

Palavras complexas na aquisição da morfologia do Português: estudo de caso

Marina Vigário & Paula Garcia

Laboratório de Fonética, FLUL-CLUL*

Abstract

We investigate morphological development in the first productions of a child (L). Until de age 2, the large majority of L's productive vocabulary is composed of simple (non-)inflected words, belonging to the Noun category (Verbs are much less frequent; Adverbs and words belonging to closed classes are residual). The consistent production of complex words appears at around 2;03, coinciding with a moment of lexical explosion. At about the same time Adjectives appear and Verbs and closed class words become increasingly frequent. Novel words coined by L (e.g., *mimpazinha* 'limpinha') appear roughly in the same period, showing that the child is already analysing words and applying morphological rules when the first complex words are consistently produced.

Keywords: acquisition, morphology, frequency, léxicon, complex words, word classes

Palavras-chave: aquisição, morfologia, frequência, léxico, palavras complexas, classes de palavras

0. Introdução

Na presente investigação, pretendemos caracterizar o desenvolvimento morfológico nas fases iniciais de produção de uma criança, L, a adquirir o Português Europeu (PE), com destaque para os processos de formação/emergência de palavras complexas. Trata-se de um tópico, até onde sabemos, ainda por explorar nos estudos publicados sobre a aquisição do PE.

Textos Seleccionados, XXVII Encontro Nacional da Associação Portuguesa de Linguística, Lisboa, APL, 2012, pp. 604-624, ISBN 978-989-97440-1-1.

* Este trabalho foi financiado por Fundos FEDER através do Programa Operacional Factores de Competitividade – COMPETE e por Fundos Nacionais através da FCT – Fundação para a Ciência e a Tecnologia no âmbito dos projectos *PTDC/LIN/70367/2006* e *PTDC/CLE-LIN108722/2008*. Parte desta investigação integrou a tese de mestrado *Palavras Complexas nas Primeiras Produções Infantis (estudo de caso)* (Garcia, 2010). Agradecemos a Amália Mendes ter-nos facultado os dados relativos à frequência de ocorrência das diferentes classes de palavras no corpus do Português incluído no C-ORAL-ROM e a Sónia Frota, que esteve na génese deste projecto.

Para além da cronologia da emergência de palavras simples e complexas, investiga-se a eventual relação entre a emergência/proporção de palavras complexas e a emergência/proporção das diferentes classes de palavras e dos diferentes tamanhos de palavra nas fases iniciais de produção linguística. Avalia-se ainda o papel da frequência no *input* na emergência/proporção dos diferentes tipos de palavras produzidos pela criança. O raciocínio subjacente a estas linhas de pesquisa é brevemente apresentado na secção 1. Os aspectos metodológicos envolvidos no tratamento dos dados de fala infantil e de fala adulta são detalhados na secção 2. Na secção 3 apresentamos os resultados da investigação, focando a atenção primeiramente nos dados da criança e depois nos do adulto. A discussão dos resultados é feita na secção 4. A secção 5 conclui o trabalho com uma breve síntese dos resultados e o levantamento de algumas linhas importantes para investigação futura.

1. Enquadramento

Como dissemos na secção anterior, até onde sabemos, a aquisição e desenvolvimento da morfologia, e em particular da formação de palavras complexas, nas primeiras produções infantis não foi antes estudada sistematicamente no Português Europeu (para outras línguas veja-se, por exemplo, Boysson-Bardies, 1999; Clark, 1998; Herschensohn, 2007; Dressler *et al.*, 2010)¹.

De acordo com Clark (1998), pelo menos no Inglês, as crianças começam a formar novas palavras muito cedo, aproximadamente na mesma altura em que usam a primeiras palavras flexionadas. Destaca-se aqui a composição como o processo usado inicialmente, entre os 8 meses e os 2 anos, para além da conversão (ou derivação zero). De acordo com a mesma autora e na mesma língua, a derivação afixal é um pouco mais tardia, observando-se uma quantidade crescente de palavras derivadas novas aproximadamente aos 3 anos.

Entre os diversos processos de formação de palavras, encontra-se na literatura, com alguma recorrência, indícios de que a composição tem um estatuto especial. Por exemplo, Booij (2007) afirma que a composição é o processo mais frequente nas línguas, e Dressler (2006) dá-o como existente em todas as línguas do mundo. Contudo, se bem que a composição seja atestada muito cedo na aquisição de línguas germânicas, ela é mais tardia em línguas como o Francês ou o Hebreu (Nicoladis, 2006; Dressler *et al.*, 2010). Esta constatação mostra que o curso do desenvolvimento morfológico depende, pelo menos até certo ponto, das características das línguas particulares.

¹ Embora ocupando-se de morfologia flexional, vale a pena mencionar Castro (2010), sobre aquisição da morfologia do plural nominal por crianças entre os 2 e os 5 anos. Um outros estudo recente, Oliveira (2011), investiga o desenvolvimento da flexão e da derivação nominal em crianças com perturbações específicas do desenvolvimento da linguagem e em crianças com desenvolvimento normal, em fases mais tardias (entre os 4;06 e os 7;10).

Na secção 3 analisamos a cronologia da emergência das palavras complexas em confronto com as palavras simples, e dentro das palavras complexas, identificamos o momento em que aparecem palavras exibindo os diferentes processos de formação morfológica da língua e com que frequência relativa.

É sabido que nem todas as categorias de palavras admitem processos de formação de palavras. Em línguas como o Português, apenas as categorias nominais, adjectivais, verbais e (algumas) adverbiais o permitem. Na literatura, os nomes aparecem frequentemente referidos como as primeiras categorias a surgir no léxico produtivo infantil, e em maior percentagem (e.g. Boysson-Bardies, 1999, e referências aí incluídas). Para o Português Europeu não conhecemos estudos específicos sobre a emergência e proporção das diferentes classes de palavra nas primeiras produções. No presente artigo investigamo-lo, tendo em vista determinar se a emergência de palavras complexas da criança está condicionada pela emergência de palavras pertencentes a categorias em que pode haver morfologia complexa.

Uma palavra complexa tem necessariamente um tamanho mínimo, de modo a poder conter um radical e outro morfema. Mesmo que não se saiba qual o formato típico das palavras complexas no Português (um contributo nesse sentido é dado na secção 3), é evidente que uma palavra complexa nesta língua não pode conter apenas uma sílaba, e não é fácil encontrar palavras deste tipo com apenas duas sílabas. Nesse sentido, importa perceber-se até que ponto a (in)capacidade de produzir palavras mais longas pode condicionar a emergência de palavras complexas. A relação entre a evolução dos diferentes formatos de palavra e o desenvolvimento morfológico será investigada neste trabalho. Importa adicionar a este respeito que também noutros campos tem sido defendido que aspectos da estrutura prosódica podem ter impacto no desenvolvimento morfológico e sintáctico – veja-se, por exemplo, Demuth (2007) e Lleó e Demuth (1999), sobre a importância do desenvolvimento prosódico para a emergência de determinantes/artigos no Espanhol, no Inglês, no Alemão e em Sesotho (língua Bantu); e Peters & Menn (1993), que defendem que, em línguas em que as fronteiras de morfemas coincidem com fronteiras de sílabas, as crianças produzem palavras polimorfémicas mais cedo.

Uma das razões intuitivas para a diferença na cronologia de emergência dos compostos nas línguas germânicas e nas línguas românicas, referida acima, é a diferença de frequência desses objectos nas línguas a que as crianças estão expostas (e.g. Nicoladis, 2006; Dressler *et al.*, 2010). Sabe-se que noutras áreas da gramática a frequência no *input* parece desempenhar um papel importante na frequência e cronologia de emergência de unidades e padrões linguísticos infantis – veja-se, por exemplo, os artigos compilados em Demuth (2006), e referências aí incluídas; e vários trabalhos recentes desenvolvidos sobre o Português, como Vigário, Freitas e Frota (2006); Vigário, Martins e Frota (2006); Vigário, Frota, Martins (2010). Por exemplo,

Vigário, Freitas e Frota (2006) mostram que a ordem de emergência e a frequência das palavras com diferentes formatos (ou tamanhos) nas primeiras produções infantis se correlacionam com a frequência desses formatos no *input* (especialmente considerando a fala adulta não dirigida à criança). A importância desta observação é tanto mais significativa quanto se sabe que as crianças que adquirem línguas como o Português, o Espanhol, ou o Francês apresentam um comportamento diferente do de crianças que adquirem o Inglês ou o Catalão, uma vez que produzem desde fases iniciais formatos de palavra menores e maiores do que um pé binário. Contudo, nem sempre a frequência no *input* parece explicar, por si só, o desenvolvimento infantil (e.g. Gülzow & Gagarina, 2007; Freitas, Frota, Vigário & Martins, 2006). Por outro lado, Vigário, Frota e Martins (2010) mostram que sempre que a frequência computada sobre *types* (ocorrência de palavras únicas) não coincide com a frequência computada sobre *tokens* (ocorrência de todas as palavras, incluindo repetições), são os valores computados sobre *tokens* que melhor se relacionam com os dados infantis.

O papel da frequência no *input* será investigado neste artigo em relação a diversos aspectos da produção infantil, designadamente à emergência/proporção de palavras simples e complexas, de palavras exibindo os diversos processos de formação de palavras permitidos no Português, de diferentes classes de palavras e de diversos formatos (ou tamanhos) de palavra. Serão considerados tanto dados relativos a contagens sobre *types* como a contagens sobre *tokens*.

Um outro aspecto aqui investigado diz respeito à evolução da dimensão do léxico de L. A relação entre a emergência de operações morfológicas e a dimensão do léxico infantil é previsível, uma vez que, para que a criança possa extrair os padrões morfológicos necessários ao estabelecimento das regras de formação de palavras, é necessário que o seu léxico tenha uma certa dimensão (Clark, 1998; Booij, 2007). Importa notar que, uma vez mais, não parece ser apenas na área da morfologia que se encontram etapas do desenvolvimento linguístico relacionadas com marcos da evolução lexical. Por exemplo, em Frota (2010) e Vigário, Frota e Matos (2011) mostra-se que existem marcos do desenvolvimento prosódico que coincidem no tempo com períodos de explosão lexical.

Para concluir este breve enquadramento, devemos fazer uma salvaguarda importante. O uso de formas complexas não implica que elas sejam analisadas. Efectivamente, as formas complexas podem ser armazenadas no léxico mental como um todo, não analisado. Isto é válido para as produções infantis, e também para as produções do adulto. É mesmo possível que uma palavra seja ao mesmo tempo susceptível de ser processada como um todo e analisada internamente (e.g. Libben, 2006). Para além disso, nem todas as palavras complexas têm o mesmo grau de transparência, e isto variará entre falantes adultos, e, certamente mais, do adulto para a criança. O que descreveremos neste trabalho será portanto, antes de mais, o *uso de formas complexas por parte da criança* (incluindo formas pouco transparentes) e não

directamente a *aquisição de processos morfológicos*. Contudo, quando a criança forma novas palavras que não coincidem com as produções do adulto, temos evidências inequívocas da extracção de generalizações e aplicação de regras (e.g. Clark, 1998; Booij, 2007). A identificação de formas não-coincidentes com as do adulto será por isso usada como o elemento mais importante para determinarmos em que ponto do desenvolvimento linguístico a criança está a formar novas palavras através dos recursos morfológicos da língua.

2. Metodologia

Neste estudo são analisadas as produções iniciais de uma criança monolíngue, L, falante do Português Europeu, registadas no diário linguístico *LumaLiDAOn* (Frota, Vigário & Jordão 2008). Foi feito um estudo longitudinal das primeiras produções de L, entre os 1;01 e os 3;03 (isto é, nos 27 primeiros meses de produções verbais), totalizando 17.229 palavras. Analisou-se a evolução mensal (i) das palavras simples e complexas (alvo), (ii) dos diferentes processos morfológicos exibidos pelas palavras complexas (alvo), (iii) das diferentes categorias gramaticais das palavras, e (iv) do tamanho das palavras (alvo e efectivamente produzidas). Em relação à generalidade dos parâmetros foi feita uma contabilização sobre *tokens* e sobre *types*. A computação dos dados relativos ao tamanho de palavra (em número de sílabas) e a separação dos dados em *tokens* e *types* foi feita com o recurso à ferramenta electrónica *FreP* (Martins, Vigário & Frota, 2009).

Os valores encontrados nas produções de L foram sistematicamente comparados com os de fala adulta. Sobre um extracto do *Corpus do Português Falado-CLUL/IC* (Portugal - Anos 90), totalizando 20 075 palavras (*tokens*), foi feita a contagem de frequência de (i) palavras simples e complexas, (ii) processos morfológicos exibidos pelas palavras complexas, (iii) tamanho das palavras (palavras-alvo e formas efectivamente produzidas). Foi feita uma contabilização sobre *tokens* e sobre *types*. A computação dos dados relativos ao tamanho de palavra e a separação dos dados em *tokens* e *types* foi feita com o recurso à mesma ferramenta electrónica. Para determinar a frequência relativa das diferentes classes de palavras foi usado um *corpus* do Português incluído no C-ORAL-ROM (*Integrated Reference Corpora for Spoken Romance Languages*), que totaliza 339 778 palavras (*tokens*) (ver Cresti & Monegnia, 2005). Este dados foram gentilmente facultados por Amália Mendes (Centro de Linguística da Universidade de Lisboa).

Para a classificação morfológica, seguiu-se genericamente o estabelecido para o Português Europeu (Cunha & Cintra, 1984; Rio-Torto, 1998; Mateus *et al.*, 2003; Villalva, 2000, 2008).

Consideraram-se as seguintes classes quanto à estrutura morfológica interna das palavras: (i) formas simples, com radical ou radical+marca de classe/vogal temática,

sem afixos (e.g. *que, não, casa, homem*) – código 0; (ii) formas simples com sufixo flexional (plural em Adjectivos e Nomes; Tempo-Modo-Aspecto e Pessoa e Número em Verbos (e.g. *peixes, podem, havia*) – código 1; (iii) formas no Infinitivo, Particípio Passado e Gerúndio (e.g. *estudar, chegado*) – código 11; (iv) formas com sufixação derivacional (e.g. *bastante, exactamente, tratamento*) – código 21; (v) formas com prefixação (e.g. *desculpa, bicicleta*) – código 22; (vi) formas com prefixação e sufixação (mas não simultânea, incluindo formas participiais) (e.g. *desligado*) – código 23; (vii) formas com sufixação e prefixação simultâneas (circunfixação, incluindo formas flexionadas) (e.g. *acalmar, empoleirar*) – código 24; (viii) formas exibindo derivação regressiva (e.g. *balanço*) – código 25; (ix) formas com dois radicais ou palavras (e.g. *televisão, fotografia*) – código 3; (x) formas com sufixos (-)avaliativos (e.g. *chuchinha, cremezinho*) – código 4; (xi) formas com reduplicação silábica (*popó, chichi, cocó*) – código 5; apenas são consideradas reduplicadas as formas em que o elemento repetido não corresponde a uma palavra existente; palavras como *pisca-pisca, chupa-chupa* foram classificadas como compostos.

Consideraram-se as seguintes classes quanto à categoria gramatical: (i) Nomes (N) (e.g. *mamã, carro*); (ii) Adjectivos (Adj) (e.g. *linda, verde*); (iii) Advérbios (Adv) (e.g. *ali, ainda*); (iv) Verbos (V) (e.g. *está, pára*); (v) Outros (Outr) – classes de palavras não susceptíveis de entrar em processos de formação de palavras (nesta classe os subtipos são irrelevantes para nossos propósitos); (vi) Amalgamas (Amlg-Adv - *daqui, dali*; Amlg-Outr - *dum, no*).

Para efeitos de trabalho, considerou-se que os processos originadores de palavras complexas eram reflectidos na composição interna morfológica. Isto é, por exemplo, uma palavra com prefixo é considerada genericamente uma instância de processo de prefixação (veja-se, apesar disso, o que é dito sobre este assunto no final da secção 1).

De igual modo, considerou-se complexa uma palavras decomponível em partes diferentes de flexão, de forma a que pelo menos um dos elementos tenha um formato e um significado aproximável a um morfema existente.

3. Resultados

3.1. Palavras simples/complexas

Nesta subsecção descrevemos o desenvolvimento de formas complexas e de formas simples, com e sem flexão, nas produções iniciais de L.

Como pode ser visto nos gráficos apresentados na Figura 1, as palavras simples flexionadas aparecem desde o início do 1.º ano de produções (constituindo entre 17% e 33% das produções mensais). Pelo contrário, até aos 2;02 praticamente não há ocorrência de palavras complexas e mesmo depois de 2;02 a proporção relativa de

palavras complexas é muitíssimo baixa (quer nas contagens sobre *tokens* quer nas sobre *types*).

Aos 2;03-2;04 dá-se um momento de explosão lexical (em *types* e em *tokens*). A partir de 2;05 as palavras complexas apresentam já valores muito estáveis (cerca de 5,6% dos *tokens*).

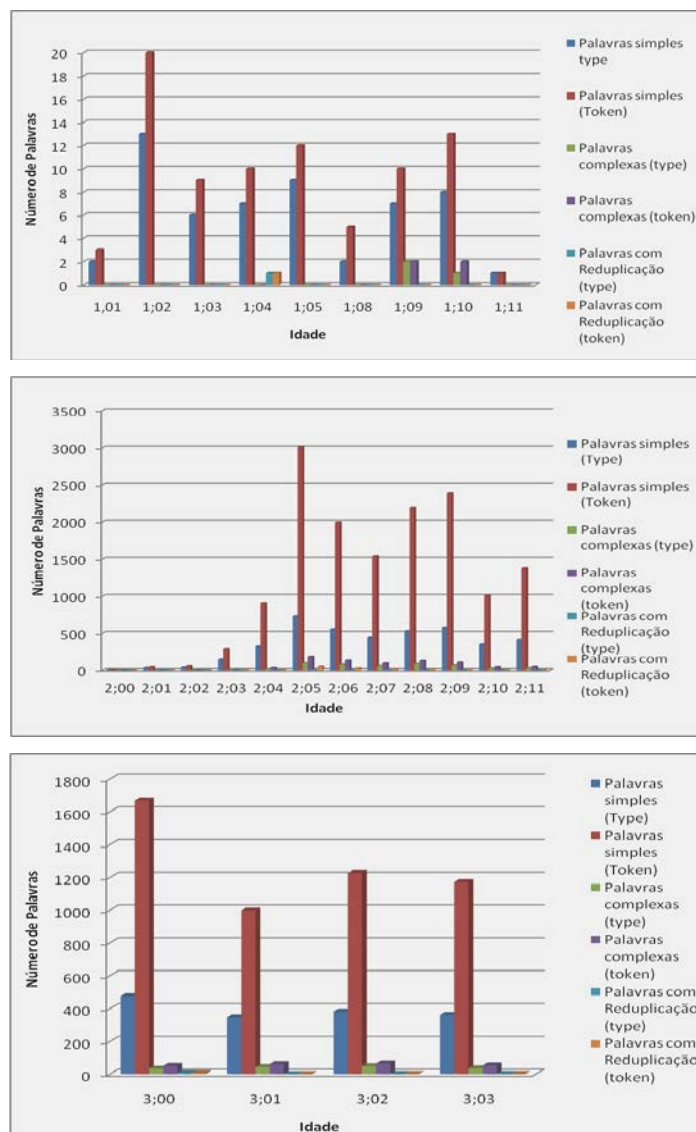


Figura 1: Evolução mensal das palavras simples, complexas e com reduplicação produzidas por L (valores absolutos em *types* e *tokens*).

A reduplicação ocorre proporcionalmente muito pouco em todos os períodos.

As palavras em (1) ilustram as formas complexas produzidas por L a partir dos 2;03.

- (1) *izão* ‘televisão’ (2;03) *fági* ‘desfaz’ (2;03)
fia ‘fotografia’ (2;04) *saco mi* ‘saco-de-dormir’ (2;05)
ada-fuva ‘guarda-chuva’ (2;05) *caço* ‘descalço’ (2;06)
futão ‘foguetão’ (2;06) *nigai* ‘desligar’ (2;07)
tepiado ‘sapateado’ (2;07) *barriguda* ‘barriguda’ (2;08)
casamento ‘casamento’ (2;08) *ventania* ‘ventania’ (3;01)

Interessantemente, palavras exibindo os diferentes processos de formação regular de palavras disponíveis para o Português são atestadas quase simultaneamente, como mostram os valores apresentados no Quadro 1.

	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

Quadro 1: Evolução mensal dos processos de formação atestados nas palavras produzidas por L (valores absolutos em *tokens* e *types*).

Também com grande significado é o facto de neste mesmo período serem atestadas formas complexas não-coincidentes com o alvo (ver Quadro 2).

Idade	Alvo	Realização	Alvo	Realização
2;04	chuchinha	chuchazinha		
2;05	sozinha	sozazinha	chuchinha	chuchazinha
	baixinho	baixoginho	pintinhas	Pintazinha
	sozinho	sozuzinho	molinho	Monzinho
	malinha	malazinha	tontinha	Tontazinha
	jeitinho	jeitozinho	ervinhas	Vevazinha
	cuequinha	quecazinha	sininho	Sinozinho
	pomadinha	pomadazinha	limpinha	mimpazinha
	gatinho	gatozinho	sequinho	Secozinho
	pouquinha	poucazinha	pianinho	pinhãozinho
	quentinha	quentizinha	bolsinho	Boçoquinho
	vaquinha	vacazinha	bolinhas	Bolazinha
	fuminhos	fumozinho	pombinho	bomboquinho
	fofinho	fofozinho	porquinhos	Poucoquinho
2;06	bonequinha	bonecazinha	pontinha	Pontazinha
	sozinha	sozazinha		

Quadro2: Formas produzidas por L não-coincidentes com o alvo

Se bem que todas as formas complexas identificadas inicialmente correspondam a formações z-avaliativas, importa notar que as produções não-coincidentes com o alvo não se confinam a este tipo de formação, já que aparecem também no mesmo período palavras flexionadas não-coincidentes com o alvo (ver 2), às quais se pode ainda juntar a palavra *biguda* ‘bigoduça’, produzida um pouco mais tarde, aos 2;08.²

- (2) *fóba* ‘sobe’ (imperativo) (2;04) *ovo* ‘ouço’ (2;05) *deteu* ‘deitou’ (2;05)
curre ‘corre’ (imperativo) (2;05) *diziu* ‘disse’ (3PSg) (2;06)

² Noutros casos, a palavra usada de modo não-coincidente com o alvo corresponde a uma forma disponível no paradigma flexional da palavra, como em *méxa* ‘mexe’ (presente indicativo, 3PSg) (2;04), *ponha* ‘põe’ (2;04), *fuba* ‘sobe’ (imperativo) (2;04), *bata* ‘bate’ (imperativo) (2;04), *moda* ‘morde’ (presente indicativo, 3PSg) (2;04), *dito* ‘disse’ (2;05), *veja* ‘vê’ (2;05), *mêta* ‘mete’ (presente indicativo, 3PSg) (2;05) (aqui como noutros lugares, o acento agudo ou circunflexo não-convencional representa uma vogal baixa ou média, respectivamente).

Para além disso, não se dá o caso de toda a sufixação avaliativa ser substituída por sufixação z-avaliativa, uma vez que a criança produz no mesmo período formas com sufixação avaliativa coincidentes com o alvo – e.g. *Fimbinho* (‘Fimbo’ é nome de personagem de história) (2;04), *Titinho* (‘Tito’ é nome de gato) (2;04), *santinha* (2;05).

Nesta fase inicial de produção de formas complexas, encontram-se tanto palavras sufixadas como prefixadas, tal como os exemplos em (1) ilustram.

Observemos agora os dados relativos à frequência de palavras simples e complexas na fala adulta (ver Figura 2).

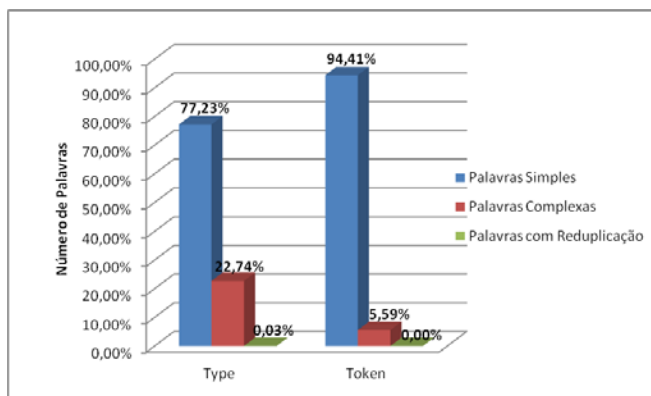
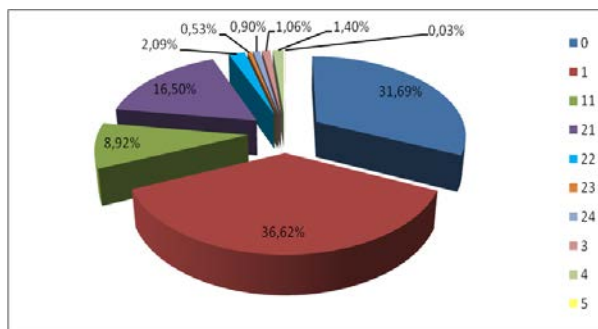


Figura 2: Palavras simples/complexas na fala adulta (valores percentuais em *tokens* e *types*)

Nestes dados é evidente a predominância de palavras simples, quer na contagem sobre *tokens* quer na contagem sobre *types*, sendo a proporção de palavras complexas especialmente baixa na contagem sobre tokens (5.6%).

A Figura 3 mostra a distribuição da frequência das palavras que exibem os diferentes processos morfológicos disponíveis na língua.



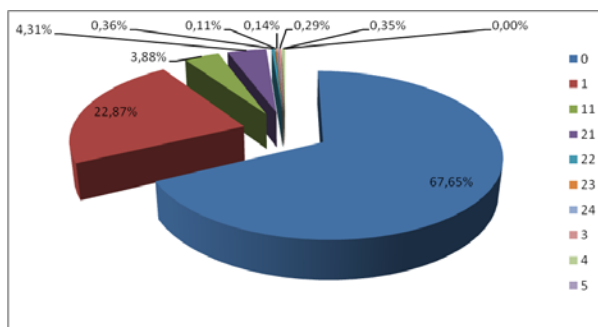
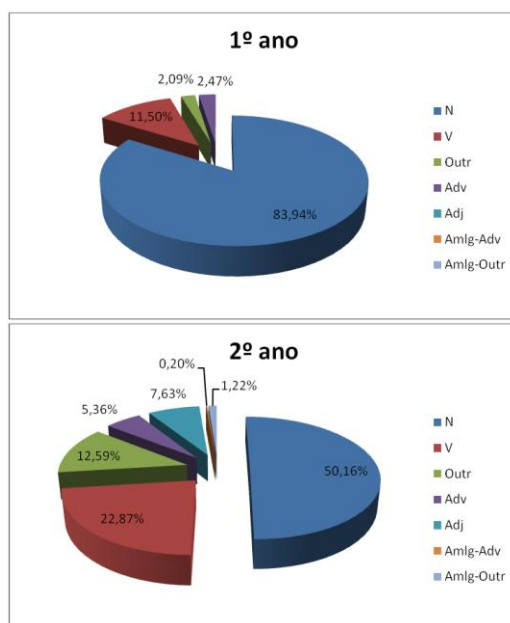


Figura 3: Processos morfológicos (Adulto) — palavras simples não-flexionadas (0), simples flexionadas (1), formas nominais verbais (11), derivadas com sufixação (21), com prefixação (22), com prefixação e sufixação (23), com circunfixação (24), compostas (3), com sufixação (z-)avaliativa (4), com reduplicação (5) (valores percentuais — painel superior: *types*; painel inferior: *tokens*).

Nos dados do adulto, as palavras formadas por derivação são claramente mais frequentes do que as compostas e com sufixos (z-)avaliativos. As formas derivadas e as formas nominais dos verbos atingem os 25.4% em *types*. Praticamente não existem formas com reduplicação.

3.2. Classes de palavras

Os gráficos da Figura 4 mostram a evolução da distribuição das palavras produzidas por L pelas classes de palavras da língua na contagem sobre *types*.



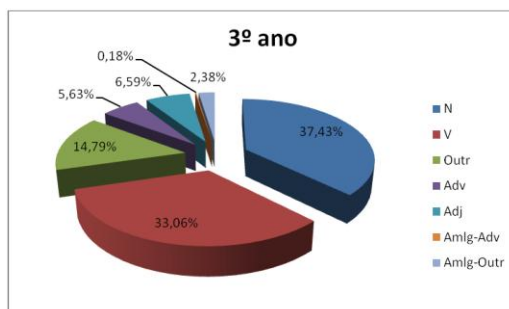
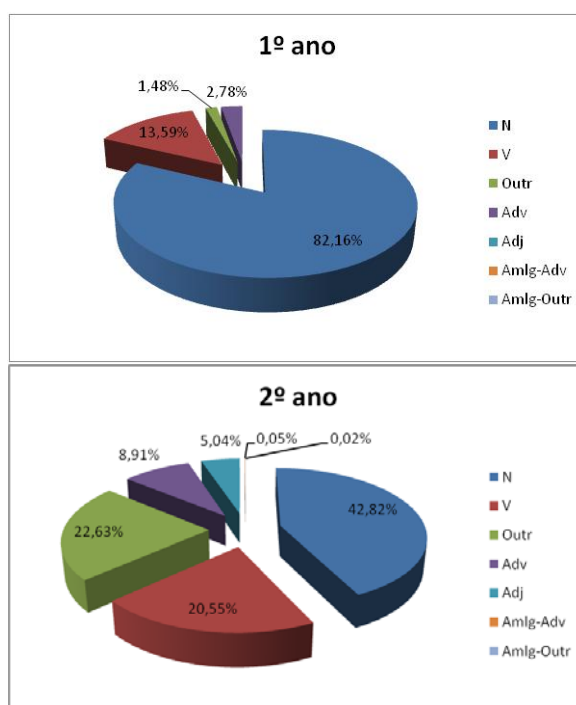


Figura 4: Evolução das classes de palavras produzidas por L por ano de produção (valores percentuais computados sobre *types*).

No 1.º ano de produção é evidente a preponderância de N sobre todas as categorias ($N \gg V (> Adv > Outr)$). Por outro lado, quase só são atestadas classes abertas, salientando-se ainda a inexistência de formas Adjectivais. No 2.º ano de produção, dá-se uma explosão categorial (desde 2;01-2;02). A categoria Adj aparece ao 2;01 e há um aumento de todas as classes, à custa da diminuição do espaço ocupado por N. No 3.º ano de produção, há um reforço do observado no ano anterior, sobretudo no que respeita ao aumento da classe V e à diminuição da classe N ($N \sim V > Outr > Adv > Adj$).

Como pode verificar-se nos gráficos da Figura 5, as mesmas tendências gerais são visíveis na contagem sobre *tokens*.



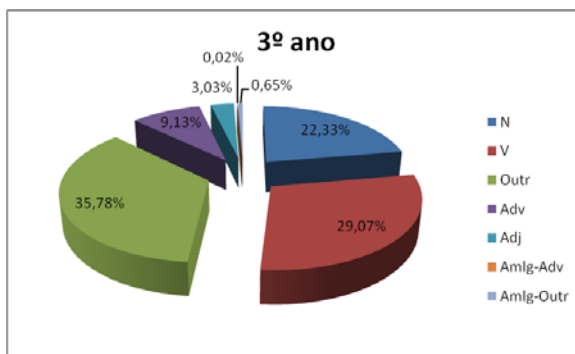


Figura 5: Evolução das classes de palavras produzidas por L por ano de produção (valores percentuais computados sobre *tokens*).

Deve salientar-se na contagem sobre *tokens* a muito diminuta ocorrência de palavras pertencentes à categoria Outr no 1.º ano de produção e a sua elevada percentagem no 2.º (N > Outr > V) e, sobretudo, no 3.º, altura em que esta se torna a categoria mais frequente (Outr > V > N > Adv > Adj).

A distribuição das palavras pelas diferentes classes na fala adulta é apresentada na Figura 6 (dados sobre *token* apenas).

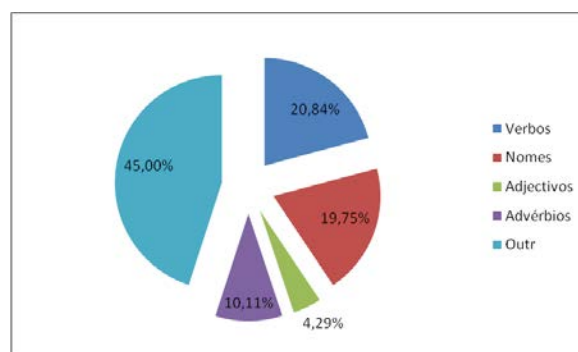


Figura 6: Distribuição das classes de palavras na fala adulta (valores percentuais computados sobre *tokens*).

A classe Outr predomina claramente sobre todas as restantes, atingindo as classes V e N menos de metade do número de ocorrências daquela classe. As palavras da classe Adv ocorrem em números razoavelmente altos, sendo as da classe Adj proporcionalmente muito menos frequentes (Outr > V~N > Adv > Adj).

3.3. Tamanho de palavra

Observemos agora a evolução dos formatos de palavras efectivamente produzidos por L (ver Figura 7).

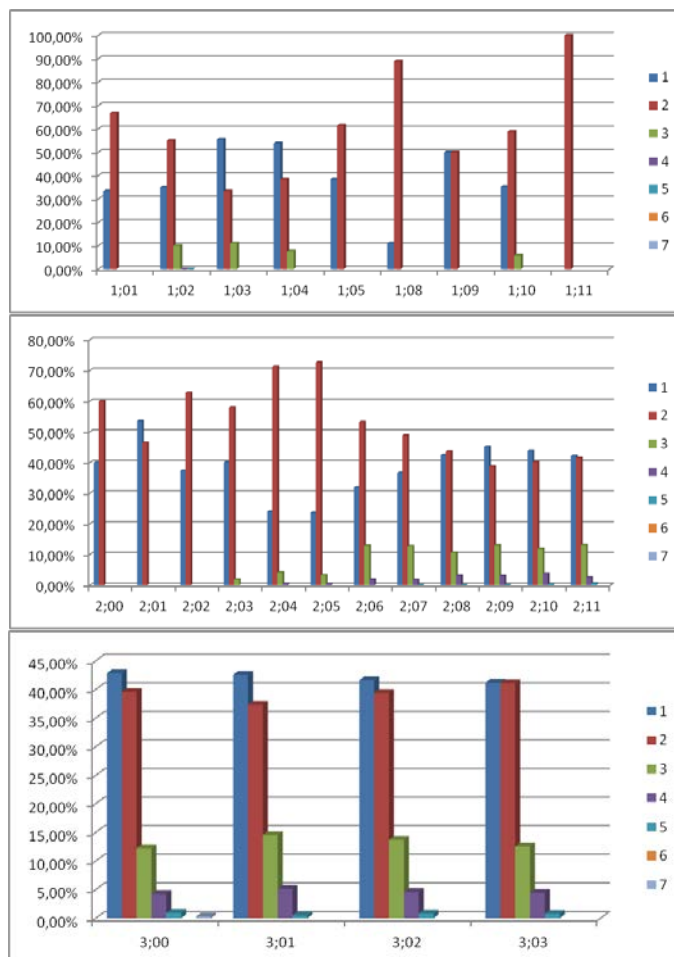


Figura 7: Evolução mensal dos formatos de palavras (em número de sílabas) efectivamente produzidos por L (valores percentuais computados sobre *tokens*).

Os tamanhos superiores a duas sílabas são muito raros no 1.º ano de produção e não ocorrem nesse período palavras com mais de 3 sílabas. Por outro lado, observa-se nessa altura grande variabilidade nas proporções dos formatos atestados, ora predominando formatos monossilábicos ora dissilábicos.

As palavras com 3 e 4 sílabas estão consistentemente presentes apenas a partir dos 2;04, apresentando valores expressivos só desde 2;06.

Desde 2;06-2;08 até ao final do período observado a proporção dos diferentes formatos de palavra apresenta muito pouca variação.

Também no que respeita às formas-alvo (formato do adulto, independentemente de como L as realiza) os tamanhos superiores a duas sílabas são pouco frequentes no 1.º ano de produção, apesar de haver ocorrências de palavras-alvo de quatro sílabas. Também aqui há grande variabilidade no 1.º ano. Os formatos com 3 e 4 sílabas estão consistentemente presentes, com valores expressivos, um pouco mais cedo do que esses formatos efectivamente produzidos, desde 2;01. É também desde os 2;01, até ao final, que a proporção dos diferentes formatos de palavra passa a apresentar pouca variação.

Vejamos agora a distribuição dos formatos de palavra no nosso *corpus* de fala adulta tendo em conta as características morfológicas das palavras.

As Figuras 8 e 9 mostram a distribuição dos formatos de palavras no universo das palavras simples e no das palavras complexas, respectivamente.

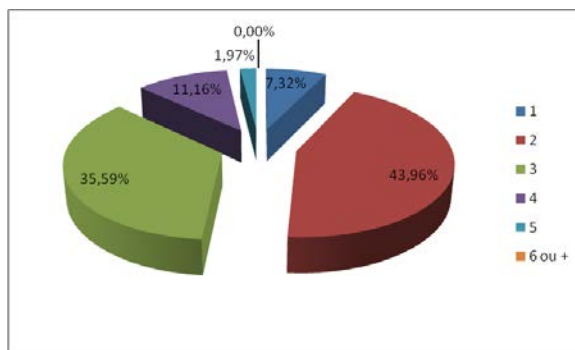


Figura 8: Distribuição dos formatos das palavras simples (em número de sílabas) na fala adulta (valores percentuais computados sobre *tokens*)

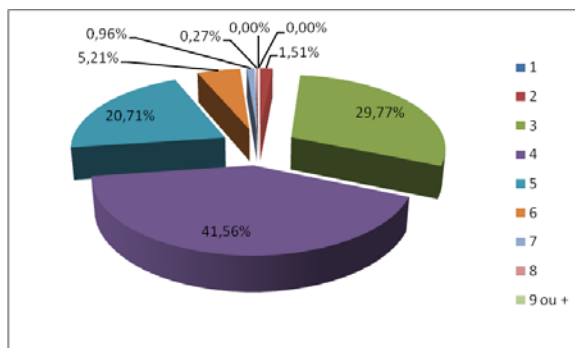


Figura 9: Distribuição dos formatos das palavras complexas (em número de sílabas) na fala adulta (valores percentuais computados sobre *tokens*)

Pode constatar-se a assimetria muito marcada na distribuição dos formatos das palavras simples e das palavras complexas: os formatos pequenos, mono ou dissilábicos, atingem os 51% das palavras simples, enquanto constituem menos de 1% das palavras complexas. Por seu turno, os formatos com 4 ou mais sílabas são relativamente pouco frequentes nas palavras simples (13%), enquanto constituem o formato claramente maioritário das palavras complexas (69%).

4. Discussão

Nesta secção discutimos sumariamente os resultados apresentados na secção precedente. Até aos 2 anos a criança usa quase exclusivamente formas simples, flexionadas ou não. A preponderância das formas simples pode ser explicada pelo desenvolvimento morfológico infantil, com formas simples a precederem a emergência das formas complexas. Contudo, é também evidente o claro predomínio de palavras simples no

input, o que pode ter impacto na cronologia da emergência das palavras complexas no Português.

Mesmo a partir dos 2 anos, a proporção de palavras complexas em L é baixa, quer na contagem sobre *types* quer sobre *tokens*. Mas tanto a proporção de palavras complexas contadas sobre *types* como a contada sobre *tokens* se aproxima dos valores do adulto contabilizados sobre *tokens* (5.6%), e não dos contabilizados sobre *types* (22.7%). Estes dados reforçam o papel preponderante das palavras-*tokens* no *input* sobre as palavras-*types* no desenvolvimento linguístico (Vigário, Frota & Martins 2010).

Para além do *input*, os resultados favorecem a hipótese de que a dimensão do léxico da criança desempenha um papel determinante no desenvolvimento morfológico (Booij 2007; ver também Frota, 2010 e Vigário, Frota & Matos, 2011 sobre a relação entre dimensão do léxico e desenvolvimento prosódico). Verificou-se que até aos 2;02 L nunca produz mensalmente mais de 43 *types* e 59 *tokens*. Aos 2;03 L produz 160 *types* e 317 *tokens* e os valores crescem muito, subsequentemente. É aos 2;05, na mesma altura que se dá uma evidente explosão lexical (que começa por volta dos 2;03-2;04 e atinge um pico em 2;05), que a proporção de palavras complexas se torna sistemática (cerca de 5,6% dos *tokens*) e muito próxima da verificada no adulto (5,6% dos *tokens*).

Quando as palavras complexas surgem sistematicamente, aparecem simultaneamente formas exibindo os diversos processos de formação de palavras da língua (derivação, composição, sufixação (z-)avaliativa). Contrariamente ao que vimos acima, aqui as propriedades do léxico usado por L não espelham as encontradas no *input* (considerados *types* ou *tokens*): as formas derivadas são muito mais frequentes no adulto do que os restantes tipos de palavras complexas, mas isso não conduz à emergência mais precoce de palavras derivadas face às compostas e com sufixação (z-)avaliativa. Importa, porém, frisar que se analisou aqui fala adulta não-dirigida à criança. Contudo, pelo menos as formações com sufixação (z-)avaliativa são certamente mais frequentes na fala dirigida à criança do que na restante fala adulta. O impacto dessa eventual diferença não pode, por isso, ser aqui descartado, pelo menos para a cronologia da emergência e frequência relativa das formações com sufixação (z-)avaliativa nas produções de L.

Entre as observações mais interessantes aqui feitas, salienta-se o facto de no momento em que aparecem as primeiras palavras complexas (2;03-2;04) surgirem já palavras formadas pela criança não-coincidentes com o alvo. Como vimos na secção 1, as formas não-coincidentes com o alvo dão-nos evidência de que a criança não está só a usar palavras eventualmente registadas como um todo no seu léxico mental, sem análise interna: ela está a usar palavras que analisa morfológicamente, reveladoras de que a criança já extraiu regularidades morfológicas que lhe permitem formar novas palavras.

Mais, formas como *sozuzinho* ‘sozinho’, *sozazinha* ‘sozinha’ ou *mimpazinha* ‘limpinha’, atestadas aos 2;05-2;06, mostram, não apenas que a criança já tem disponível o processo de formação de z-avaliativos, mas também que ela sabe que o

sufixo *-zinh-* se junta a bases adjectivais com marcas de concordância de género, com estatuto de palavra morfológica (e.g. Villalva 2000).

Interessantes são também as realizações do alvo ‘sombriinha’ (sinónimo de *guarda-chuva*) como *fomba*, *fombazinha* e *fombainha* (todas co-ocorrendo aos 2;05), tal como as realizações *ventozinha* e *ventinha* para ‘ventoinha’. Estas formas permitem discutir a pertinência de se considerar na análise das primeiras produções infantis palavras complexas pouco transparentes no léxico adulto. Tanto *sobrinha* como *ventoinha* são formações semanticamente pouco transparentes: se bem que o radical seja reconhecível em ambos os casos, o sufixo não tem os valores semânticos produtivos de diminutivo ou avaliativo. Contudo, isso não impediu a criança de decompor as palavras, como as formas acima demonstram. Estes dados sugerem que a opção feita aqui de analisar o léxico infantil à luz dos dados adultos (como se faz na pesquisa do desenvolvimento de outras áreas da gramática), sem exclusão à partida de formas semanticamente opacas ou pouco transparentes, foi adequada.

Em relação à evolução da distribuição das diferentes classes de palavras, verifica-se que no 1.º ano de produção quase só surgem palavras pertencentes a classes abertas. Este facto pode ser relacionado com as fases iniciais do desenvolvimento sintáctico, mas não só. Efectivamente, muitas das palavras de classe fechada que nesta fase não são produzidas são clíticos fonológicos. Estes objectos, que do ponto de vista fonológico se destacam pouco das palavras a que se cliticizam, são de mais difícil segmentação, por um lado, e são por natureza não-proeminentes, por outro, o que pode favorecer a sua emergência mais tardia.

Interessantemente, aqui, novamente, o desenvolvimento infantil não é convergente com os dados de frequência no *input*: no 1.º ano de produção estão praticamente ausentes palavras das categorias *Outr*, sendo só no 3.º ano que essa categoria atinge uma predominância próxima da verificada na fala adulta (36% em L e 45% na fala adulta).

A predominância inicial da categoria N no léxico produzido por L é convergente com o reportado para outras línguas (e.g., Boysson-Bardies, 1999), o que sugere que ela pode ter motivações mais gerais, eventualmente relacionadas com o desenvolvimento linguístico e/ou cognitivo da criança, independentemente das características da língua particular, como a frequência no *input*.

Sem surpresa, os resultados reportados mostram que o facto de no 1.º ano de produção quase só estarem representadas classes abertas, que são as que podem estar envolvidas em processos regulares de formação de palavras, significa que a quase inexistência de palavras complexas antes dos 2 anos não é explicável pela ausência das classes de palavras que contêm palavras complexas.

No que respeita aos formatos de palavra, os dados revelam diferenças na evolução dos formatos das palavras-alvo e dos formatos das palavras efectivamente produzidas. As palavras-alvo com 3 e 4 sílabas aparecem consistentemente e com valores expressivos desde 2;01, precedendo proximamente a produção sistemática de formas complexas, aos 2;03-2;04. Os formatos maiores nas palavras efectivamente produzidas,

por seu turno, apenas aparecerem consistentemente mais tarde (2;06). Tal parece sugerir que existe uma relação entre o aparecimento da morfologia complexa e os formatos maiores das palavras no léxico *mental* da criança. L não precisa de ter a capacidade de *produzir* efectivamente o formato das palavras complexas, para que elas ocorram.

Os resultados referentes aos formatos efectivamente produzidos pela criança não reflectem inteiramente a distribuição dos formatos de palavras na fala adulta, em particular, as proporções sistemáticas dos diversos tamanhos de palavra na criança entre 2;01-3;03, uma vez que palavras com 1 e 2 sílabas ocorrem com frequência semelhante e são muito mais frequentes do que os restantes formatos e os trissílabos representam cerca de 10% dos formatos produzidos. Esta observação é válida quer considerando a distribuição dos formatos das palavras simples, quer a distribuição das palavras complexas (ver dados na secção precedente), quer ainda ignorando o tipo de palavra – olhando para o mesmo corpus de fala adulta que nós, Vigário, Freitas & Frota (2006) encontram 44% de dissílabos, 29% de monossílabos e cerca de 19% de trissílabos.

Em termos gerais, as palavras complexas, e em particular as compostas, parecem ser mais tardias do que o reportado para línguas como o Alemão (e.g. Grimm 2007). Tal pode ser entendido à luz da hipótese de Peters & Menn (1993) de que, em línguas em que as fronteiras de morfemas coincidem com fronteiras de sílabas, as crianças produzem palavras polimorfémicas mais cedo do que em línguas em que essa coincidência não se verifique. No Português, como nas restantes línguas Românicas, não só as fronteiras de morfema não se alinham frequentemente com as fronteiras de sílaba (a não ser nos limites das palavras), como a ressilabificação opera entre palavras, eliminando muitas vezes esse alinhamento. Menos pistas prosódicas para a segmentação dos morfemas podem, pois, resultar na emergência mais tardia da morfologia complexa nesta língua.

A emergência mais tardia da composição é, no entanto, também compatível com o papel desempenhado pela frequência no *input* no desenvolvimento linguístico (e.g. Nicoladis 2006, Dressler *et al.* 2010): no Português, o facto de este processo ser muito pouco frequente no *input* faz das palavras formada por composição candidatas a aparecer mais tarde do que nas línguas Germânicas, onde esse processo é muito produtivo.

5. Conclusão

Neste artigo, apresentamos o que julgamos ser a primeira descrição do desenvolvimento da morfologia derivacional nos primeiros estádios de aquisição no Português Europeu. Para além de identificarmos a evolução das produções de uma criança no que se refere à proporção palavras simples/palavras complexas, às diferentes classes de palavras e aos diversos formatos de palavras produzidas/alvo, confrontámos essa evolução com os dados correspondentes na fala adulta e tendo em conta a evolução do léxico.

Verificou-se que a frequência no *input* se relaciona bem com o padrão de emergência/proporção das palavras simples e complexas, mas não com a cronologia relativa de emergência sistemática das palavras derivadas, compostas e com sufixação (z-)avaliativa, nem com ordem de emergência/proporção dos diversos tipos de palavras, ou os formatos das produções da criança.

Por seu turno, identificámos um momento de explosão lexical (aos entre os 2;03 e os 2;05) que acompanha muito proximamente o momento em que aparecem palavras complexas consistentemente e com valores expressivos (aos 2;03-2;04), bem como um momento de explosão de classes de palavras (aos 2;01-2;02). Também na mesma altura (2;01) aparecem sistematicamente e com valores expressivos as palavras-alvo maiores, com 3 ou mais sílabas.

Porque este estudo é baseado em dados de uma única criança, é fundamental alargar a observação dos aspectos aqui analisados a outras crianças. Isso é tanto mais importante quanto a dimensão e constituição do léxico parecem desempenhar um papel muito importante na emergência de palavras complexas e no estabelecimento das generalizações morfológicas que resultam na extracção das regras morfológicas da língua pela criança, e esta é uma área em que se espera alguma variação entre crianças.

Para além disso, importa alargar o escopo temporal coberto. Aqui confinámos as observações aos momentos iniciais da produção de L até aos 3;03. Contudo, o conhecimento morfológico está longe de concluído nesta fase (e.g. Castro 2010).

Para uma aferição mais fina do efeito do *input* no desenvolvimento morfológico, será importante estender a observação a dados da fala dirigida a cada criança.

Finalmente, sobressai desta investigação a necessidade de reforçar o estudo do desenvolvimento morfológico em interacção com o desenvolvimento do léxico e de outras áreas da gramática, bem como do papel desempenhado pela frequência no *input* no desenvolvimento linguístico.

Referências

- Booij, G. (2007) *The Grammar of words: An Introduction to Morphology*. Oxford Linguistics: Second Edition.
- Boysson-Bardies, B. (1999) *How Language Comes to Children: From Birth to Two Years*. Cambridge, MA: MIT Press.
- Castro, A. (2010) Aquisição de morfologia de plural em português europeu: sobre a produtividade das regras. In Ana Maria Brito, Fátima Silva, João Veloso e Alexandra Fiéis (orgs.) *XXV Encontro Nacional da Associação Portuguesa de Linguística. Textos seleccionados*. Porto: Associação Portuguesa de Linguística, pp. 277-289.
- Clark, E. (1998) Morphology in Language Acquisition. In A. Spencer & A. Zwicky (eds.) *The Handbook of Morphology*. Blackwell Publishers, pp. 374-389.
- Corpus do Português Falado. *Documentos Autênticos*. Centro de Linguística da Universidade de Lisboa e Instituto Camões.

- Cresti, E. & M. Monegna (eds.) (2005) *C-ORAL-ROM: Integrated ReferEnce Corpora for Spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins Publishing Company, Studies in Corpus Linguistics n° 15 (com DVD).
- Cunha, C. & L. Cintra (1984) *Nova Gramática do Português Contemporâneo* (2.^a edição). Lisboa: Sá de Costa.
- Demuth, K. (2007) Acquisition at the prosody-morphology interface. In A. Belikova, L. Meroni & M. Umeda (eds.) *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition (GALANA)*. Somerville, MA.: Cascadilla Proceedings Project, pp. 84-91.
- Demuth, K. (ed.) (2006) Special on Issue Crosslinguistic Perspectives on the Development of Prosodic Words. *Language and Speech* 49(2).
- Dressler, W.U., L.E Lettner & K. Korecky-Kröl (2010) First language acquisition of compounds. With special emphasis on early German child language. In S. Scalese & I. Vogel (eds.) *Cross-Disciplinary Issues in Compounding*. Amsterdam: John Benjamins, pp. 323-344.
- Freitas, M. J., S. Frota, M. Vigário & F. Martins (2006) *Efeitos prosódicos e efeitos de frequência no desenvolvimento silábico em Português Europeu*. In XX Encontro Nacional da Associação Portuguesa de Linguística. Textos Seleccionados. Lisboa: APL, pp. 397-412.
- Frota, S. (2010) Prosodic structure in early child speech: Evidence from intonation, tempo and coda production. Comunicação proferida no *Workshop on Prosodic Development*. Universitat Pompeu Fabra, Abril.
- Frota, S., M. Vigário, F. Martins & M. Cruz (2010) *FrePOP*. Version 1.0. (ISBN 978-989-95713-2-7). (<http://frepop.fl.ul.pt>)
- Frota, S., Vigário, M. & Jordão, R. (2008) *LumaLiDaOn*. Lisboa: Laboratório de Fonética da FLUL (<http://www.fl.ul.pt/laboratoriofonetica/lumalidaon.htm>).
- Garcia, Paula (2010) *Palavras Complexas nas Primeiras Produções Infantis (estudo de caso)*. Dissertação de Mestrado, Universidade Católica Portuguesa / Escola Superior de Saúde do Alcoitão.
- Grimm, A. (2007) The development of Early Prosodic Word Structure in Child German: Simplex Words and Compounds. Dissertação de Doutoramento, Potsdam University.
- Gülzow, I & N. Gagarina (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.
- Libben, G. (2006) Why Study Compound Processing? An overview of issues. In G. Libben & G. Jarema (eds.) *The Representation and Processing of Compound Words*. Oxford Linguistics, pp.1-22.
- Lleó C. & K. Demuth (1999) Prosodic constraints on the emergence of grammatical morphemes: Crosslinguistic evidence from Germanic and Romance languages. In A.

- Greenhill, H. Littlefield & C. Tano (eds.) *Proceedings of the 23rd Annual Boston University Conference on Language Development*. Somerville, MA.: Cascadilla Press, pp. 407-418.
- Martins, F., M. Vigário & S. Frota (2009) *FreP - Frequências no Português*. Version 2.0. Beta. Software in CD-ROM (registered at IGAC, 18th Dec. 2009, no. 209/2010)
- 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 Português* (6.^a edição). Lisboa: Editorial Caminho.
- Nicoladis, E. (2006) Preschool Children's Acquisition of Compounds. In G. Libben & G. Jarema (eds.) *The Representation and Processing of Compound Words*. Oxford Linguistics, pp. 96-124.
- Oliveira, Ana Cristina (2011) *Flexão Nominal e Nominalizações Agentivas e Instrumentais em Crianças com Perturbações Específicas do Desenvolvimento da Linguagem Estudo de Caso Comparativo*. Dissertação de Mestrado, Instituto Politécnico de Setúbal, Escola Superior de Setúbal e Universidade Nova de Lisboa.
- Peters, A. & L. Menn (1993) False starts and filler syllables: Ways to learn grammatical morphemes. *Language* 69: 742-777.
- Rio-Torto, G. (1998) *Morfologia Derivacional. Teoria e Aplicação ao Português*. Porto: Porto Editora.
- 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., Freitas, M. J. & Frota, S. (2006) Grammar and frequency effects in the acquisition of prosodic words in European Portuguese. (Special on Issue Crosslinguistic Perspectives on the Development of Prosodic Words, guest-edited by K. Demuth). *Language and Speech* 49(2), 175-203.
- Vigário, M., S. Frota & F. Martins F. (2010) A frequência que conta na aquisição da fonologia: *types* ou *tokens*. In Ana Maria Brito, Fátima Silva, João Veloso & Alexandra Fiéis (orgs.) *XXV Encontro Nacional da Associação Portuguesa de Linguística. Textos seleccionados*. Porto: Associação Portuguesa de Linguística, 749-767.
- Vigário, Marina Sónia Frota & Nuno Matos (2011) Early prosodic development in European Portuguese: Evidence from intonation and tempo. Comunicação apresentada na *The International Child Phonology Conference*, Universidade de York, Junho.
- Villalva, Alina (2000) *Estruturas Morfológicas. Unidades e Hierarquias nas Palavras do Português*. Lisboa: Fundação Calouste Gulbenkian.
- Villalva, Alina (2008) *Morfologia do Português*. Lisboa: Universidade Aberta.