subsecção foi analisado o tamanho das palavras simples e complexas que constituem o discurso do adulto, em termos de número de sílabas das palavras.

Considerou-se esta informação relevante, pois comparando os dados do adulto com os da criança poderemos avaliar a este respeito o papel do *input* no comportamento infantil.

Serão de seguida apresentados os resultados referentes á frequência dos diferentes formatos de palavras (isto é, tamanhos, em número de sílabas) das palavras simples na fala adulta, em valores absolutos e percentuais.

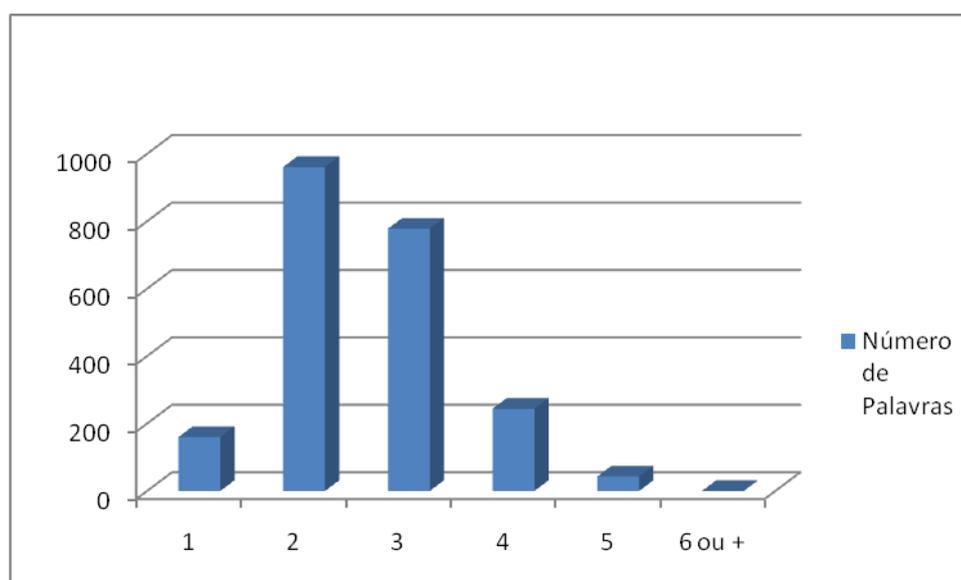


Figura III.32. Número de sílabas (valores absolutos) - Adulto (palavras simples).

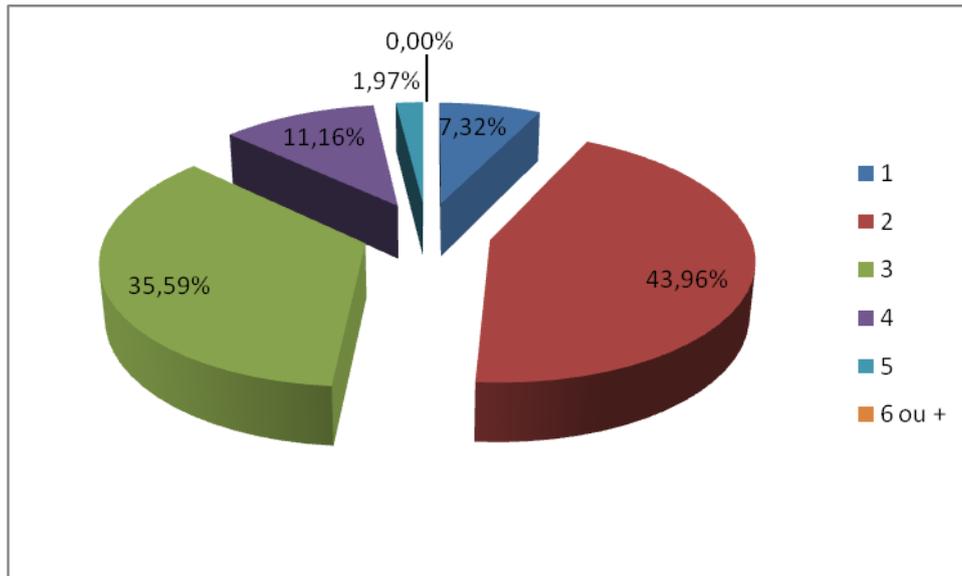


Figura III.33. Número de sílabas (valores percentuais) - Adulto (palavras simples).

As palavras simples do adulto são predominantemente dissilábicas e trissilábicas. As palavras com os formatos mais pequenos, com uma ou duas sílabas, representam claramente a maior fatia das palavras simples produzidas (51%). E estas juntamente com as palavras com três sílabas representam 87% do total de palavras.

As palavras com quatro sílabas representam 11,16%, sendo as palavras com cinco sílabas muito raras (1,97%).

No discurso do adulto, no *corpus* em observação, não se registam palavras simples com mais do que cinco sílabas.

Seguidamente apresentamos os valores relativos à frequência dos diversos formatos de palavras das palavras complexas produzidas pelo adulto, em valores percentuais e absolutos.

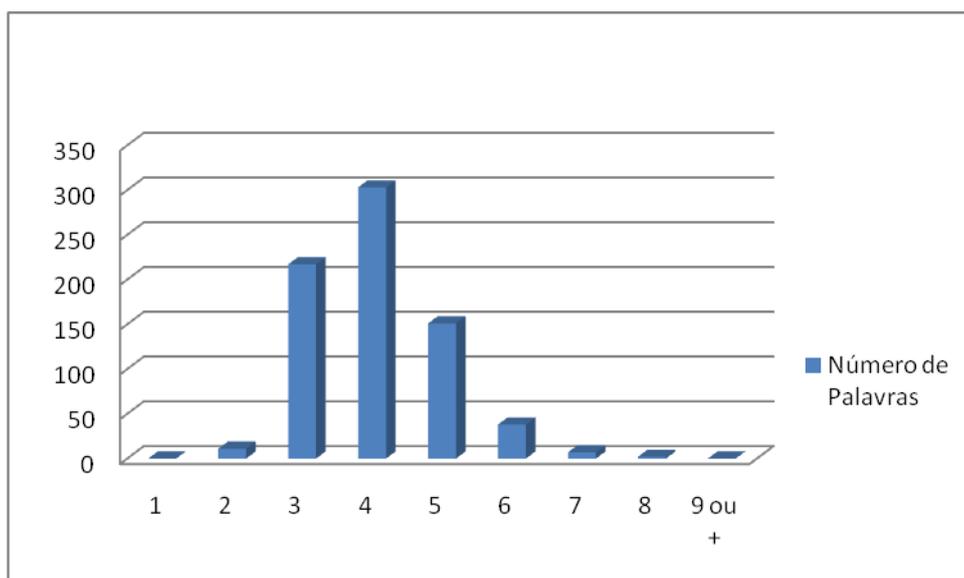


Figura III.34. Número de sílabas (valores absolutos) - Adulto (palavras complexas).

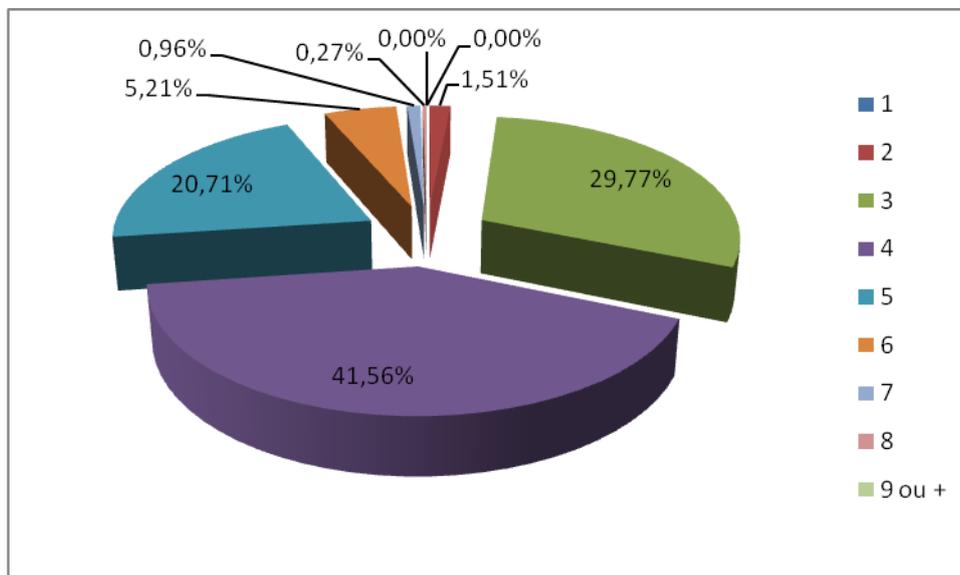


Figura III.35. Número de sílabas (valores percentuais) - Adulto (palavras complexas).

Relativamente às palavras complexas, o formato de palavras mais utilizada pelo adulto são as palavras quadrissilábicas, revelando um valor bastante elevado 41,56% do total de palavras complexas no adulto.

Seguidamente, as palavras complexas mais utilizadas pelo adulto contêm três sílabas (29,77%) e cinco sílabas (20,71%).

As palavras com seis sílabas correspondem a 5,21% do total de palavras complexas.

Em número diminuto encontram-se as palavras com duas, e com mais de seis sílabas, correspondendo a cerca de 1% das palavras complexas produzidas pelo adulto.

Neste tipo de palavras (complexas) produzidas pelo adulto são quase inexistentes os formatos inferiores a três sílabas, e as formas com quatro ou mais sílabas são claramente preponderantes, representando 69% do total de palavras complexas.

II) 3.5 Afixos utilizados

De seguida, são apresentados os vinte afixos mais utilizados nas palavras do adulto, entre os 123 afixos diferentes que isolámos.

Os resultados desta observação permitir-nos-ão determinar qual o grau de coincidência entre os resultados obtidos para o adulto e o encontrados na fala da criança.

Os vinte afixos mais frequentes na fala adulta e a sua taxa de incidência na *corpus* em estudo são apresentados no gráfico seguinte, valores absolutos e percentuais, contento os dados *type* e *token*.

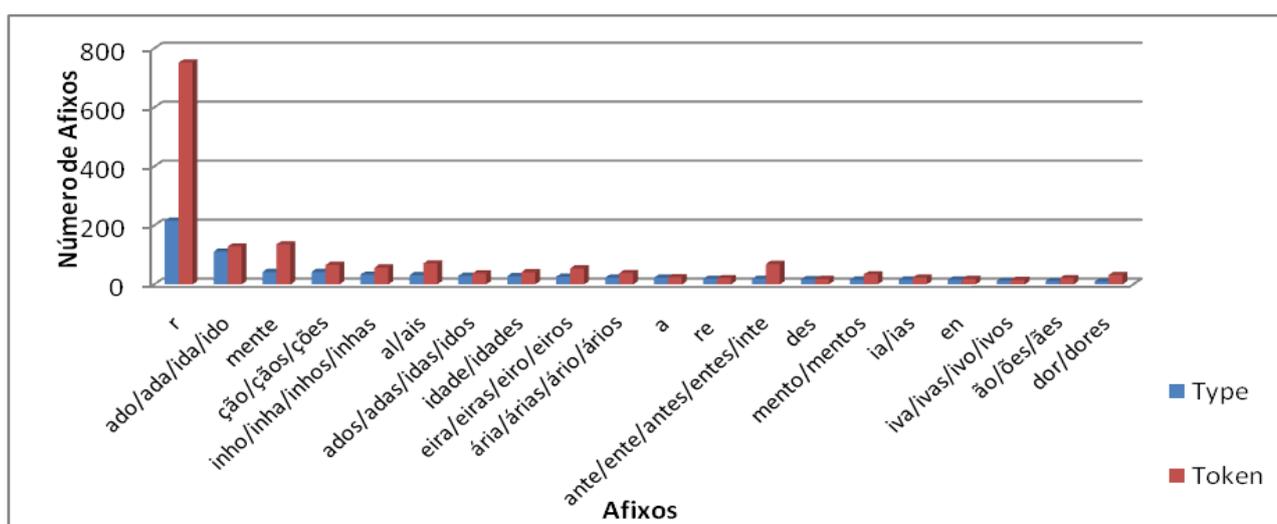


Figura III.36. Os 20 afixos mais frequentes na fala adulta e sua frequência (valores absolutos) (type/token).

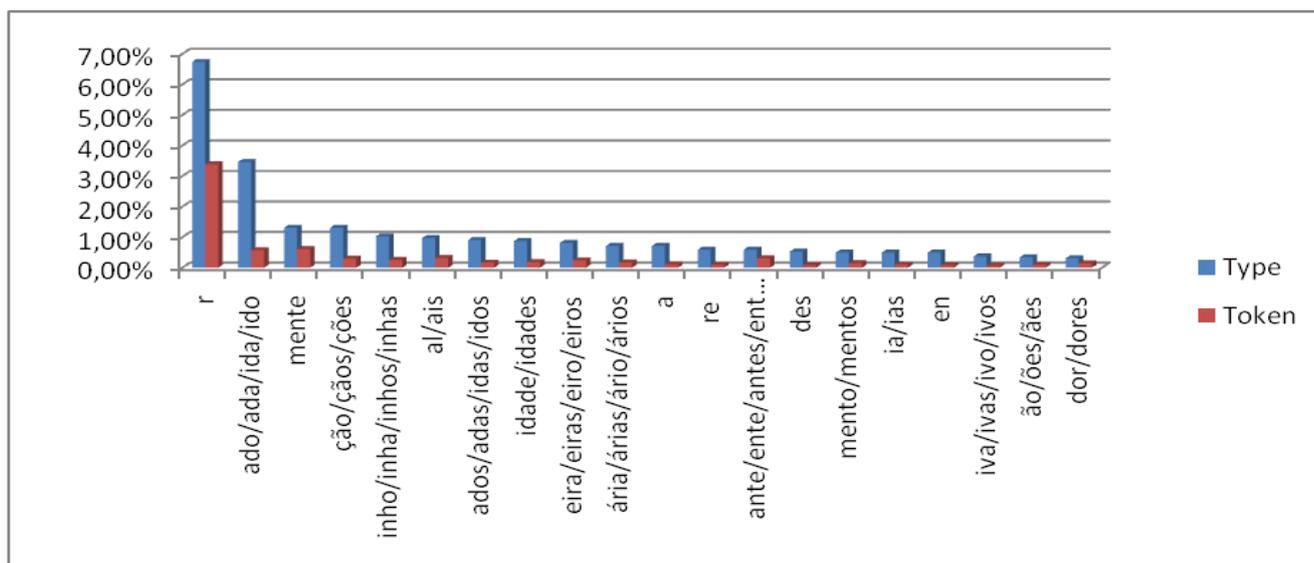


Figura III.37. Os 20 afixos mais frequentes na fala adulta e sua frequência (valores percentuais) (type/token).

Entre os resultados observados, salienta-se a preponderância do morfema do infinitivo. Esta saliência é ainda mais evidente na contagem feita sobre *token*.

Seguidamente, os três afixos mais frequentes para o adulto, tanto na contagem sobre *tokens* como na contagem sobre *types*, são *-ado/ada/ida/ido*, *-mente* e *-ção/çãos/ções*. Para além destes três afixos, os valores de ocorrência dos restantes, em termos absolutos, são bastante baixos, pois ocorrem menos de 50 vezes em todo o *corpus* nos *types* e menos de 70 nos *tokens*. Dos 123 afixos diferentes usados pelo adulto, apenas 22 (vinte e dois) apresentaram um valor igual ou superior a 10 (*type*). Toda a informação desciminada sobre a utilização de afixos encontra-se no Anexo 8 (*Afixos Adulto*).

Considerando a sua proporção face ao *corpus*, ela é ainda notoriamente mais diminuta, os afixos mais frequentes apresentam os seguintes valores: *-r* (6,74% e 3,39%), *-ado/ada/ida/ido* (3,46% e 0,58%), *-mente* (1,31% e 0,61%) e *-ção/çãos/ções* (1,31% e 0,30%), de acordo com as contagens considerando *types* e *tokens*, respectivamente. Os restantes afixos encontram-se abaixo destes valores, que já são bastante baixos. Note-se os valores de frequência dos afixos é sistematicamente bastante mais elevada na contagem feita sobre *types* do que na contagem feita sobre *tokens*.

Importa referir também a diminuta ocorrência de prefixos, quer na contagem sobre *types*, quer sobre *tokens*, pois os prefixos encontrados *re-*, *des-* e *-en* foram os únicos prefixos encontrados entre os vinte afixos mais utilizados. Estes prefixos apresentam os seguintes valores para *types* e *tokens* respectivamente: *re-* 0,59% e 0,09%, *des-* 0,53% e 0,08%, e *-en* 0,44% e 0,07%.

IV. DISCUSSÃO DOS RESULTADOS

Antes de iniciarmos a discussão dos resultados em relação aos objectivos propostos, importa lembrar que todos os resultados apresentados, à excepção dos resultados em relação ao tamanho de palavra, correspondem aos alvos da sua produção e não às suas efectivas produções, como já foi demonstrado na secção 2.3 do capítulo *Metodologia*.

Tendo em mente os objectivos e hipóteses formulados na secção anterior *Metodologia*, discutimos no presente capítulo os resultados descritos no capítulo anterior.

Em cada uma das subsecções contempladas neste capítulo os resultados do discurso da criança e do adulto serão comparados, tentando-se avaliar o papel do *input* na evolução do uso das palavras complexas pela criança.

4.1. Palavras simples versus palavras complexas

Nesta subsecção faremos uma discussão dos resultados descritos nas subsecções correspondentes do capítulo anterior, *Análise de Resultados*, relativas aos dados da criança e do adulto. Posteriormente, relacionamos os resultados dos tipos de palavras utilizados pela criança e pelo adulto.

Tal como foi sustentado anteriormente no capítulo *Enquadramento*, considerámos como tipos de palavras em função da sua complexidade morfológica, palavras simples, palavras complexas e palavras com reduplicação de sílabas.

De acordo com o discurso da criança, observámos que as palavras complexas encontram-se, para todas as idades, em número diminuto relativamente às palavras simples.

Nos primeiros meses das produções da criança, as palavras complexas são inexistentes e surgem as primeiras ocorrências aos 1;09 e 1;10 em número diminuto.

Verificámos que a criança utiliza a primeira palavra com reduplicação de sílabas aos 1;04 meses de idade, mas de modo episódico.

As palavras complexas, no primeiro ano da criança encontram-se entre os 10 a 20% do discurso total da criança, tanto com base nos dados *type* como *token*. Estes valores elevados podem ser justificados por dois factores. Primeiro, porque nesta idade, a criança produz poucas palavras e a dimensão do seu léxico é ainda reduzida, o que significa que as palavras complexas produzidas correspondem a uma grande percentagem a palavras únicas. Segundo, porque estes valores ainda que elevados apenas correspondem a duas idades (1;09 e 1;10) em que este tipo de palavras foram produzidas.

Embora durante o segundo ano de idade da criança, o número de palavras complexas se mantenha proporcionalmente muito abaixo do número de palavras simples, observámos que a partir dos 2;03 a criança inicia a utilização constante deste tipo de palavras.

Aos 2;03 e a partir desta idade, a criança utiliza palavras complexas através dos processos composição e derivação de forma sistemática. Nesta fase, o uso sistemático de palavras complexas pode ser explicada por dois factos concomitantes: o aumento de palavras no léxico da criança e o aumento do uso de formatos de palavra maiores (questão discutida mais à frente). Efectivamente, nesta idade (2;03) a criança produz 144 palavras (*type*) e 292 (*token*), verificando-se um aumento significativo de palavras pois, na idade anterior (2;02 anos), a criança produz menos de um terço de palavras únicas (41) e uma quantidade proporcionalmente muito inferior de *tokens* (57).

O número de palavras complexas vai aumentando de acordo com o crescimento da criança e com o aumento no número de palavras mensais. Pelo contrário, o número de palavras consideradas na categoria de palavras com reduplicação e que reflectem as palavras com reduplicação de sílabas, diminuem ao longo do crescimento da criança e do léxico. Este padrão aproxima-se do observado no adulto, onde verificámos que a presença de palavras reduplicadas é virtualmente inexistente. Este facto confirma a reduplicação como tendo um lugar particular no léxico inicial, onde existe uma maior incidência de palavras deste tipo do que em estádios mais avançados e do que no léxico do adulto.

No segundo ano de idade, as palavras complexas ocorrem entre os 3,18% do discurso total (2;03 anos) e os 14,45% (2;08 anos) (contagem sobre *types*). Quanto aos valores para os *tokens*, o mesmo tipo de palavras atinge os 2,24% (2;03), como valor mínimo e os 6,17% aos 2;06 anos (valor máximo nesta idade).

Por fim, no terceiro ano de idade a percentagem de palavras complexa ocorrem entre os 6,58% e os 11,37% do discurso total (valores *type*) e os 2,95% e 5,85% (valores *token*). Em valores médios, no terceiro ano as palavras complexas correspondem a 9,57% dos *types* e a 4,57% dos *tokens*.

Estes valores relativos aproximam-se dos exibidos pelo adulto, que também apresenta um número diminuto de palavras complexas relativamente às palavras simples: estas últimas correspondem nas suas produções a 77% dos *types* e a 94% dos *tokens*, contra apenas 22,74% de palavras complexas no total de *types* e 6% no total de *tokens*.

Os dados do adulto e da criança convergem também na evidente maior expressão de palavras complexas em todo o *corpus* quando considerados os *types*, relativamente aos *tokens*. Contudo, a proporção de palavras complexas no discurso da criança até ao final do período em observação é claramente inferior à verificada no adulto quando considerada a contagem sobre *types*, onde os valores máximos observados atingem pouco mais de metade dos contabilizados no *corpus* do adulto. Pelo contrário, os valores máximos de ocorrência de palavras complexas nas contagens sobre *tokens* são idênticos aos do adulto, e a média não se afasta já muito dos valores do adulto. Neste sentido, o tipo de dados do adulto que melhor se correlaciona com os observados na criança parece ser o proveniente da computação sobre *tokens* e não sobre *types*.

Vigário, Frota & Martins (2010) investigam também sobre que tipo de dados a criança computa a frequência, admitindo que ela desempenha efectivamente um papel importante no processo de aquisição: sobre o vocabulário do *input* (listagem de palavras únicas – *types*) ou sobre a ocorrência de todas as instâncias das palavras (*tokens*).

Existe informação de que os valores de frequência para *types* e para *tokens* numa mesma língua podem divergir muito. O facto de nuns trabalhos se mostrar a relevância da frequência relativamente às contagens computadas sobre *types* e noutros relativamente às contagens baseadas em *token* pode significar que ambos os tipos de dados podem ser importantes para extrair a informação de frequência relevante, podendo eventualmente até variar em função do aspecto fonológico em aquisição ou da fase de aquisição (Vigário, Frota & Martins 2010). Contudo, na sua avaliação do poder predictor, para a emergência e desenvolvimento linguístico, da frequência no *input* e observando vários aspectos da fonologia, os autores verificaram que, sempre que os valores de frequência computados sobre *types* e sobre *tokens* variam, fazendo previsões

diferentes, o melhor preditor para o comportamento encontrado na criança é o computado sobre *tokens*. Efectivamente, os resultados do estudo apontam fortemente para que seja a frequência em *tokens* a relevante para a aquisição da fonologia (Vigário, Frota & Martins 2010). Os dados agora apresentados parecem sugerir o mesmo em relação à ocorrência de palavras complexas na criança.

Um aspecto importante observado é o aumento bastante significativo do número de palavras e da sua frequência, em termos de *token e type*, à medida que a criança cresce. À medida, que a criança vai crescendo, também o seu léxico vai aumentando. Se, no início desta análise (1;01 anos) a criança utiliza um reduzido número de palavras, pertencentes a poucas classes gramaticais, com o aumento da idade, o número de palavras aumenta e a variedade de classes gramaticais também.

Segundo Clark (1995), nos primeiros meses enquanto falantes, as crianças falam sobre conceitos, actividades e rotinas do dia-a-dia. Quando vão adicionando palavras ao seu vocabulário, elaboram cada domínio semântico, reorganizando e dividindo cada termo disponível para a produção. Em cada ano que vai passando, as crianças vão acrescentando novos conceitos ao seu vocabulário e encaixando essas novas palavras em categorias e subgrupos.

Na bibliografia analisada, não se encontraram referências ao momento específico do aparecimento de palavras complexas na aquisição das línguas, a não ser a identificação dos dois anos como idade para a produção das primeiras palavras compostas para várias línguas (Booij, 2007; Libben & Jarema, 2006).

Neste estudo, a criança produziu a primeira palavra complexa ao 1;09 anos através do processo de derivação e não por composição como a bibliografia sugere. Este ponto é explorado na subsecção seguinte.

4.2. Tipos de Palavras em função da sua estrutura interna

Analisámos as palavras simples e complexas, quer da criança quer do adulto, de acordo com a sua estrutura interna. Os resultados foram apresentados nas secções I.3.2 e II.3.2 do capítulo *Análise de Resultados*.

Nesta subsecção discutiremos os resultados obtidos e relacionamo-los.

Foram considerados todos os processos de formação de palavras para o Português Europeu bem como a reduplicação de sílabas. Apenas o processo

morfológico *conversão* foi ignorado, uma vez que não foi utilizado pela criança e, no adulto, não temos a informação se foi utilizado ou não, pois como foi referido anteriormente, os processos morfológicos não foram analisados de acordo com as classes gramaticais.

O discurso da criança é marcado por palavras simples, sem afixo ou com sufixo flexional. As palavras complexas encontram-se em número reduzido, como vimos. Estes resultados são observáveis tanto nos *types* como nos *tokens*.

Verificámos que as palavras que contêm reduplicação de sílabas surgem aos 1;04 anos, idade em que ainda não usa palavras complexas. A partir dos dois anos de idade, a criança utiliza sempre este tipo de palavras. A utilização deste tipo de palavras vai diminuindo à medida que a criança cresce, tomando os valores mais baixo nos três anos.

Um das primeiras observações efectuadas neste estudo, é a de que esta criança utiliza as suas primeiras palavras complexas ao 1;09 (um ano e nove meses).

Uma das questões que colocámos no início da elaboração deste estudo foi se a emergência de palavras complexas estaria relacionada com o tamanho do léxico e/ou com o tamanho de palavra.

Em relação ao tamanho de palavra, discutiremos o assunto mais à frente, em secção específica. Em relação ao tamanho do léxico, analisando as idades imediatamente anteriores, não se verifica que exista um aumento de palavras no léxico da criança que justifique a emergência de palavras complexas (tanto nos dados *type* como *token*), nesta idade.

As primeiras palavras derivadas surgem em idades bastante precoces aos 1;09 anos, utilizando simultaneamente a sufixação e a prefixação.

As palavras compostas surgem aos 2;03 anos, sete meses depois.

A partir desta idade, a criança utiliza sempre estes dois tipos de palavras complexas.

De acordo com vários autores, referidos anteriormente, o processo de composição é dos mais frequentes em várias línguas, surgindo na aquisição por volta dos dois anos de idade. Para a língua portuguesa, tendo em conta os resultados desta criança, esta assumpção é semi-verdadeira. Se por um lado, as palavras formadas por composição surgem, efectivamente, numa idade que ronda os dois anos de idade, este não é o processo de formação mais frequente nas palavras complexas que a criança usa.

Observando os valores percentuais de palavras derivadas e compostas produzidas pela criança, verificamos que apenas na idade 2;04 as palavras compostas apresentam valores superiores às derivadas, tanto nos *types* como nos *tokens*.

Nas restantes idades, as palavras derivadas são sempre mais frequentes do que as compostas, facto evidente tanto nos *types* como nos *tokens*. Os dados de aquisição aqui descritos para o Português Europeu divergem, portanto, a este respeito dos colhidos na literatura sobre outras línguas.

No interior da classe de palavras derivadas, as mais frequentes, de modo geral, são as derivadas por sufixação. Contudo, os dois tipos de palavras derivadas mais utilizados, derivadas por sufixação e derivadas por prefixação, emergem na mesma idade. As ocorrências de formas com sufixação derivacional e prefixação, mas não simultânea (23) e as formas com sufixação derivacional e prefixação simultânea, circunfixação, (24), são raras e tardias.

A partir da idade de dois anos e quatro meses a criança usa as primeiras palavras com sufixação (z-) avaliativa, e utiliza-as em todas as idades subsequentes.

As formas no Infinitivo, Particípio Passado e Gerúndio surgem ao 1;09. Contudo, a produção constante deste tipo de palavras inicia aos 2;03. A sua evolução ao longo da idade da criança é positiva, pois a criança produz mais palavras deste tipo à medida que cresce.

A partir dos dois anos e cinco meses a criança utiliza palavras exibindo mais do que um processo morfológico; ainda que com uma frequência bastante baixa, é notável a utilização deste tipo de palavras em idades precoces.

Relativamente às palavras complexas, as formas mais utilizadas pela criança vão alternando, de acordo com a idade, entre as que exibem morfologia derivacional e sufixação (z-)avaliativa, que apresentam quase sempre valores aproximados.

As palavras compostas são em geral claramente menos frequentes do que os dois restantes tipos de palavras. Contudo, elas ocorrem na mesma altura em que aparecem as palavras que exibem os outros dois processos de formação de palavras de modo constante, ocorrendo em todos os meses a partir desta altura.

Nos últimos quatro meses observados, as palavras complexas atingem uma média de 9,57% (*type*) e 4,5% (*token*).

Verifica-se nos dados da criança uma preponderância das palavras simples não flexionadas sobre as flexionadas. Também aqui os dados do adulto computados sobre

tokens parecem ser os que se correlacionam com o comportamento infantil, uma vez que estes, mas não os *types*, apresentam a mesma tendência.

No discurso do adulto o processo de derivação é mais utilizado (20,02% *type* e 4,91% *token*) contra os (1,06% *type* e 0,29% *token*) do processo de composição.

Em relação à frequência dos diferentes tipos de palavras complexas na criança, comparativamente ao adulto, como já foi referido anteriormente, o processo de derivação é mais utilizado em relação ao processo de composição, em ambos os casos.

De acordo com os resultados obtidos verificamos dois factos. O primeiro é relativo aos dados que melhor se correlacionam com a criança e que podem funcionar como predictores do comportamento linguístico da criança. Trata-se da observação de que na frequência de palavras complexas, os valores *token* do adulto se relacionam melhor com os dados da criança, tal como observámos na secção anterior acerca da complexidade morfológica das palavras. O segundo diz respeito à formação de palavras complexas e à estrutura interna das palavras complexas. Verificámos no presente estudo que no discurso desta criança os compostos são produzidos em idades precoces, mas não de forma frequente e que o processo mais utilizado para formar palavras é a derivação.

Um aspecto que merece discussão adicional diz respeito ao significado que pode ter o uso que a criança faz de palavras complexas. Como salientámos no capítulo inicial desta tese, o facto de uma palavra ter estrutura interna não implica que os falantes tenham necessariamente consciência da complexidade da palavra. De igual modo, o facto de a criança usar palavras complexas, não implica necessariamente que ela adquiriu o processo morfológico exibido pela palavra. Efectivamente, a criança pode ter aprendido a palavras como uma unidade não-analisada, dessa forma armazenada no léxico mental.

Contudo, verificámos que esta criança entre os 2;04 e os 2;06 formou palavras formadas não coincidentes com as formas do adulto (como *quentezinho*, *chuchazinha*), sugerindo que a criança já tem adquirido algum tipo de regra de formação de palavras complexas. Interessantemente, o período em que ocorrem estas distinções coincide com aquele em que a criança começa a utilizar palavras complexas, sugerindo que se encontra em fase de aquisição da regra.

Não é claro se este é o momento do processo de aquisição em que a criança adquire os diversos processos de formação de palavras, ou se o verificado é apenas válido para a sufixação avaliativa. Este assunto merece exploração em investigação

futura. Em particular, havendo outros dados recolhidos por outras fontes relativas aos mesmos períodos de produção desta mesma criança, será importante verificar se em todas as palavras produzidas que se encontram registadas existirão palavras claramente formadas pela criança, não coincidentes com as formas adultas, em que estejam exibidos outros processos de formação de palavras.

4.3. Classes de Palavras

O estudo da classe das palavras utilizadas pela criança desde as primeiras produções justificou-se pelo facto de esta informação poder ser relevante para a formação de palavras complexas.

Discutiremos nesta subsecção os dados provenientes dos resultados do discurso e da criança e posteriormente comparamo-los com os dados resultantes do discurso do adulto.

Analisando as classes de palavras mais utilizadas pela criança e pelo adulto, verificou-se que existem resultados divergentes.

De acordo com os dados *type*, a criança dá primazia aos *Nomes* e, em segundo lugar, utiliza mais *Verbos* no seu discurso. Mas nos *tokens*, estes não uniformes em todas as idades: até aos 2;07 anos a criança utiliza mais *Nomes* e *Verbos*; a partir desta idade, a categoria mais utilizada é *Outr*, seguida de *Verbos* e em terceiro lugar *Nomes*.

Em relação às categorias gramaticais e de acordo com Boysson-Bardies (1999) (citado por Herschensohn, 2007), as crianças adquirem primeiro os nomes, constituindo estes a maior parte do seu vocabulário na maioria das línguas.

No caso do Inglês Americano, a predominância de nomes está directamente relacionada com os aspectos linguísticos e culturais do *input*. (Herschensohn, 2007).

Correia (2009) estudou o discurso de cinco crianças incluído Luma, mas usando uma outra fonte de dados, e concluiu que em todas elas, a produção de nomes é superior em idades mais precoces do que os verbos. À medida que as crianças crescem, a quantidade de verbos aumenta.

Segundo Turkey (2006) os nomes são adquiridos primeiro porque correspondem a categorias básicas de objectos. Contudo existem estudos que apontam para que os

verbos possam ser adquiridos primeiro do que os nomes se as crianças estiverem expostas a mais verbos do que nomes.

De acordo com Gentner (1978, 1982), as crianças adquirem universalmente nomes antes dos verbos porque os nomes são mais fáceis de codificar do que os verbos; os nomes correspondem a entidades ou coisas e as crianças pequenas tem essa percepção do seu mundo físico.

As classes de palavras encontradas nas crianças podem estar relacionadas com as características individuais de cada língua.

Esta mudança na categoria gramatical mais utilizada até à idade 2;07 anos no Português pode estar relacionada com o aumento de palavras no léxico da criança ao longo da idade. Além disso, a categoria *Outr* é constituída por palavras que pertencem à classe fechada (como determinantes, preposições, artigos,...). O uso destas palavras na criança inicia-se durante o segundo ano de idade. Em idades mais precoces, a criança omite frequentemente este tipo de palavras. O uso crescente das palavras pertencentes a esta categoria pode estar relacionado com o desenvolvimento da sintaxe.

A classe dos *Adjectivos* é aquela que surge mais tardiamente aos 2;01 anos, que também pode correlacionar-se com o desenvolvimento sintáctico: enquanto *Nomes* e *Verbos* podem ocorrer sozinhos, tipicamente os *Adjectivos* acompanham outras palavras *Nomes*, pelo que o seu aparecimento pode estar mais dependente do desenvolvimento de estruturas sintácticas mais complexas e por isso mais tardias.

Quanto à classe dos *Advérbios*, é utilizada pela primeira vez aos 2;00 e é a classe menos frequente na maioria das idades, de acordo com os dados *types*. Considerando os valores computados sobre *token*, a classe menos frequente são os *Adjectivos*.

A partir da idade 2;01 a criança utiliza todas as classes gramaticais no seu discurso.

Tal como na criança, também no adulto a classe mais utilizada é a de palavras *Outr*, correspondendo a quase metade das palavras produzidas pelo adulto (45%). Estes dados suportam aqueles encontrados na criança, mas relativamente aos dados *token* e para um período de produções mais tardio, em que a classe mais utilizada é já a de *Outr*.

O mesmo é verificado para a ordem de classes mais frequentes, pois o adulto utiliza, em segundo lugar, *Verbos* e com valor bastante próximo da classe de *Nomes*, que surge em terceiro lugar.

As palavras da classe dos *Adjectivos* são as menos utilizadas pelo adulto, tal como a criança (*token*).

Como verificámos, relativamente à complexidade morfológica e estrutura interna das palavras, estes resultados sugerem que a computação sobre *tokens* constitui um melhor predictor dos resultados da criança, pois todos os resultados encontrados para o discurso do adulto (também computados sobre *tokens*) se correlacionam com os dados da criança nas contagens sobre *token*, mas não necessariamente sobre *types*.

Relativamente a esta criança, e tendo em consideração os resultados *token* (pois são estes que são confirmados pelos dados obtidos pelo adulto), a categoria de palavras mais utilizada é a categoria gramatical *Outr*. Este resultado poderá ser explicado pelas características do Português Europeu (por exemplo, marcando com preposições o que outras línguas marcam com desinências casuais).

Os valores exibidos pela criança, relativamente às classes abertas e fechadas (84% e 65% - classe aberta e 14% e 35% - classe fechada (*types* e *tokens*, respectivamente) durante o período entre os 2;07 e os 3;03 vão a caminho dos resultados encontrados no adulto (55% classe aberta e 45% classe fechada). O aumento no uso das palavras de classe fechada também reflectem o desenvolvimento sintáctico, com o uso de palavras gramaticais, que no adulto representam mais de 30% do total de palavras produzidas (cf. Vigário *et al.* 2006).

4.4. Tamanho de Palavras Simples e Complexas

Concentremo-nos agora no papel do tamanho de palavra na emergência de palavras complexas.

Ao iniciar este trabalho, colocámos a hipótese de o tamanho das palavras poder vir a influenciar a emergência de palavras complexas, isto é, questionámo-nos se a criança apenas começa a utilizar palavras complexas quando já tem adquirido/produziu palavras maiores (com três sílabas ou mais).

De acordo com a informação disponível na literatura, em línguas tão diversas como o Inglês, o Yidin, o Árabe, o Japonês, o Lardil, o Estónio, o Bengali, o Baule, o Chamicuro, o Alemão ou o Catalão, a palavra prosódica apresenta um tamanho mínimo definido, sendo pelo menos dissilábica ou bimoraica. (Vigário, Martins & Frota, 2005).

Apesar de, segundo Vigário *et al.* (2006) esta hipótese não ser confirmada no Português Europeu, visto o nosso objectivo não ser a aquisição deste tipo de palavras,

fomos verificar se a emergência de palavras com três ou mais sílabas está directamente relacionada com a emergência de palavras complexas.

Verificámos que relativamente ao tamanho de palavra, e considerando os *alvos* e as produções efectivas, que existem diferenças de acordo com a idade em que a criança começa a produzir palavras complexas.

Nas palavras alvo, e até aos 2;03 anos, a criança realiza maioritariamente palavras monossilábicas e dissilábicas, mas também possui palavras trissilábicas e até palavras com quatro sílabas no seu discurso. Apesar de a partir desta idade observarmos um aumento do tamanho de palavras, o seu discurso é constituído por uma variedade considerável de palavras em relação ao número de sílabas em idades precoces.

A relação entre o surgimento de palavras complexas e o tamanho de palavra (em número de sílabas) é evidenciada quando temos em conta as palavras efectivamente produzidas pela criança. Até aos 2;03 (idade em que começam a surgir as palavras complexas de forma constante), a criança utiliza preferencialmente formas monossilábicas e dissilábicas. As formas com três sílabas encontram-se em idades pontuais. A partir desta idade existe um aumento significativo de palavras com três sílabas e surgem pela primeira vez as palavras com mais de três sílabas.

Quanto ao adulto, verificámos que existem claras disparidades no tamanho de palavras simples e complexas. Enquanto as palavras simples são maioritariamente monossilábicas e dissilábicas, as palavras complexas têm mais frequentemente quatro, três e cinco sílabas (por ordem de frequência). Neste tipo de palavras, não são observadas palavras monossilábicas, e as dissilábicas são muito raras, ocorrendo apenas 1,51%. Isto significa que o tamanho distingue em larga medida os dois tipos de palavras.

De acordo com os resultados obtidos para o tamanho de palavra no discurso das crianças e adultos podemos salientar dois aspectos.

Primeiro, o tamanho das palavras *alvo* e das realizações da criança são similares, sendo que as palavras realizadas são ligeiramente mais pequenas do que as palavras *alvo* e que as palavras com maior número de sílabas surgem mais tardiamente nas palavras realizadas.

Segundo, e mais importante, a criança apresenta palavras maiores no seu discurso, quer em termos de palavras *alvo*, quer considerando as realizações efectivas e de forma constante por volta dos dois anos de idade. Esta idade coincide com aquela em que a criança começa a utilizar palavras complexas no seu discurso.

A partir destes resultados, é concebível que para a emergência de palavras complexas na aquisição da criança possa desempenhar um papel importante, o desenvolvimento dos formatos de palavra maiores no discurso da criança.

4.5. Afixos mais utilizados

Por se ter considerado relevante para a economia deste trabalho, realizou-se o estudo de todos os afixos utilizados pela criança nas palavras derivadas, que se encontra em anexo (Anexo 3 e 4). Para a análise de resultados e para efeitos comparativos entre todas as idades, utilizámos os cinco afixos mais utilizados pela criança.

Quanto aos adultos, e por possuírem um conjunto mais vasto de informação, colocámos para análise e comparação de resultados os vinte afixos mais utilizados.

No primeiro ano de vida da criança, esta utiliza afixos em apenas três idades. Este facto poderá dar-se porque a criança produz poucas palavras derivadas nesta faixa etária.

Durante o segundo ano, a quantidade e diversidade de afixos aumenta significativamente, resultados, que estão directamente relacionados com o aumento de palavras derivadas que ocorre nesta idade.

Nesta faixa etária são também observados muitos afixos resultantes das palavras avaliativas, que são muito utilizadas pela criança nesta altura.

No terceiro ano, composto por quatro meses de análise das produções da criança, os resultados acompanham os dados anteriores. Existe um aumento significativo do número e diversidade de afixos utilizados pela criança, justificado pelo aumento de palavras utilizadas e pelo aumento do uso do processo de derivação como processo de formação de palavras.

Estabelecendo uma comparação entre os dados da criança e os dados do adulto, verificamos que estes suportam os resultados dos afixos mais utilizados na criança. Os vinte afixos mais utilizados pelo adulto, disponíveis na análise de resultados, contêm alguns dos afixos mais utilizados pela criança, nomeadamente, *r*, *ada/ado/ida/ido* e *inho/inha/inhos/inhas*.

O afixo *r* é o mais utilizado tanto pelas crianças como pelos adultos.

Pelo contrário, os afixos *mente*, *cão/çãos/çães* e *al/ais* utilizados frequentemente pelo adulto, não o são pela criança de forma frequente. Contudo, há que ressaltar que,

no adulto são considerados vinte afixos enquanto para a criança apenas são considerados cinco afixos.

Os valores de frequência no adulto são diminutos quando considerada a proporção de formas dos diferentes afixos relativamente ao total de palavras.

De acordo com Clark (1995), as crianças são também sensíveis às produções do adulto que as rodeia. Os afixos que elas utilizam em primeiro lugar parecem coincidir com aqueles que são mais produzidos e realizados pelos adultos (Fletcher, P. & MacWhinney, B., 1995). Demonstrando que existe uma clara influência do meio externo nas aquisições das crianças.

Por não ser comparável, não se realizou a relação de afixos entre o Português Europeu e outras línguas; Não foram encontradas referências específicas sobre os afixos mais utilizados na língua portuguesa.

V. Conclusão

Este trabalho assenta numa base de dados, LumaLiDa, com registos quase diários das produções verbais de uma criança, Luma, durante dois anos e dois meses, correspondentes aos primeiros 3 anos e 4 meses da vida da criança. Esta base de dados revela-se extremamente completa, quantitativa e qualitativamente, englobando diversas informações para além das meras produções verbais, extremamente úteis para o aprofundar de diversas áreas de investigação.

Para além de todas as informações que constituíam a base de dados inicial, este trabalho deu um ainda maior contributo, analisando cada uma das palavras ao nível da sua morfologia. Em cada palavra foi realizado um extenso trabalho de análise, quanto à sua classe gramatical, estrutura interna e existência ou ausência de afixos.

A investigação efectuada nos últimos anos sobre aquisição da gramática e da fonologia, especialmente, tem demonstrado que, a frequência de unidades e padrões linguísticos correlaciona-se com a ordem de emergência e as frequências dessas unidades e padrões nas produções das crianças (e.g. Ingram, 1988; Roark & Demuth, 2000, Demuth & Johnson, 2003, Levelt & Van de Vijver, 2004, Freitas, Frota, Martins & Vigário, 2006; Gülzow & Gagarina, 2007, entre muitos outros).

Dada a importância dos dados respeitantes ao adulto, como predictores dos resultados encontrados na criança, nesta investigação utilizámos a base de dados *Corpus* do Português Falado com o intuito de comparar os dados do adulto com os da criança.

Apesar de se tratar de um banco de dados com uma extensão considerável, esta investigação disponibilizou também informação nova referente ao discurso do adulto, nomeadamente em relação à frequência, classes gramaticais, estrutura interna e afixos utilizados.

Após a análise de resultados, revelam-se agora as conclusões retiradas. Assim uma visão global sobre as palavras que constituem o discurso da criança Luma, mostra que as palavras complexas se encontram em número diminuto face às palavras simples. No entanto, ao longo da idade a quantidade de palavras complexas vai aumentando, significando que a tendência destas palavras é ampliarem o seu número ao longo do crescimento da criança.

Uma categoria particular de palavras, a das formas reduplicadas, parece especialmente presente em fases precoces, registando-se uma diminuição da sua

frequência com o crescimento infantil. A diminuição do uso de formas com reduplicação em idades mais avançadas espelha a quase total ausência deste tipo de palavras nos dados do adulto.

A primeira conclusão deste estudo é a de que a criança utiliza as suas primeiras palavras complexas ao 1;09 anos através da derivação e passados sete meses utiliza a sua primeira palavra através do processo de composição.

Seguindo o percurso de análise efectuado, podemos dizer que no *corpus* analisado LumaLiDa, dentro das palavras complexas existem mais palavras derivadas do que compostas e com sufixação (z-) avaliativa; as palavras mono e dissilábicas caracterizam o discurso desta criança quer na realização efectiva, quer nos *alvos*; a classe de palavras mais utilizada é variável entre *Nomes* e *Outr* (categoria composta por palavras de classe fechada tal como, artigos, preposições, conjunções,...).

Relativamente ao discurso do adulto, a investigação sobre a base de dados analisada *Corpus* do Português Falado permitiu verificar que, também neste discurso, as palavras complexas se encontram em número diminuto relativamente às palavras simples. As palavras constituídas por reduplicação de sílabas estão limitadas a apenas uma palavra, sugerindo que este é um tipo de palavras característico do discurso da criança e do adulto quando dirigido à criança.

Também no discurso do adulto, concluímos que dentro das palavras complexas, as palavras derivadas encontram-se em maior número do que as palavras compostas e com sufixação (z-) avaliativa; quanto às classes gramaticais, este discurso apresentou um maior número de palavras *Outr*; relativamente ao tamanho de palavra, o discurso do adulto revelou uma predominância de palavras dissilábicas para as palavras simples e de palavras com quatro sílabas para as palavras complexas.

Concluimos também, que de acordo com as análises efectuadas aos *corpus* da criança e do adulto, que os morfemas mais utilizados pela criança e pelo adulto se correlacionam de um modo geral.

Tomando como ponto de partida as predições por nós estabelecidas no início da elaboração desta dissertação, acerca da importância da frequência do *input* e do tamanho de palavra, concluímos que os resultados encontrados no adulto predizem, na generalidade, os resultados que iremos encontrar na criança. Dentro dos dados do adulto, os melhores predictores do comportamento infantil são os computados sobre *tokens* (que englobam todas as palavras do discurso, repetições incluídas) e não sobre *types* (contabilizando apenas palavras únicas).

Quanto ao tamanho de palavra, verificámos que dele pode depender a emergência de palavras complexas, pois por um lado observámos que no discurso do adulto, as palavras complexas apresentam um maior número de sílabas, tendo quase exclusivamente três ou mais sílabas; por outro lado, observámos que é nas idades em que a criança realiza (ou tenta realizar) *alvos* com maior número de sílabas, que começa a utilizar palavras complexas.

Importa salientar, para concluir, que o facto de a criança produzir palavras complexas não implica que ela as tenha construído, ou mesmo que analise a sua estrutura interna. Pelo contrário, a criança pode armazenar no seu léxico mental as formas lexicais inanalizadas. Contudo, pelo facto de termos verificado que a criança produziu palavras diferentes daquelas encontradas no input, a partir dos 2;04 anos de idade, pressupõe-se que a partir desta idade, já inferiu regras de formação de palavras.

Não sendo claro o momento do processo de aquisição em que a criança adquire os diversos processos de formação. Este assunto merece exploração em investigação futura, nomeadamente, seria importante verificar se em todas as palavras produzidas que se encontram registadas existirão palavras claramente formadas pela criança, isto é, não coincidente com as formas do adulto.

Outro aspecto que julgamos que seria interessante explorar, diz respeito às eventuais especificidades do discurso do adulto dirigido à criança, o que não foi possível fazer nesta tese por limitações de tempo.

Esta dissertação constitui um estudo de caso. Os resultados reportados relativos ao uso de palavras complexas na criança foram obtidos através de uma análise descritiva, pelo que devem ser entendidos como tal. Acredita-se que o essencial do que foi aqui descrito não é específico desta criança, de acordo com a informação existente sobre a emergência relativamente precoce de palavras complexas nas línguas. Contudo, tal como referido, não existem estudos disponíveis que se debrucem sobre a emergência e evolução das palavras complexas durante a aquisição e desenvolvimento do Português Europeu. Efectivamente, este é, até onde sabemos, o primeiro estudo sistemático feito sobre a aquisição da morfologia (derivacional) do Português Europeu. Esperamos que ele possa abrir o caminho a outros estudos sobre a aquisição deste e doutros aspectos da aquisição da morfologia desta língua.

VI Referências

Aguiar, J. .2009. **Sílabas e tipos silábicos mais frequentes: análise de corpora**. *Diacrítica* 23/1, 7-22.

Booij, G. (2007) **The Grammar of words: An Introduction to Morphology**. Oxford Linguistics: Second Edition.

Corpus do Português Falado. Documentos Autênticos. Centro de Linguística da Universidade de Lisboa e Instituto Camões.

Corpus do Português C-ORAL-ROM (Integrated Reference Corpora for Spoken Romance Languages). RePort do Centro de Linguística da Universidade de Lisboa.

Correia, Susana (2009) **The Acquisition of Primary Word Stress in European Portuguese**. Dissertação de Doutoramento, Universidade de Lisboa.

Costa, T., M. J. Freitas, S. Frota, M. Vigário & F. Martins. 2007. **Sobre o PA na periferia esquerda da palavra**. Submitted to Textos Seleccionados do XXI Encontro Nacional da Associação Portuguesa de Linguística.

Cresti, E. & Monegnia, M. (eds.) (2005) **C-ORAL-ROM: Integrated ReferEnce Corpora for Spoken Romance Languages**. Amsterdam/Philadelphia: John Benjamins Publishing Company, Studies in Corpus Linguistics nº 15 (com DVD).

Cunha, C.; Cintra, L. **Nova Gramática do Português Contemporâneo**. Edições J. Sá de Costa (1994).

Demuth, K., & Johnson, M. (2003) Truncation to subminimal words in Early French. *Canadian Journal of Linguistics*, 48, 211-241.

Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa. Lisboa: Academia das Ciências de Lisboa. Editorial Verbo 2001.

Fletcher, Paul; MacWhinney, Brian, eds. (1995). **The Handbook of Child Language.** Oxford, Cambridge: Blackwell.

Freitas, M. J., S. Frota, M. Vigário & F. Martins. (2006) **Efeitos prosódicos e efeitos de frequência no desenvolvimento silábico em Português Europeu.** In Selecção de Comunicações apresentadas no XX Encontro Nacional da Associação Portuguesa de Linguística.

Fromkin, V.; Rodman, R.; Casanova, I. (1993) **Introdução à Linguagem.** Edições de Almedina.

Frota, S., Vigário, M., Martins, F., (2004) **FreP.** Lisboa: Laboratório de Fonética da FLUL (<http://www.fl.ul.pt/LaboratorioFonetica/FreP>).

Frota, S., Vigário, M. & Jordão, R. (2008) **LumaLiDaOn.** Lisboa: Laboratório de Fonética da FLUL (<http://www.fl.ul.pt/laboratoriofonetica/lumalidaon.htm>).

Frota, S., Vigário, M., Martins, F., Cruz, M. (em curso) **FrePOP – Frequency of Phonological Objects in Portuguese.** Laboratório de Fonética da Faculdade de Letras de Lisboa.

Frota, S., (2010) **Prosodic structure in early child speech: Evidence from intonation, tempo and coda production.** Workshop on Prosodic Development. Universitat Pompeu Fabra, April 16.

Gülzow I; Gagarina, N. (eds.) (2007) **Frequency effects in language acquisition – Defining the Limits of Frequency as an Explanatory Concept.** Berlin:Walter de Gruyter.

Herschensohn, J. (2007) **Language Development and Age.** Cambridge University Press.

Huber, J.; **Gramática do Português Antigo (tradução portuguesa)**. Lisboa: Fundação Calouste Gulbenkian 1986.

Ingram, J. (2007) **Neurolinguistics – An introduction to spoken Language Processing and its disorders**. Cambridge Textbooks in Linguistics.

Jordão, R (2009) **A Estrutura Prosódica e a Emergência de Segmentos em Coda no PE. Um Estudo de Caso**. Dissertação de Mestrado, Universidade de Lisboa

Lléo, C. (2006) **The Acquisition of Prosodic Word structures in Spanish by monolingual and Spanish-German Bilingual Children**. *Language and Speech*, 2006, 49(2), 205-229. Kingston Press Ltd.

Libben, G.; Jarema, G.; (2006) **The Representation and Processing of Compound Words**. Oxford Linguistics.

Mateus, M. H. M., A. M. Brito, I. Duarte, I. H. Faria, S. Frota, G. Matos, F. Oliveira, M. Vigário & A. Villalva (2003) **Gramática da Língua Portuguesa**, 6.^a edição, Lisboa, Editorial Caminho, 1127 pp.

McCrostie, J.; (2007) **Investigating the Accuracy of Teachers Word Frequency Intuitions**. *RELC Journal*: 2007; 38 (1): 53-66

Nunes, A.M; R.L. Coimbra (2007) **“Um Estudo da Amálgama e do seu Valor Metafórico em Mia Couto”**. Cano López, Pablo (coord.), *Actas del VI Congreso de Lingüística General*. Vol 2 Tomo 1, Madrid: Arco Libros (ISBN 84-7636-671-5), pp. 1465-1474

Prieto, Pilar (2006) **The Relevance of Metrical Information in Early Prosodic Word Acquisition. A Comparison of Catalan and Spanish**. *Language & Speech* 49 (2):233-261 (Special issue on the Crosslinguistic Perspectives on the Development of Prosodic Words).

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

Schmitt, N.; Dunham, B.; (1999) **Exploring native and non-native intuitions of word frequency**. *Second Language Research Online* (1999) 15: 389-411

Spencer, Andrew; Zwicky, Arnorld M., ed. (2001). **The Handbook of Morphology**. Blackwell Publishers.

Vigário, Marina (2003) **The Prosodic Word in European Portuguese**. Berlin/New York: Mouton de Gruyter

Vigário, M., F. Martins & S. Frota. (2005) **Frequências no Português: a ferramenta FreP**. In *Actas do XX Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, APL, 897-908.

Vigário, M., F. Martins & Frota. S. (2006) **A ferramenta FreP e a frequência de tipos silábicos e classes de segmentos no Português**. In *Seleccção de Comunicações apresentadas no XX Encontro Nacional da Associação Portuguesa de Linguística*.

Vigário, M., Freitas, M. J. & Frota, S. (2006) **Grammar and frequency effects in the acquisition of prosodic words in European Portuguese**. *Language and Speech* (Special Issue Crosslinguistic Perspectives on the Development of Prosodic Words, guest-edited by K. Demuth) 49(2), 175-203.

Vigário, M., F. Martins & S. Frota. (2006) **A ferramenta FreP e a frequência de tipos silábicos e de classes de segmentos no Português**. In Fátima Oliveira & Joaquim Barbosa (orgs.). *Textos Seleccionados do XXI Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: APL, pp. 675-687.

Vigário, M.; Frota, S.; & Martins F. (2010) **A frequência que conta na aquisição da fonologia: types ou tokens?** Universidade de Lisboa.

Villalva, Alina; **Estruturas Morfológicas. Unidades e Hierarquias nas Palavras do Português**. Lisboa: Fundação Calouste Gulbenkian 1994,2000.

Villalva, Alina; **Morfologia do Português**. Lisboa: Universidade Aberta 2008

Weissenborn, Jürgen; Hohle, Barbara (2001). **Approaches to Bootstrapping: Phonological, Lexical, Syntactic and Neurophysiological Aspects of Early Language Acquisition**. Amsterdam: John Benjamins.

<http://www.ling.upenn.edu/Events/PLC/plc33/abstracts/Dhillon.pdf>

<http://www.ddl.ish-lyon.cnrs.fr/Divers/index.asp?Action=Edit&Langue=EN&Page=Agenda&Jour=All&Mois=4&Annee=2006&Img=CarteDouble->

<http://webh01.ua.ac.be/apil/apil112/apil112.pdf>

<http://www.dwhyjzx.com/uploadfile/2007103182222265.pdf>

www.portaldalinguaportuguesa.org/

VII Anexos